Collapsing of non-centered parameterised MCMC algorithms with applications to epidemic models

Peter Neal (Lancaster University) and Fei Xiang (University of Cambridge)

May 17, 2016

Running title: Collapsing of MCMC algorithms

Abstract

Data augmentation is required for the implementation of many MCMC algorithms. The inclusion of augmented data can often lead to conditional distributions from well-known probability distributions for some of the parameters in the model. In such cases, collapsing (integrating out parameters) has been shown to improve the performance of MCMC algorithms. We show how integrating out the infection rate parameter in epidemic models leads to efficient MCMC algorithms for two very different epidemic scenarios, final outcome data from a multitype SIR epidemic and longitudinal data from a spatial SI epidemic. The resulting MCMC algorithms give fresh insight into real life epidemic data sets.

Keywords: Collapsing; measles; non-centred MCMC algorithms; spatial epidemics; stochastic epidemic models.

1 Introduction

A key aim of parametric Bayesian statistics is, given data $\mathbf{x}$ which are assumed to arise from a model $\mathcal{M}$ with unknown parameters $\boldsymbol{\theta}$, to obtain the posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta}|\mathbf{x})$. For all, but the simplest of problems this is not analytically tractable and often an MCMC algorithm is used to obtain samples from $\pi(\boldsymbol{\theta}|\mathbf{x})$. Furthermore, the implementation of an MCMC algorithm will often require data augmentation, $\mathbf{y}$, with the resulting algorithm producing samples from $\pi(\boldsymbol{\theta}, \mathbf{y}|\mathbf{x})$ with the marginal distribution $\pi(\boldsymbol{\theta}|\mathbf{x})$ of primary interest. This raises the question of how to construct an efficient MCMC algorithm to obtain samples from $\pi(\boldsymbol{\theta}, \mathbf{y}|\mathbf{x})$?

One approach which can assist in developing an efficient MCMC algorithm is collapsing, Liu (1994), that is, to integrate out some of the parameters from the model and construct an MCMC algorithm for a subset of the parameters. Specifically, suppose that $\boldsymbol{\theta} = (\lambda, \boldsymbol{\phi})$ and that the conditional distribution of λ

given $\boldsymbol{\phi}$ and $\mathbf{x}$, $\pi(\lambda|\boldsymbol{\phi}, \mathbf{y}, \mathbf{x})$ is known. In this case, we can simply integrate out λ to leave $\pi(\boldsymbol{\phi}|\mathbf{x})$. In Liu (1994) particular focus is placed upon collapsing for the Gibbs sampler but the approach can easily be applied to any MCMC algorithm, where $\pi(\lambda|\boldsymbol{\phi}, \mathbf{y}, \mathbf{x})$ is known, see, for example Neal and Roberts (2005) for an epidemic example.

In this paper λ is the infection rate of an epidemic model. For a number of epidemic models the augmented data can be chosen independently of λ leading to a non-centered parameterisation, Papaspiliopoulos *et al.* (2003). We show how the augmented data can be chosen to give a straightforward to compute likelihood. Then by integrating out not only λ but also a subset of the augmented data we obtain a tractable likelihood which can be utilised within an efficient MCMC algorithm. The generic approach is introduced in Section 2 with the details being model specific. The methodology is illustrated with two distinct epidemic models; final outcome data from a multitype SIR epidemic (Section 3) and longitudinal data from a spatial SI epidemic (Section 4). These highlight the ease with which the collapsing of the MCMC algorithm can be implemented and the significant efficiency gains that it offers. Finally, we briefly summarise the findings of the paper in Section 5.

2 Generic collapsing setup

In this Section we outline the generic collapsing approach taken in this paper. This allows us to highlight the key elements in choosing the data augmentation and implementing the collapsing for epidemic models.

Let $\boldsymbol{\theta} = (\lambda, \boldsymbol{\phi})$ and $\mathbf{y} = (\mathbf{v}, \mathbf{w})$ denote the parameters of the model and the augmented data, respectively. The parameters and augmented data are each divided into two sets with λ and $\mathbf{v}$ denoting parameters and augmented data which are to be integrated out and $\boldsymbol{\phi}$ and $\mathbf{w}$ denoting the remaining parameters and augmented data. Throughout this Section we assume that λ is one-dimensional, for ease of exposition and since this is the case in the examples in Sections 3 and 4 and generally likely to be the case in practice. However, the following discussion straightforwardly extends to λ being multidimensional and even in the one-dimensional case the effect on the performance of the MCMC algorithm can be dramatic as we highlight in Section 3.

