Bayesian inference for generalized stochastic population growth models with application to aphids

Colin S. Gillespie and Andrew Golightly

Newcastle University, Newcastle Upon Tyne, UK.

Summary. In this paper we analyse the effects of various treatments on cotton aphids (*Aphis gossypii*). The standard analysis of count data on cotton aphids determines parameter values by assuming a deterministic growth model and combines these with the corresponding stochastic model to make predictions on population sizes, depending on treatment. Here, we use an integrated stochastic model to capture the intrinsic stochasticity, of both observed aphid counts and unobserved cumulative population size for all treatment combinations simultaneously. Unlike previous approaches, this allows us to explicitly explore and more accurately assess treatment interactions. Markov chain Monte Carlo methods are used within a Bayesian framework to integrate over uncertainty associated with the unobserved cumulative population size and estimate parameters. We restrict attention to data on aphid counts in the Texas High Plains obtained for three different levels of irrigation water, nitrogen fertiliser and block, but note that the methods we develop can be applied to a wide range of problems in population ecology.

Key Words: Cotton aphid, Markov jump process, Moment closure approximation, Markov chain Monte Carlo.

1. Introduction

The importance of the cotton aphid as an insect pest is well documented (see for example Leclant and Deguine, 1994). Consequently, accurate modelling of cotton aphid populations is of economic importance (Rummel et al., 1995; Xia et al., 1999). In the recent literature, attempts to model cotton aphid dynamics have included the use of dispersion models (Celini and Vaillant, 2004), computer simulation models (Giarola et al., 2006) and nonlinear regression models (Matis et al., 2007).

Whilst deterministic population growth models for insect counts are reasonably well developed (see for example Matis et al., 2006, 2008) little work has taken into account stochastic effects that are particularly important when population sizes are small (Zheng and Ross, 1991). Matis et al. (2007) propose a stochastic model for aphid growth, and use this to make predictions on population sizes using parameter values obtained from the analogous deterministic model. The model itself assumes a birth rate dependent on current aphid numbers and a death rate dependent on the latter plus the unobserved cumulative population size, giving a bivariate Markov jump process with an unobserved component. We build on the work of Matis et al. (2007) by using the stochastic model to infer unknown parameters (namely birth and death rates) using discrete time observations on aphid counts only. The task of inferring parameters governing Markov jump processes with discrete state space has been examined by Boys et al. (2008) amongst others. It is found therein that working with the discrete stochastic model can be computationally prohibitive for models involving more than three reactions and two species. We therefore approximate the

model and perform exact Bayesian inference for this approximate model. Such approximations have, in the past, focused on the use of coupled stochastic differential equations to replace the underlying discrete valued process (see for example Golightly and Wilkinson, 2005; Heron et al., 2007; Purutcuoglu and Wit, 2007). Such an approach, however, is by no means straightforward, and can be computationally prohibitive for large datasets (Golightly and Wilkinson, 2008). Our approach also ignores discreteness but assumes that transition densities for the number of aphids (and the cumulative population size) are Normal, with parameters determined using moment closure techniques (Milner et al., 2009). Hence, given complete information on both aphid counts and cumulative population size, Bayesian inference is straightforward for our approximate model. However, we do not observe cumulative population size and therefore integrate over the uncertainty associated with it via a computationally efficient Markov chain Monte Carlo (MCMC) algorithm. In brief, we implement an MCMC scheme to alternate between draws of the latent data (conditional on the current parameter value and the observed data), and the parameter value (conditional on the observed and latent data). The methods we develop are general and can in theory be applied to any stochastic kinetic model.

As well as using synthetic data to validate the methodology, the algorithm is implemented using the data in Matis et al. (2008) (see also Matis et al., 2007), that is, using measurements on cotton aphid counts in the Texas High Plains. The data were obtained for several controlled conditions including various irrigation water and nitrogen fertiliser levels. We apply our method to estimate the effect of these treatments on the stochastic birth and death rates. In addition, we integrate over the uncertainty associated with the rate constants to give predictive distributions of the total number of aphids at future time points. This is particularly effective for comparing the effects of different treatments on aphid numbers and is therefore of practical importance.

The remainder of this paper is organised as follows: in Section 2, a description of the model and the data is given. We also outline the moment closure approximation which will form the basis of the inference algorithm described in Section 3. The algorithm is applied to real and synthetic data in Section 4 before conclusions are drawn in Section 5.

2. The model

2.1. Description of the Model

Following Matis et al. (2006), we assume a linear aphid population birth rate of $\lambda N(t)$ where $N(t)$ denotes the size of the aphid population at time t. As discussed in Prajneshu (1998), aphids excrete honey-dew, forming a cover on the leaf surface and causing aphid starvation. As the area covered by excretion at time t is proportional to the cumulative population size at time t, $C(t)$, we assume a death rate of $\mu N(t)C(t)$ and for simplicity, assume that there is no removal or decomposition of honey-dew (Matis et al., 2006, 2007, 2008). The model can be represented by the coupled (pseudo) reactions,

$$N \xrightarrow{\lambda} 2N + C \tag{1}$$

$$N + C \xrightarrow{\mu} C \tag{2}$$

since clearly, an occurrence of (1) will lead to a unit increase in both N and C whereas the reaction in (2) will give a unit decrease in N but leave C unchanged. Such models are known in the literature as stochastic kinetic models; see Wilkinson (2006) for a concise

introduction. We can express the probabilistic laws governing the time evolution of the process by considering a small interval $(t, t+dt]$, namely

$$\Pr\{N(t+dt) = n(t)+1,\, C(t+dt) = c(t)+1 \,|\, n(t), c(t)\} = \lambda n(t)dt + o(dt) \quad (3)$$

$$\Pr\{N(t+dt) = n(t)-1,\, C(t+dt) = c(t) \,|\, n(t), c(t)\} = \mu n(t)c(t)dt + o(dt), \quad (4)$$

and we adopt the convention that upper case denotes random variable and lower case denotes the realisation. The expressions in (3) and (4) describe a Markov jump process where each event occurs at a particular rate dependent on the current state of the system. Note that ignoring stochasticity leads to a deterministic model where the time course behaviour of $N(t)$ and $C(t)$ is described by the set of differential equations,

$$\begin{cases} \dfrac{dN(t)}{dt} = \lambda N(t) - \mu N(t)C(t) \\ \dfrac{dC(t)}{dt} = \lambda N(t) \end{cases}$$

Let $p_{n,c}(t)$ denote the probability that there are n aphids in the population at time t and a cumulative population size of c. Then, in principle, inference may proceed by solving the forward Kolmogorov equation to obtain $p_{n,c}(t)$. This is achieved by writing $p_{n,c}(t+\Delta t)$ as the sum of the probabilities of arriving in state $(n,c)'$ at time $t+\Delta t$:

$$p_{n,c}(t+\Delta t) = \lambda(n-1)p_{n-1,c-1}(t)\Delta t + \mu c(n+1)p_{n+1,c}(t)\Delta t + \{1 - n(\lambda+\mu c)\Delta t\}\, p_{n,c}(t)\ . \quad (5)$$

Plainly, the first two terms on the RHS of (5) give the probability that the system is one event (either a birth or a death) removed from the state $(n,c)'$ at time t and then undergoes such an event in $(t, t+\Delta t]$. The last quantity in (5) is the probability that there are no events in $(t, t+\Delta t]$. Hence, equation (5) leads to the forward Kolmogorov equation

$$\frac{dp_{n,c}(t)}{dt} = \lambda(n-1)p_{n-1,c-1}(t) + \mu c(n+1)p_{n+1,c}(t) - n(\lambda+\mu c)p_{n,c}(t), \quad (6)$$

which is also known in the context of stochastic chemical kinetic models as the "Chemical Master Equation". If equation (6) could be solved to obtain $p_{n,c}(t)$ then Bayesian inference would be straightforward. Even in the absence of a solution to (6), Boys et al. (2008) show that it is possible to directly estimate the parameters in Markov jump processes using Markov chain Monte Carlo algorithms. However, the methods proposed are not suitable for any reaction system of reasonable size (say, more than three reactions with two species) or when there is a large amount of data. The dataset we describe in Section 2.3 involves twenty-seven treatment-block combinations, giving rise to a model with fifty-four reactions and species. Therefore, it seems unlikely that we can carry out Bayesian inference based on the exact underlying model. Instead, working with an approximate model but making exact Bayesian inferences, appears to be a promising approach. Developments on approximating the underlying model in the recent literature have focused on the diffusion approximation (see for example Golightly and Wilkinson, 2005, 2006; Heron et al., 2007; Purutcuoglu and Wit, 2007). The diffusion approximation of the stochastic kinetic model considered here is obtained by calculating the infinitesimal mean and second moment of the process defined by (3)–(4) and matching these to the drift and diffusion coefficient of an Itô stochastic differential equation. We obtain

$$\begin{cases} dN(t) = (\lambda N(t) - \mu N(t)C(t))\, dt + \sqrt{\lambda N(t)}\, dW_1(t) - \sqrt{\mu N(t)C(t)}\, dW_2(t) \\ dC(t) = \lambda N(t)dt + \sqrt{\lambda N(t)}\, dW_1(t) \end{cases} \quad (7)$$

where $W_1(t)$ and $W_2(t)$ are uncorrelated Brownian motions. Whilst it is possible to base inference algorithms around the diffusion approximation, absence of an analytic solution precludes closed form expressions for the associated transition densities. Consequently, Bayesian inference is not straightforward. The approach of Eraker (2001) (see also Golightly and Wilkinson, 2005) involves augmenting observations with latent data to allow a sufficiently accurate Euler-Maruyama approximation of the underlying but unavailable transition densities. Whilst this approach has been shown to work well, computational cost scales almost linearly with the number of latent values employed. Therefore, we adopt a different approximation. As with the diffusion approximation, we also ignore the discreteness associated with the Markov jump process but approximate transition densities as Normal, with mean and variance determined by moment closure methods. Bayesian inference can then be performed using these approximate transition densities. Since the moment closure approximation does not require the need for imputing latent values, the technique is computationally efficient. The method is outlined in Section 2.2.

2.2. Moment Closure Approximation

For the forward Kolmogorov equation in (6), it is not possible to derive an analytic solution. One common method of approximating the first few moments of the process is to examine the moment equations of the system (see for example Krishnarajah et al., 2005; Singh and Hespanha, 2007; Gillespie, 2009). We first define the bivariate moment generating function as

$$M(\theta,\phi;t) \equiv \sum_{n,c=0}^{\infty} e^{n\theta}\, e^{c\phi} p_{n,c}(t)$$

and the associated cumulant generating function as

$$K(\theta,\phi;t) \equiv \log[M(\theta,\phi;t)] = \sum_{n,c=0}^{\infty} \frac{\theta^n}{n!}\frac{\phi^c}{c!}\kappa_{nc}\,.$$

When dealing with only the first few moments, cumulants are a particularly useful representation since κ_{10} and κ_{01} are the marginal means of the $n(t)$ and $c(t)$ and $\{\kappa_{20},\kappa_{02},\kappa_{11}\}$ are the marginal variances and covariance, respectively.

On multiplying equation (6) by $e^{n\theta}e^{c\phi}$, summing over $\{n,c\}$ and converting to the cumulant generating function, we obtain the partial differential equation

$$\frac{\partial K}{\partial t} = \lambda(e^{\theta+\phi}-1)\frac{\partial K}{\partial \theta} + \mu(e^{-\theta}-1)\left(\frac{\partial^2 K}{\partial\theta\partial\phi} + \frac{\partial K}{\partial\theta}\frac{\partial K}{\partial\phi}\right) \tag{8}$$

after dropping the notational dependence of $K(\theta,\phi;t)$ on θ, ϕ and t. A complete derivation of equation (8) can be found in Appendix A. An ordinary differential equation for the mean population can now be obtained by differentiating (8) with respect to θ and setting $\theta=\phi=0$. Likewise, to obtain an expression for the covariance we differentiate (8) with

respect to θ and ϕ and set $\theta = \phi = 0$. Thus we obtain the following cumulant equations

$$\begin{aligned}
\frac{d\kappa_{10}}{dt} &= \lambda\kappa_{10} - \mu(\kappa_{10}\kappa_{01} + \kappa_{11}) \\
\frac{d\kappa_{01}}{dt} &= \lambda\kappa_{10} \\
\frac{d\kappa_{20}}{dt} &= \lambda(\kappa_{10} + 2\kappa_{20}) + \mu(\kappa_{11} - 2\kappa_{10}\kappa_{11} - 2\kappa_{21} + \kappa_{01}(\kappa_{10} - 2\kappa_{20})) \\
\frac{d\kappa_{11}}{dt} &= \lambda(\kappa_{10} + \kappa_{20} + \kappa_{11}) - \mu(\kappa_{10}\kappa_{02} + \kappa_{01}\kappa_{11} + \kappa_{12}) \\
\frac{d\kappa_{02}}{dt} &= \lambda(\kappa_{10} + 2\kappa_{11}) \,. \qquad (9)
\end{aligned}$$

Since the forward Kolmogorov equation contains a non-linear rate μnc, the i^{th} equation depends on the $(i+1)^{\text{th}}$ equation. To circumvent this dependency problem we assume an underlying bivariate Normal distribution, that is we close higher-order moments. This implies that

$$\kappa_{12}(t) = \kappa_{21}(t) = 0$$

in equations (9). Matis et al. (2006) extensively explores the validity of assuming an underlying Normal distributions for this model, and concludes that the approximation appears to be suitable. In particular, they conclude

(a) The mean values $\kappa_{10}(t)$ and $\kappa_{01}(t)$ are well approximated and the expected final cumulative count $\kappa_{01}(\infty)$ has an error rate of less than 2.5%
(b) The approximation for the variance is also very accurate, with an error rate of less than 2.5%.
(c) The skewness for the population and cumulative number of aphids can be substantive, but the limiting distribution is approximated well by a Normal distribution. It is worth noting that since we are using a truncated Normal distribution, then our approximating distribution will have non-zero third cumulant.

Overall, the moment closure approximation provides a computationally efficient method for describing aphid population and final cumulative count distributions (Matis et al., 2006).

2.3. Description of the Data

In this paper we consider the data described in Matis et al. (2008) consisting of five observations on cotton aphid counts on twenty randomly chosen leaves in each plot, for twenty-seven treatment-block combinations. The data were recorded in July 2004 in Lamesa, Texas. The treatments consisted of three nitrogen levels (blanket, variable and none), three irrigation levels (low, medium and high) and three blocks, each being a distinct area. Irrigation treatments were randomly assigned within each block as whole plots. Nitrogen treatments were randomly assigned within each whole block as split plots. The data are plotted in Figure 1. Note that the sampling times are t=0, 1.14, 2.29, 3.57 and 4.57 weeks (i.e. every 7 to 8 days).