The joint posterior distribution of $\boldsymbol{\theta}$ and $\mathbf{y}$ satisfies

$$\pi(\boldsymbol{\theta}, \mathbf{y}|\mathbf{x}) \quad \propto \quad \pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})\pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{2.1}$$

We make a few assumptions about the form of the righthand side of (2.1) in order to implement the collapsing. Firstly, that the priors on λ and $\boldsymbol{\phi}$ are independent. Secondly, that the augmented data

$\mathbf{y} = (\mathbf{v}, \mathbf{w})$ is independent of λ. Thirdly, that the augmented data to be integrated out, $\mathbf{v}$ is independent of $\boldsymbol{\phi}$ and $\mathbf{w}$. Under these assumptions (2.1) satisfies

$$\begin{aligned} \pi(\lambda, \boldsymbol{\phi}, \mathbf{v}, \mathbf{w}|\mathbf{x}) &\propto \pi(\mathbf{x}|\mathbf{v}, \mathbf{w}, \lambda, \boldsymbol{\phi})\pi(\mathbf{v}, \mathbf{w}|\lambda, \boldsymbol{\phi})\pi(\lambda, \boldsymbol{\phi}) \\ &\propto \pi(\mathbf{x}|\mathbf{v}, \mathbf{w}, \lambda, \boldsymbol{\phi})\pi(\mathbf{w}|\boldsymbol{\phi})\pi(\mathbf{v})\pi(\boldsymbol{\phi})\pi(\lambda). \end{aligned} \tag{2.2}$$

The final assumption that we make regards the form of $\pi(\mathbf{x}|\mathbf{v}, \mathbf{w}, \lambda, \boldsymbol{\phi})$, we assume that there exists a deterministic function $h(\cdot, \cdot, \cdot, \cdot)$ such that, if $\mathbf{V}$ and $\mathbf{W}$ are drawn from $\pi(\mathbf{v})$ and $\pi(\mathbf{w}|\boldsymbol{\phi})$, respectively, $h(\lambda, \boldsymbol{\phi}, \mathbf{V}, \mathbf{W})$ gives a realisation of $\mathbf{X}$ from the model $\mathcal{M}$ with parameters $\boldsymbol{\theta} = (\lambda, \boldsymbol{\phi})$. Note that different realisations of the process can be generated by changing some or all of λ, $\boldsymbol{\phi}$, $\mathbf{V}$ and $\mathbf{W}$. In this case

$$\pi(\mathbf{x}|\lambda, \boldsymbol{\phi}, \mathbf{v}, \mathbf{w}) = 1_{\{h(\lambda, \boldsymbol{\phi}, \mathbf{v}, \mathbf{w}) = \mathbf{x}\}}. \tag{2.3}$$

This is a general situation observed in both Sections 3 and 4. The discontinuous density in (2.3) can make constructing an efficient MCMC algorithm difficult, see for example Neal *et al.* (2012). The integration out of $\mathbf{v}$ and λ gives

$$P(\boldsymbol{\phi}, \mathbf{w}) = \int_{\lambda} \int_{\mathbf{v}} 1_{\{h(\lambda, \boldsymbol{\phi}, \mathbf{v}, \mathbf{W}) = \mathbf{x}\}} \pi(\lambda)\pi(\mathbf{v})\, d\mathbf{v}\, d\lambda, \tag{2.4}$$

the probability that, given $\boldsymbol{\phi}$ and $\mathbf{w}$, λ and $\mathbf{v}$ sampled from $\pi(\lambda)$ and $\pi(\mathbf{v})$, respectively, will result in $h(\lambda, \boldsymbol{\phi}, \mathbf{v}, \mathbf{w}) = \mathbf{x}$. Note that the inclusion of augmented data $\mathbf{v}$ into the model to then simply integrate it out again may seem unnecessary but it is helpful in understanding the model dynamics and enabling us to exploit (2.3) directly. By constructing an MCMC algorithm based on (2.4) rather than (2.3), we are exploiting a Rao-Blackwellisation, see, for example, Smith and Roberts (1993). Specifically, we are replacing an unbiased, indicator function estimate of the likelihood ($E[1_{\{h(\lambda, \boldsymbol{\phi}, \mathbf{V}, \mathbf{W}) = \mathbf{x}\}}|\lambda, \boldsymbol{\phi}] = \pi(\mathbf{x}|\lambda, \boldsymbol{\phi})$) by an unbiased, probability estimate of the likelihood ($E[P(\boldsymbol{\phi}, \mathbf{W})|\boldsymbol{\phi}] = \pi(\mathbf{x}|\boldsymbol{\phi})$). As we shall observe in Section 3 this substantially improves the performance of the MCMC algorithm.