Let i, j, k denote the respective level of water, nitrogen and block, with $i, j, k \in \{1, 2, 3\}$ where for simplicity 1 denotes low water/blanket nitrogen, 2 denotes medium water/variable nitrogen and 3 denotes high water/zero nitrogen. Now denote the time series obtained for combination ijk as $\mathcal{D}_{ijk}$, the number of aphids for combination ijk at time t by $N_{ijk}(t)$ and

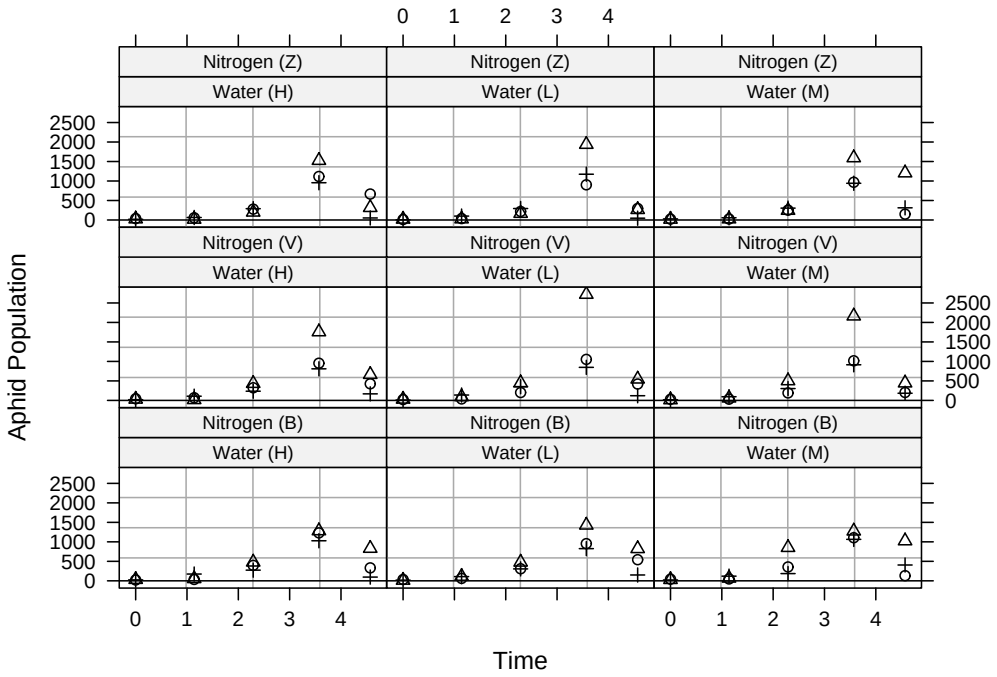

Fig. 1. Aphid counts against time obtained for 27 different scenarios – three levels of water (low, medium and high) and nitrogen (blanket, variable and zero) and three blocks (with the data from block 1 given by a circle, block 2 by a triangle and block 3 by a cross).

the cumulative population size at time t by $C_{ijk}(t)$. We then model each dataset D_{ijk} with a stochastic kinetic model of the form

$$N_{ijk} \xrightarrow{\lambda_{ijk}} 2N_{ijk} + C_{ijk}$$
$$N_{ijk} + C_{ijk} \xrightarrow{\mu_{ijk}} C_{ijk}$$

and impose the following nested structure on the birth and death rates

$$\lambda_{ijk} = \lambda + \alpha_i^l + \beta_j^l + (\alpha\beta)_{ij}^l + \rho_k^l + (\alpha\rho)_{ik}^l + (\beta\rho)_{jk}^l \tag{10}$$
$$\mu_{ijk} = \mu + \alpha_i^m + \beta_j^m + (\alpha\beta)_{ij}^m + \rho_k^m + (\alpha\rho)_{ik}^m + (\beta\rho)_{jk}^m \tag{11}$$

subject to the constraints

$$\alpha_1^l = \alpha_1^m = \beta_1^l = \beta_1^m = \rho_1^l = \rho_1^m = 0$$

and

$$(\alpha\beta)_{11}^l = (\alpha\beta)_{11}^m = 0, \qquad (\alpha\rho)_{11}^l = (\alpha\rho)_{11}^m = 0, \qquad (\beta\rho)_{11}^l = (\beta\rho)_{11}^m = 0\ .$$

Thus, inference from these data amounts to making statements about thirty-eight parameters. The interpretation of equations (10) and (11) is clear — all $(5 \times 3^3 =)135$ data points are used to learn about the baseline birth and death rates λ and μ. The $(5 \times 3^2 =)45$ observations recorded for the medium water scenario inform us about the effect of medium

water relative to baseline via α_2^l, α_2^m and so on. Hence, we allow the overall birth and death rates to vary with each treatment combination but assume underlying (baseline) birth and death rates of λ and μ for all combinations. Naturally, the structure in equations (10) and (11) allows us to explore possible interactions between treatments. The analysis of this dataset is given in Section 4. In the next section we formulate the inference problem for the stochastic growth model considered here and outline an MCMC approach.

3. Inference

For simplicity, consider first the task of inferring the birth and death rates λ and μ in the model given by equations (1)–(4) from a single time series. Let $\mathbf{X}(t_u) = (N(t_u), C(t_u))'$ be the random 2-vector of observed aphid counts and unobserved cumulative population size at time t_u. Let $\boldsymbol{\eta} = (\lambda, \mu)'$ be the 2-vector of parameters. Then, using the moment closure approximation and assuming a Normal distribution, we have

$$\mathbf{X}(t_u) \,|\, \mathbf{X}(t_{u-1}), \boldsymbol{\eta} \sim N\left(\boldsymbol{\psi}_{u-1}\,,\, \boldsymbol{\Sigma}_{u-1}\right) \tag{12}$$

where $\boldsymbol{\psi}_{u-1}$ and $\boldsymbol{\Sigma}_{u-1}$ are calculated via the moment closure approximation outlined in Section 2.2 and may depend on the parameters $\boldsymbol{\eta}$ and state $\mathbf{X}(t_{u-1})$ in a nonlinear way. Note that $N(\boldsymbol{\psi}\,,\, \boldsymbol{\Sigma})$ denotes the multivariate Normal distribution with mean vector $\boldsymbol{\psi}$ and covariance matrix $\boldsymbol{\Sigma}$. Now suppose that observations on aphid counts $\mathbf{n} = \{n(t_u) : u = 0, \ldots, T\}$ are available. Summarising our *a priori* beliefs about $\boldsymbol{\eta}$ and the initial unobserved cumulative population size $C(t_0)$ via $p(\boldsymbol{\eta})$ and $p\,(c(t_0))$, we obtain the joint posterior distribution for parameters and unobserved states as

$$p\,(\boldsymbol{\eta}, \mathbf{c} \,|\, \mathbf{n}) \propto p(\boldsymbol{\eta})\, p\,(c(t_0)) \prod_{u=1}^{T} p\,(\mathbf{x}(t_u) \,|\, \mathbf{x}(t_{u-1}), \boldsymbol{\eta}) \tag{13}$$

where $p\,(\mathbf{x}(t_u) \,|\, \mathbf{x}(t_{u-1}), \boldsymbol{\eta})$ is the transition density associated with (12) and $\mathbf{c} = \{c(t_u) : u = 0, \ldots, T\}$ denotes the vector of unobserved states.

Since interest lies largely in the posterior distribution of parameters given aphid counts, $p(\boldsymbol{\eta} \,|\, \mathbf{n})$, we integrate over our uncertainty for the unobserved states via MCMC. That is, we sample the joint posterior density in equation (13) via a Gibbs sampler in which we alternate between draws of $\boldsymbol{\eta}$ conditional on $\mathbf{x} = \{\mathbf{x}(t_u) : u = 0, \ldots, T\}$ and draws $\mathbf{c}$ conditional on the current $\boldsymbol{\eta}$ and $\mathbf{n}$. Retaining only those sampled values of $\boldsymbol{\eta}$ gives a sample from the target distribution $p(\boldsymbol{\eta} \,|\, \mathbf{n})$ (Tanner and Wong, 1987). Since the required full conditionals typically preclude analytic tractability, we use a Metropolis-Hastings step where necessary.

Extension of this framework to the case of multiple datasets as in Section 2.3 is straightforward. Recall that we have twenty-seven datasets $\mathcal{D}_{ijk}$, for $i, j, k \in \{1, 2, 3\}$, one for each water, nitrogen and block combination. Let $\mathbf{x}_{ijk}(t_u) = (n_{ijk}(t_u), c_{ijk}(t_u))'$ be the vector of aphid counts and cumulative population size at time t_u in dataset $\mathcal{D}_{ijk}$. Define $\mathbf{n}_{ijk} = \{n_{ijk}(t_u) : u = 0, \ldots, T\}$ and $\mathbf{c}_{ijk} = \{c_{ijk}(t_u) : u = 0, \ldots, T\}$ to be the respective vectors of aphid counts and cumulative population sizes in dataset $\mathcal{D}_{ijk}$. Redefine $\mathbf{n} = \{\mathbf{n}_{ijk} : i, j, k \in \{1, 2, 3\}\}$ and $\mathbf{c} = \{\mathbf{c}_{ijk} : i, j, k \in \{1, 2, 3\}\}$. Finally, let η be the thirty-six vector of parameters given in equations (10) and (11). We then obtain the joint posterior for parameters and latent data as

$$p\,(\boldsymbol{\eta}, \mathbf{c} \,|\, \mathbf{n}) \propto p(\boldsymbol{\eta})\, p\,(\mathbf{c}(t_0)) \prod_{i=1}^{3} \prod_{j=1}^{3} \prod_{k=1}^{3} \prod_{u=1}^{4} p\,(\mathbf{x}_{ijk}(t_u) \,|\, \mathbf{x}_{ijk}(t_{u-1}), \boldsymbol{\eta}) \tag{14}$$

where $\mathbf{c}(t_0)$ denotes the vector of initial cumulative population sizes for each dataset. Note that we take independent prior distributions for each $C_{ijk}(t_0)$. Since the cumulative population size should be at least as large as the initial observed population size $n_{ijk}(t_0)$, and we do not expect $C_{ijk}(t_0)$ to differ hugely from $n_{ijk}(t_0)$, we specify our prior beliefs via $C_{ijk}(t_0) = n_{ijk}(t_0) + \epsilon$ where ϵ follows a Gamma distribution with shape 1 and scale 0.2. As for the simpler case of a single dataset, sampling of (14) may proceed by alternating between parameter draws and latent state draws. In what follows, we describe in detail the parameter and state updates. For clarity and notational simplicity, we give details of the MCMC scheme in the context of a single time series. To validate the methodology, we apply the scheme to synthetic data comprising six time-series (arising from three treatments and two blocks) in Section 4.1. We analyse the aphid data, for all twenty-seven treatment-block combinations in Section 4.2.

Parameter updates

Since the birth and death rates λ and μ must be strictly positive, we sample their full conditional distribution via a random walk Metropolis algorithm (O'Hagan and Forster, 2004) on the log scale. To aid the mixing of the resulting Markov chain, we sample λ and μ in a single block. Let $\log(\boldsymbol{\eta}) = (\log(\lambda), \log(\mu))'$ and we assume independent proper uniform $U(-10, 10)$ priors for $\log(\lambda)$ and $\log(\mu)$. If at some iteration of the MCMC algorithm, the current parameter values are λ and μ, we propose λ_* and μ_* jointly, where

$$\begin{aligned} \log(\lambda_*) &= \log(\lambda) + \epsilon_1\,, \qquad \epsilon_1 \sim N(0, \omega_1^2) \\ \log(\mu_*) &= \log(\mu) + \epsilon_2\,, \qquad \epsilon_2 \sim N(0, \omega_2^2) \end{aligned}$$

and ϵ_1 is independent of ϵ_2. Note that ω_1 and ω_2 are tuning parameters whose values determine the mixing properties of the chain. The values of ω_1 and ω_2 are chosen to achieve an optimal acceptance probability of around 0.23 (Roberts and Rosenthal, 2001).

A move to $\log(\boldsymbol{\eta}_*)$ is accepted with probability

$$\min\left\{1, \frac{\displaystyle\prod_{u=1}^{T} p\left(\mathbf{x}(t_u) \,|\, \mathbf{x}(t_{u-1}), \boldsymbol{\eta}_*\right)}{\displaystyle\prod_{u=1}^{T} p\left(\mathbf{x}(t_u) \,|\, \mathbf{x}(t_{u-1}), \boldsymbol{\eta}\right)}\right\}$$

provided that the proposed values are consistent with the prior. Note that for the multiple aphid datasets described in Section 2.3 and analysed in Section 4.2, only the baseline parameters need be strictly positive. We therefore update the remaining parameters using a Metropolis random walk update without transforming to the log-scale.

Latent process updates

Since we only have observations on the population level, that is, $N(t)$ and not $C(t)$, we have to sample the latent process at every iteration of the MCMC algorithm. Here we adopt the simplest strategy which updates the latent process one value at a time by sampling $C(t_u)$ from its full conditional distribution for $u = 0, \ldots, T$. Using the joint posterior in

Table 1. Parameters used in the simulation study. The underlying (baseline) birth and death are $\{\lambda = 1.75, \mu = 0.00095\}$. Treatment 2 increases the death rate by 0.0004 (but leaves the birth rate unchanged) and Treatment 3 increases the birth rate by 0.35 (but leaves the death rate unchanged). The block effect reduces the death rate by 0.0003 (but leaves the birth rate unchanged).

	Treatment 1	*Treatment 2*	*Treatment 3*
Block 1	$\{1.75, 0.00095\}$	$\{1.75, 0.00135\}$	$\{2.1, 0.00095\}$
Block 2	$\{1.75, 0.00065\}$	$\{1.75, 0.00105\}$	$\{2.1, 0.00065\}$

equation (13), we have

$$p\left(c(t_u) \,|\, \mathbf{x}(t_{u-1}), \mathbf{x}(t_{u+1}), n(t_u), \boldsymbol{\eta}\right) \propto p\left(\mathbf{x}(t_{u+1}) \,|\, \mathbf{x}(t_u), \boldsymbol{\eta}\right) p\left(\mathbf{x}(t_u) \,|\, \mathbf{x}(t_{u-1}), \boldsymbol{\eta}\right)$$
$$= N\left(\mathbf{x}(t_{u+1}) \,;\, \boldsymbol{\psi}_u \,,\, \boldsymbol{\Sigma}_u\right) N\left(\mathbf{x}(t_u) \,;\, \boldsymbol{\psi}_{u-1} \,,\, \boldsymbol{\Sigma}_{u-1}\right) \quad (15)$$

where $N(\cdot \,;\, \boldsymbol{\psi} \,,\, \boldsymbol{\Sigma})$ denotes the probability density function of a Normal random variable with mean $\boldsymbol{\psi}$ and covariance matrix $\boldsymbol{\Sigma}$. Since $\boldsymbol{\psi}_u$ and $\boldsymbol{\Sigma}_u$ typically depend on $C(t_u)$ in a nonlinear way, the form of (15) precludes analytic tractability. We therefore use a Metropolis-Hastings step to sample the density in (15). To construct a suitable proposal distribution we note that the data are equispaced. Hence, smoothing between the neighbours of $N(t_u)$ we arrive at

$$q\left(\mathbf{x}(t_u) \,|\, \mathbf{x}(t_{u-1}), \mathbf{x}(t_{u+1}), \boldsymbol{\eta}\right) = N\left(\mathbf{x}(t_u) \,;\, \frac{1}{2}\left(\mathbf{x}(t_{u-1}) + \mathbf{x}(t_{u+1})\right) \,,\, \frac{1}{2}\boldsymbol{\Sigma}_{\mathbf{u-1}}\right)$$

and condition this density on $N(t_u) = n(t_u)$ using standard multivariate Normal conditioning arguments. We denote the resulting density by $q\left(c(t_u) \,|\, \mathbf{x}(t_{u-1}), \mathbf{x}(t_{u+1}), n(t_u), \boldsymbol{\eta}\right)$ and use this as a proposal density inside a Metropolis-Hastings step. If the current value of the chain is $c(t_u)$, we accept a move to $c_*(t_u) \sim q\left(\cdot \,|\, \mathbf{x}(t_{u-1}), \mathbf{x}(t_{u+1}), n(t_u), \boldsymbol{\eta}\right)$ with probability

$$\min\left\{1, \frac{p\left(\mathbf{x}(t_{u+1}) \,|\, \mathbf{x}_*(t_u), \boldsymbol{\eta}\right) p\left(\mathbf{x}_*(t_u) \,|\, \mathbf{x}(t_{u-1}), \boldsymbol{\eta}\right)}{p\left(\mathbf{x}(t_{u+1}) \,|\, \mathbf{x}(t_u), \boldsymbol{\eta}\right) p\left(\mathbf{x}(t_u) \,|\, \mathbf{x}(t_{u-1}), \boldsymbol{\eta}\right)} \times \frac{q\left(c(t_u) \,|\, \mathbf{x}(t_{u-1}), \mathbf{x}(t_{u+1}), n(t_u), \boldsymbol{\eta}\right)}{q\left(c_*(t_u) \,|\, \mathbf{x}(t_{u-1}), \mathbf{x}(t_{u+1}), n(t_u), \boldsymbol{\eta}\right)}\right\}$$

where $\mathbf{x}_*(t_u) = (n(t_u), c_*(t_u))'$. We repeat this step for each latent value in turn, thereby updating the whole latent path. The form of $q\left(\cdot \,|\, \mathbf{x}(t_{u-1}), \mathbf{x}(t_{u+1}), n(t_u), \boldsymbol{\eta}\right)$ is given in Appendix B. Other updating schemes are possible. For example we find that a random walk also update works well, but this scheme requires the specification of additional tuning parameters.

4. Results

4.1. Simulation Study

To examine the performance of our inference algorithm, we consider a simple example in which we have three treatments and two blocks. Let i, j represent the block and treatment level, $i \in \{1, 2\}$, $j \in \{1, 2, 3\}$. Denote the synthetic dataset for each block-treatment combination by $\mathcal{D}_{ij}$. For each dataset we assume birth and death rates of the form

$$\lambda_{ij} = \lambda + \alpha_i^l + \beta_j^l \quad \text{and} \quad \mu_{ij} = \mu + \alpha_i^m + \beta_j^m$$

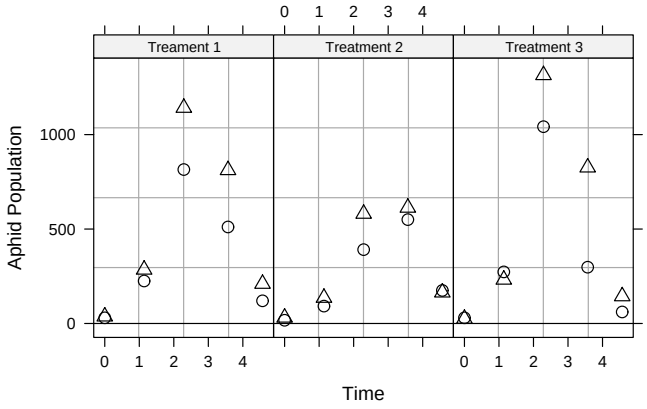

Fig. 2. Simulated data sets based on the parameter values in Table 1. The circles and triangle correspond to Blocks 1 and 2 respectively.

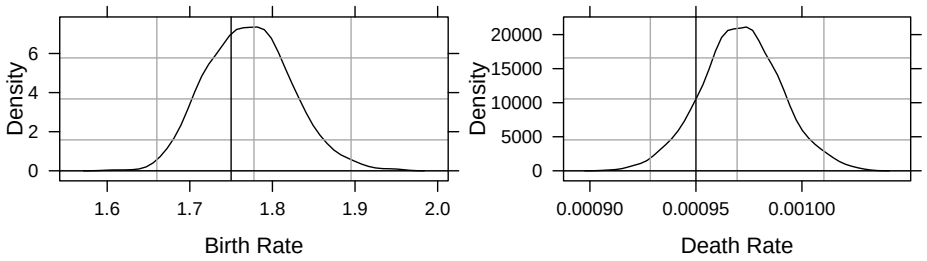

Fig. 3. Kernel density estimates of the marginal posterior distributions of baseline birth and death rates (λ and μ) for the synthetic data shown in Figure 2. The true values are the black vertical lines.

with $\alpha_1^l = \alpha_1^m = \beta_1^l = \beta_1^m = 0$. Note for simplicity that we assume no interaction between treatment and block. The six datasets were simulated from the underlying exact discrete stochastic model (and not our approximate inferential model) using the Gillespie algorithm (Gillespie, 1977) by taking the values of λ_{ij} and μ_{ij} given in Table 1. Note in particular that $\lambda = 1.75$, $\alpha_2^l = 0$, $\beta_2^l = 0$, $\beta_3^l = 0.35$, $\mu = 0.00095$, $\alpha_2^m = -0.0003$, $\beta_2^m = 0.0004$ and $\beta_3^m = 0$.

To mirror the real dataset in Section 2.3, the initial population sizes were chosen from a Poisson distribution with mean 28 and data points were recorded at times $0.0, 1.14, 2.29, 3.57$ and 4.57. The resulting dataset is shown in Figure 2.

We ran the MCMC algorithm described in Section 3 for two million iterations and thinned by a factor of 1000 iterations. This yielded a sample of 2000 iterates with low correlation to be used as the main monitoring run. The posterior densities for the baseline birth and death rates (λ and μ) are shown in Figure 3. Plainly, the sampler gives values of the birth and death rates that are consistent with the true values. In other simulation studies (not reported here), in general we noticed that the death rate is harder to recover, and we believe that this is because the parameter is strongly linked with the latent process, that is, the reaction hazard for aphid death (with rate constant μ) depends on the unobserved cumulative population size.

The posterior distributions for the treatment and block effects for birth and death are

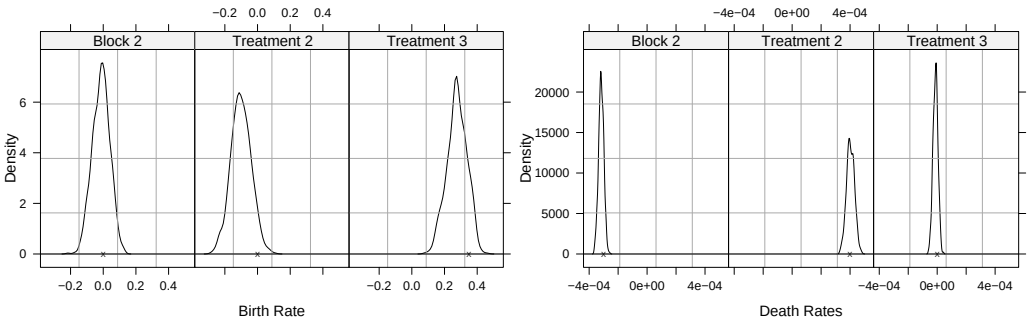

Fig. 4. Kernel density estimates of the marginal posterior distributions of treatment and block birth and death rates for the synthetic data shown in Figure 2. Left panel: α_2^l, β_2^l and β_3^l respectively. Right panel: α_2^m, β_2^m and β_3^m respectively. An X corresponds to the true value.

shown in Figure 4. As with the baseline parameters shown in Figure 3, samples of the birth and death rates are clearly consistent with the true values.

4.2. Application to the Aphid Data Set

The complete aphid dataset described in Section 2.3 comprises twenty-seven treatment combinations and involves thirty-eight parameters. We believe that this system is the largest stochastic kinetic model (in terms of numbers of reactions and parameters) that has been fitted to real data. The inference scheme of Section 3 was run for 2 million iterations with a thin of 1000. This produced (near) uncorrelated samples from the parameter posterior distributions. Marginal posterior distributions for the baseline birth and death rates are shown in Figure 5. In Table 2 we give 95% credible intervals for all parameters. Rather surprisingly, some of the strongest effects present are in the Block 2 interaction terms. Further inspection of Table 2 shows that nitrogen plays an important part in the aphid population dynamics. As levels of nitrogen were decreased (from blanket to variable to zero), both the birth rate and death rates increased. For example, 95% credible intervals for β_2^l and β_3^l (corresponding to the change in birth rate from baseline for variable and zero nitrogen respectively) are $(0.0553, 0.3220)$ and $(0.236, 0.529)$. This result is in agreement with Matis et al. (2008).

Since the purpose of this study was to determine how aphid populations are affected by the different treatments, simply using Table 2 is insufficient. Indeed, it is difficult to explicitly describe the effect of a particular treatment simply through inspection of Table 2, since, for example, the birth *and* the death rate can increase (as with the medium water effect) and this may have little effect on the total number of aphids. We therefore consider the effect of different treatment combinations by examining the distributions of the total number of aphids at future time points in Section 4.2.1.

The sensitivity of the posterior distribution to the choice of prior distribution was assessed by using repeated runs of the inference scheme each with small changes in the prior specification. We found that such changes resulted in only small changes to the posterior distributions.

Naturally, the validity of any conclusions that can be drawn from the modelling process can be questioned if the model does not fit the data sufficiently well. To test model fit, that is, to see if the model is consistent with the observed data, we compared the predictive

Table 2. Posterior 95% credible intervals for the parameters in equations (10) and (11), obtained using the aphid dataset.

Effect	Birth Rates		Death Rates	
Baseline	λ:	(1.64, 1.86)	μ:	(0.000904, 0.000987)
Water (M)	α_2^l:	(0.501, 0.790)	α_2^m:	(5.73×10^{-5}, 1.64×10^{-4})
Water (H)	α_3^l:	(0.0888, 0.3590)	α_3^m:	(-0.000204, -0.000111)
Nitrogen (V)	β_2^l:	(0.0553, 0.3220)	β_2^m:	(1.84×10^{-5}, 1.14×10^{-4})
Nitrogen (Z)	β_3^l:	(0.236, 0.529)	β_3^m:	(0.000109, 0.000208)
Block 2	ρ_2^l:	(-0.2420, 0.0112)	ρ_2^m:	(-0.000411, -0.000326)
Block 3	ρ_3^l:	(-0.169, 0.124)	ρ_3^m:	(0.000127, 0.000254)
Water (M) Nitrogen (V)	$(\alpha\beta)_{22}^l$:	(-0.3410, -0.0614)	$(\alpha\beta)_{22}^m$:	(0.000110, 0.000185)
Water (M) Nitrogen (Z)	$(\alpha\beta)_{23}^l$:	(-0.818, -0.525)	$(\alpha\beta)_{23}^m$:	(-1.43×10^{-4}, -5.53×10^{-5})
Water (H) Nitrogen (V)	$(\alpha\beta)_{32}^l$:	(-0.538, -0.266)	$(\alpha\beta)_{32}^m$:	(0.000021, 0.000102)
Water (H) Nitrogen (Z)	$(\alpha\beta)_{33}^l$:	(-0.550, -0.253)	$(\alpha\beta)_{33}^m$:	(-9.21×10^{-5}, 3.51×10^{-6})
Water (M) Block 2	$(\alpha\rho)_{22}^l$:	(-0.697, -0.395)	$(\alpha\rho)_{22}^m$:	(-0.000247, -0.000144)
Water (M) Block 3	$(\alpha\rho)_{23}^l$:	(-0.613, -0.278)	$(\alpha\rho)_{23}^m$:	(-0.000275, -0.000134)
Water (H) Block 2	$(\alpha\rho)_{32}^l$:	(-0.2140, 0.0824)	$(\alpha\rho)_{32}^m$:	(0.000183, 0.000276)
Water (H) Block 3	$(\alpha\rho)_{33}^l$:	(-0.0784, 0.2430)	$(\alpha\rho)_{33}^m$:	(0.000166, 0.000306)
Nitrogen (V) Block 2	$(\beta\rho)_{22}^l$:	(0.497, 0.781)	$(\beta\rho)_{22}^m$:	(-0.000213, -0.000121)
Nitrogen (V) Block 3	$(\beta\rho)_{23}^l$:	(-0.111, 0.187)	$(\beta\rho)_{23}^m$:	(-1.11×10^{-4}, 2.52×10^{-5})
Nitrogen (Z) Block 2	$(\beta\rho)_{32}^l$:	(0.587, 0.884)	$(\beta\rho)_{32}^m$:	(-1.72×10^{-6}, 9.19×10^{-5})
Nitrogen (Z) Block 3	$(\beta\rho)_{33}^l$:	(0.0907, 0.4030)	$(\beta\rho)_{33}^m$:	(-1.78×10^{-4}, -4.75×10^{-5})

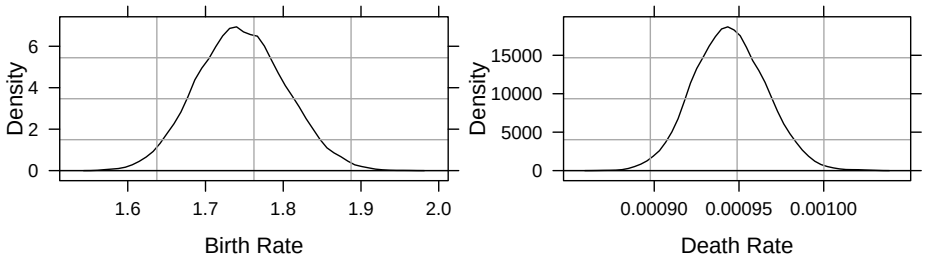

Fig. 5. Kernel density estimates of the marginal posterior distributions of baseline birth and death rates (λ and μ) for the aphid data shown in Figure 1.

distributions for aphid population size $N(t)$ with the actual observations. The predictive distribution is obtained by combining the model at a particular time point with the parameter posterior and integrating over the uncertainty associated with the parameter value. In other words, the predictive distribution is taken to be the posterior average of realisations of the population growth process, where the average is taken over the posterior distribution. Therefore, obtaining predictive distributions is straightforward, as for each randomly sampled parameter value $(\lambda_{ijk}, \mu_{ijk})$ and the unobserved state $C_{ijk}(t_0)$ from the MCMC output, we can forward simulate at each observation time from the underlying discrete stochastic model using the Gillespie algorithm. Figure 6 shows box and whisker plots of the predictive distributions of aphid population size $N(t)$ (for a variety of treatment scenarios) at times 1.14, 2.29, 3.57 and 4.57 weeks, plotted against the actual data. Inspection of Figure 6 suggests that the model captures the underlying characteristics of the data sufficiently well. Further improvements to the model are discussed in Section 5.

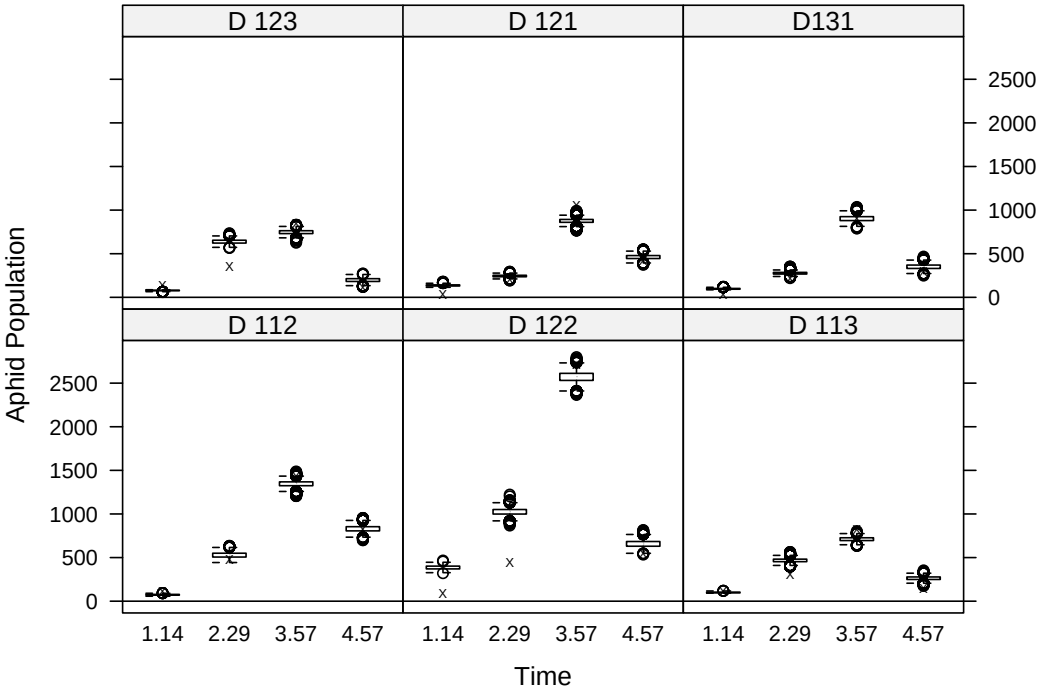

Fig. 6. Predictive distributions for aphid population sizes (represented by Box and Whisker plots) and observations for 6 aphid datasets, corresponding to 6 treatment combinations. The data points are represented by an X.

4.2.1. Additional number of Aphids by treatment

Recall that $C_{ijk}(t_n)$ denotes the cumulative population size at time t_n in dataset $\mathcal{D}_{ijk}$. Hence, for some future time point t_n and any treatment combination ijk, we can sample the predictive distribution of $C_{ijk}(t_n)$ by averaging over each randomly sampled value of λ_{ijk}, μ_{ijk} and $C_{ijk}(t_0)$ (from the MCMC scheme) with fixed $N_{ijk}(t_0) = n_{ijk}(t_0)$ and forward simulating via the Gillespie algorithm to time t_n. Since we wish to compare the effects of each treatment and their interactions, we fix $t_0 = 0$, $C_{ijk}(0) = 1$, $N_{ijk}(0) = 1$ and $t_n = 6$ weeks. Hence, for each treatment effect and interaction, we can compute the distributions of cumulative population numbers at time 6 *in addition* to the numbers we would see at baseline. For example, to obtain samples from the distribution of additional aphid numbers due to the effect of medium water, we use each MCMC iterated value of the parameters

$$\lambda_{211} = \lambda + \alpha_2^l \quad \text{and} \quad \mu_{211} = \mu + \alpha_2^m$$

combined with the Gillespie algorithm, to generate values of $C_{211}(6)$. Denote the sampled values of $C_{211}(6)$ by $\{c_{211}^{(i)}(6), i = 1, \ldots, N\}$. We then take

$$c_{211}^{(i)}(6) - c_{111}^{(i)}(6), \qquad i = 1, \ldots, N$$

as a sample from the distribution of additional aphids due to medium water. Similar calculations can be performed to obtain the distributions of additional aphid numbers (over baseline) for all 18 treatment combinations. These distributions are shown in Figure 7.

Inspection of Figure 7 suggests that the variable and zero nitrogen treatments have little effect over baseline with distributions of additional numbers of aphids centred around zero. Similarly, it appears that water has little effect by itself. This is not entirely unexpected since in the year that the data were recorded, there was moderately high rainfall. Plainly, there are strong interactions. In particular, block 2 appears to have a very strong interaction with nitrogen.

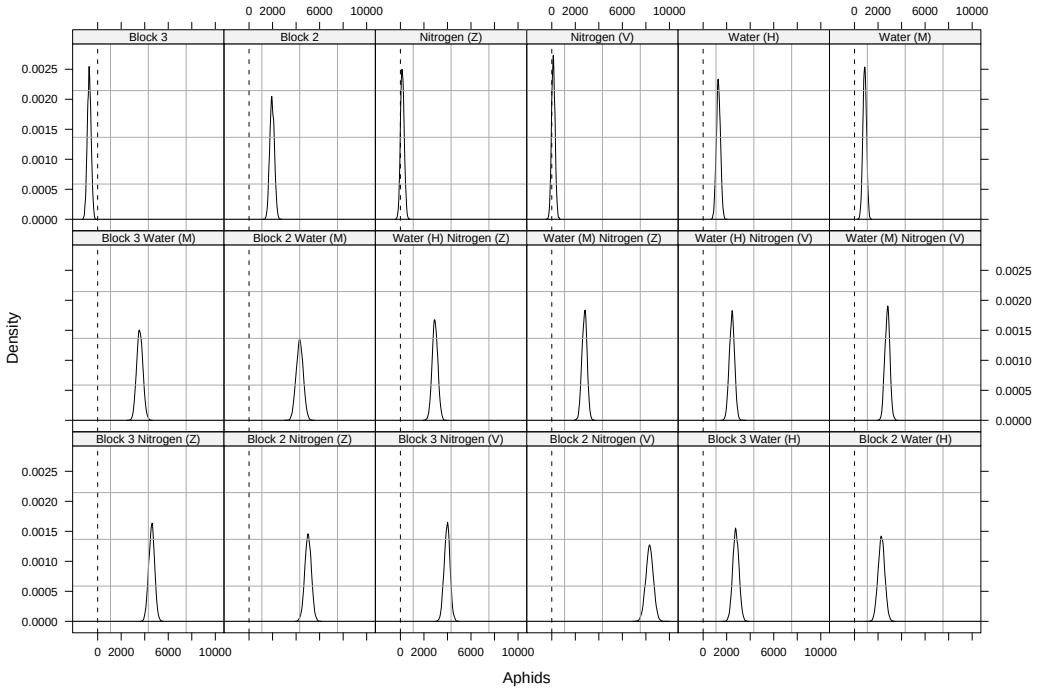

Fig. 7. Distribution of additional number of aphids due to each treatment over baseline at time 6.

5. Conclusion

In this paper we have considered the task of inference for a stochastic population growth model of aphid numbers. Whilst such models are well known, interest has usually focused on their deterministic counterparts (Matis et al. (2006, 2007)). The model considered here is formulated as a Markov jump process with an unobserved component representing cumulative population size. The associated (but unavailable) transition densities were approximated using a moment closure approach and an MCMC scheme used to sample the posterior distribution of parameters and latent data. It is important to note that the methodology considered here could in theory be applied to any Markov jump process (Milner et al., 2009). The model was fitted to the data sets of Matis et al. (2008) and consisted of cotton aphid counts for twenty-seven treatment combinations. To allow the possibility of interactions, and rather than fitting a model to all datasets separately, we used a nested structure for the birth and death parameters (specific to each treatment combination) resulting in thirty-eight unknown parameters to be estimated.

As well as confirming the results of Matis et al. (2008), we also find strong interaction effects, in particular in block 2. These effects were quantified probabilistically by calculating the distribution of additional aphid numbers at time $t = 6$ over baseline. Note that in computing these distributions, we take into account parameter uncertainty by integrating over the unknown parameter vector.

Finally, it is worth noting that although the predictive distributions in Figure 6 suggest adequate model fit, it may be possible to improve the model by adding in an immigration term, resulting in

$$\begin{aligned} \emptyset &\xrightarrow{\gamma} N + C \\ N &\xrightarrow{\lambda} 2N + C \\ N + C &\xrightarrow{\mu} C \ . \end{aligned}$$

This is the subject of ongoing research.

Acknowledgements

We are grateful to the Associate Editor and two referees for their constructive comments and suggestions which led to improvements in the paper. We also wish to thank Prof Richard Boys, Dr Malcolm Farrow and Prof Jim Matis for helpful discussions.

A. Constructing the Moment Equations

Recall that the forward Kolmogorov equation governing the probability $p_{n,c}(t)$ is

$$\frac{dp_{n,c}(t)}{dt} = \lambda(n-1)p_{n-1,c-1}(t) + \mu c(n+1)p_{n+1,c}(t) - n(\lambda + \mu c)p_{n,c}(t), \tag{16}$$

and the bivariate moment generating function is

$$M(\theta, \phi; t) \equiv \sum_{n,c=0}^{\infty} e^{n\theta}\ e^{c\phi} p_{n,c}(t) \ .$$

On multiplying equation (16) by $e^{n\theta}e^{c\phi}$, we obtain

$$\begin{aligned} e^{n\theta}e^{c\phi}\frac{dp_{n,c}(t)}{dt} \quad = \quad & \lambda(n-1)e^{\theta}e^{\phi}e^{(n-1)\theta}e^{(c-1)\phi}p_{n-1,c-1}(t) \\ & +\mu c(n+1)e^{-\theta}e^{(n+1)\theta}e^{c\phi}p_{n+1,c}(t) \\ & -n(\lambda + \mu c)e^{n\theta}e^{c\phi}p_{n,c}(t). \end{aligned}$$

Hence, summing both sides of the preceding equation over $\{n, c\}$ gives

$$\frac{\partial M}{\partial t} = \lambda(e^{\theta+\phi} - 1)\frac{\partial M}{\partial \theta} + \mu(e^{-\theta} - 1)\frac{\partial^2 M}{\partial \theta \partial \phi} \tag{17}$$

where we have dropped the notational dependence of $M(\theta, \phi; t)$ on θ, ϕ and t. Now, we have the cumulant generating function as

$$K(\theta, \phi; t) \equiv \log[M(\theta, \phi; t)] = \sum_{n,c=0}^{\infty} \frac{\theta^n}{n!}\frac{\phi^c}{c!}\kappa_{nc} \ ,$$

and arrive at the partial differential governing $K(\theta, \phi; t)$ by noting that

$$\frac{\partial}{\partial t}\left(e^K\right) = e^K \frac{\partial K}{\partial t}, \quad \frac{\partial}{\partial \theta}\left(e^K\right) = e^K \frac{\partial K}{\partial \theta}, \quad \frac{\partial}{\partial \theta}\left\{\frac{\partial}{\partial \phi}\left(e^K\right)\right\} = e^K \frac{\partial^2 K}{\partial \theta \partial \phi} + e^K \frac{\partial K}{\partial \theta}\frac{\partial K}{\partial \phi}$$

where we have dropped the notational dependence of $K(\theta, \phi; t)$ on θ, ϕ and t. Hence, substituting $M = e^K$ into equation (17) gives

$$\frac{\partial K}{\partial t} = \lambda(e^{\theta+\phi} - 1)\frac{\partial K}{\partial \theta} + \mu(e^{-\theta} - 1)\left(\frac{\partial^2 K}{\partial \theta \partial \phi} + \frac{\partial K}{\partial \theta}\frac{\partial K}{\partial \phi}\right).$$

This equation can then be used to derive ordinary differential equations for the cumulants κ_{10}, κ_{01}, κ_{11}, κ_{02} and κ_{20}.

B. Constructing the Proposal Density for the Latent Data

Recall that $\mathbf{X}(t_u) = (N(t_u), C(t_u))'$ denotes the random 2-vector of population size and cumulative size, and partition $q\left(\mathbf{x}(t_u) \,|\, \mathbf{x}(t_{u-1}), \mathbf{x}(t_{u+1}), \boldsymbol{\eta}\right)$ in the following way,

$$N\left\{\begin{pmatrix} n(t_u) \\ c(t_u) \end{pmatrix}; \frac{1}{2}\begin{pmatrix} n(t_{u-1}) + n(t_{u+1}) \\ c(t_{u-1}) + c(t_{u+1}) \end{pmatrix}, \frac{1}{2}\begin{pmatrix} \sigma^{nn}_{u-1} & \sigma^{nc}_{u-1} \\ \sigma^{cn}_{u-1} & \sigma^{cc}_{u-1} \end{pmatrix}\right\}.$$

Then, conditioning on $N(t_u) = n(t_u)$ gives the proposal density as

$$q\left(c(t_u) \,|\, \mathbf{x}(t_{u-1}), \mathbf{x}(t_{u+1}), n(t_u), \boldsymbol{\eta}\right) = N\left(c(t_u); \psi, \Sigma\right)$$

where

$$\psi = \frac{1}{2}\left(c(t_{u-1}) + c(t_{u+1})\right) + \sigma^{cn}_{u-1}\left(\sigma^{nn}_{u-1}\right)^{-1}\left(n(t_u) - \frac{1}{2}\left[n(t_{u-1}) + n(t_{u+1})\right]\right)$$
$$\Sigma = \frac{1}{2}\sigma^{cc}_{u-1} - \frac{1}{2}\sigma^{cn}_{u-1}\left(\sigma^{nn}_{u-1}\right)^{-1}\sigma^{nc}_{u-1}.$$

References

Boys, R. J., D. J. Wilkinson, and T. B. L. Kirkwood (2008). Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing 18*, 125–135.

Celini, L. and J. Vaillant (2004). A model of temporal distribution of Aphis gossypii glover (homoptera: Aphididae) on cotton. *Journal of Applied Entomology 128*, 133–139.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics 19*, 177–191.

Giarola, L. T. P., S. G. F. Martins, and M. C. P. Toled Costa (2006). Computer simulation of Aphis gossypii insects using penna aging model. *Physica A: Statistical Mechanics and its Applications 368*, 147–154.

Gillespie, C. S. (2009). Moment-closure approximations for mass-action models. *IET Systems Biology, 3*, 52–58.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry 81*, 2340–2361.

Golightly, A. and D. J. Wilkinson (2005). Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics 61*(3), 781–788.

Golightly, A. and D. J. Wilkinson (2006). Bayesian sequential inference for stochastic kinetic biochemical network models. *Journal of Computational Biology 13*(3), 838–851.

Golightly, A. and D. J. Wilkinson (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis 52*, 1674–1693.

Heron, E. A., B. Finkenstadt, and D. A. Rand (2007). Bayesian inference for dynamic transcriptional regulation; the Hes1 system as a case study. *Bioinformatics 23*, 2596–2603.

Krishnarajah, I., A. Cook, G. Marion, and G. Gibson (2005). Novel moment closure approximations in stochastic epidemics. *Bulletin of Mathematical Biology 67*, 855–873.

Leclant, F. and J. P. Deguine (1994). Cotton aphids. In G. A. Matthews and J. P. Tunstall (Eds.), *Insect Pests of Cotton*, pp. 285–323. U.K.: CAB International.

Matis, J. H., T. R. Kiffe, T. I. Matis, J. A. Jackman, and H. Singh (2007). Population size models based on cummulative size, with application to aphids. *Ecological Modelling 205*, 81–92.

Matis, J. H., T. R. Kiffe, T. I. Matis, and D. E. Stevenson (2006). Application of population growth models based on cumulative size to Pecan aphids. *Journal of Agricultural, Biological, and Environmental Statistics 11*, 425–449.

Matis, J. H., T. R. Kiffe, T. I. Matis, and D. E. Stevenson (2007). Stochastic modeling of aphid population growth with nonlinear power-law dynamics. *Mathematical Biosciences 208*, 469–494.

Matis, J. H., T. R. Zhou, T. R. Kiffe, and T. I. Matis (2007). Fitting cumulative size mechanistic models to insect population data: A nonlinear random effects model analysis. *Journal of the Indian Society of Agricultural Statistics 61*, 147–155.

Matis, T. I., M. N. Parajulee, J. H. Matis, and R. B. Shrestha (2008). A mechanistic model based analysis of cotton aphid population dynamics data. *Agricultural and Forest Entomology 10*, 1–8.

Milner, P., C. S. Gillespie, and D. J. Wilkinson (2009). Parameter inference using moment closure models. In preparation.

O'Hagan, A. and J. Forster (2004). *Kendall's Advanced Theory of Statistics* (2nd ed.), Volume Volume 2B. Arnold.

Prajneshu (1998). A nonlinear statistical model for aphid population growth. *Journal of the Indian Society of Agricultural Statistics 51*, 73–80.

Purutcuoglu, V. and E. Wit (2007). Bayesian inference of the kinetic parameters of a realistic MAPK/ERK pathway. *BMC Systems Biol. 1*, P19.

Roberts, G. O. and J. Rosenthal (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science 16*(4), 351–367.

Rummel, D. R., M. D. Arnold, J. E. Slosser, K. C. Neece, and W. E. Pinchak (1995). Cultural factors influencing the abundance of Aphis gossypii glover in Texas High Plains cotton. *Southwestern Entomologist 20*, 395–406.

Singh, A. and J. Hespanha (2007). A derivative matching approach to moment closure for the stochastic logistic model. *Bulletin of Mathematical Biology 69*, 1909–1925.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association 82*(398), 528–540.

Wilkinson, D. J. (2006). *Stochastic Modelling for Systems Biology*. Chapman and Hall/CRC Press, London.

Xia, J. Y., W. Van der Werf, and R. Rabbinge (1999). Influence of temperature on bionomics of cotton aphid, Aphis gossypii, on cotton. *Entomologia Experimentalis et Applicata 90*, 25–35.

Zheng, Q. and J. Ross (1991). Comparison of deterministic and stochastic kinetics for nonlinear systems. *Journal of Chemical Physics 94*, 3644–3648.