By integrating out λ and $\mathbf{v}$, it follows from (2.2) and (2.4) that

$$\pi(\boldsymbol{\phi}, \mathbf{w}|\mathbf{x}) \propto P(\boldsymbol{\phi}, \mathbf{w})\pi(\mathbf{w}|\boldsymbol{\phi})\pi(\boldsymbol{\phi}). \tag{2.5}$$

Therefore it is straightforward to construct an MCMC algorithm which alternates between updating the parameters $\boldsymbol{\phi}$ and the augmented data $\mathbf{w}$. However, we want samples from $\pi(\lambda, \boldsymbol{\phi}|\mathbf{x})$. This can easily be done using a sample $(\boldsymbol{\phi}, \mathbf{w})$ from $\pi(\boldsymbol{\phi}, \mathbf{w}|\mathbf{x})$. Then

$$\pi(\lambda|\boldsymbol{\phi}, \mathbf{w}, \mathbf{x}) \propto \pi(\lambda) \int_{\mathbf{v}} 1_{\{h(\lambda, \boldsymbol{\phi}, \mathbf{w}, \mathbf{v}) = \mathbf{x}\}} \pi(\mathbf{v})\, d\mathbf{v}, \tag{2.6}$$

and we can sample λ using (2.6) provided that the integral can be computed. The sampling of a one-dimensional parameter λ is usually straightforward. For the examples considered in Sections 3 and 4 direct simulation from the conditional distribution is possible using a sample $\mathbf{v}$ from $\pi(\mathbf{v})$.

We summarise the generic MCMC algorithm with details being model specific and given in Sections 3 and 4.

MCMC algorithm

i) Propose $\boldsymbol{\phi}'$ from $q_{\boldsymbol{\phi}}(\boldsymbol{\phi}, \cdot)$ and accept the proposed move with probability

$$\frac{\pi(\boldsymbol{\phi}', \mathbf{w}|\mathbf{x}) q_{\boldsymbol{\phi}}(\boldsymbol{\phi}', \boldsymbol{\phi})}{\pi(\boldsymbol{\phi}, \mathbf{w}|\mathbf{x}) q_{\boldsymbol{\phi}}(\boldsymbol{\phi}, \boldsymbol{\phi}')} \wedge 1.$$

ii) Propose $\mathbf{w}'$ from $q_{\mathbf{w}}(\mathbf{w}, \cdot)$ and accept the proposed move with probability

$$\frac{\pi(\boldsymbol{\phi}, \mathbf{w}'|\mathbf{x}) q_{\mathbf{w}}(\mathbf{w}', \mathbf{w})}{\pi(\boldsymbol{\phi}, \mathbf{w}|\mathbf{x}) q_{\mathbf{w}}(\mathbf{w}, \mathbf{w}')} \wedge 1.$$

iii) Sample $\pi(\lambda|\boldsymbol{\phi}, \mathbf{w}, \mathbf{x})$ using (2.6).

iv) Store $\boldsymbol{\theta} = (\lambda, \boldsymbol{\phi})$ as a sample from $\pi(\boldsymbol{\theta}|\mathbf{x})$.

In practice steps (i) and (ii) of the algorithm might comprise multiple steps for updating different sets of parameters and augmented data, respectively.

In the examples considered in Sections 3 and 4, λ represents the infection rate and the integrating out the infection rate allows the MCMC algorithm to move efficiently to effectively determine an appropriate infection rate given the other parameters and the augmented data. In both examples the augmented data $\mathbf{w}$ consists of $\boldsymbol{\omega}$, the order of infection, $\mathbf{L}$, the additional infectious pressure required for successive infections and for the measles example in Section 3 the infectious periods $\mathbf{I}$. In both cases the data $\mathbf{v}$ is a subset of the infectious pressures $\mathbf{L}$, chosen to ensure that the correct number of infections take place. The other parameters ($\boldsymbol{\phi}$) are model specific, vaccine efficacy in the measles example in Section 3 and spatial and background risk to infection in the spatial SI epidemic in Section 4.

3 Final size epidemic data

In this Section we illustrate collapsing of an MCMC algorithm using a non-centered parameterisation for final size epidemic data. The motivating example is the outbreak of measles in a Finnish school (a small, assumed to be homogeneously mixing, community), Paunio *et al.* (1998), where individuals are grouped by vaccination status. A second measles example where there is missing vaccination statuses of individuals (van Boven *et al.* (2010)) is considered in the supplementary material. We start by describing the data. The data are analysed using a Sellke (Sellke (1983)) construction of the epidemic process. This

involves extending the non-centered parameterisation for SIR epidemic models employed in Neal (2012), Section 3 to multiple types of individuals.

The data consist of how many school pupils were infected in a measles outbreak in Honkajoki, a small rural Finnish municipality in 1989, Paunio *et al.* (1998). Pupils belong to one of three types, 0, 1 or 2, where a type k ($k = 0, 1, 2$) individual has received k doses of measles vaccines. Let x_k and n_k denote the total number of infected individuals and the total number of individuals of type k, respectively, with $m = x_0 + x_1 + x_2$ and $N = n_0 + n_1 + n_2$. The data are summarised in Table 1.

Table 1 about here.

A vaccine can affect both an individual's susceptibility to and infectivity with a disease, Halloran *et al.* (2010). Given the data, there is insufficient information to model both variable susceptibility and infectivity, see, for example, van Boven *et al.* (2010), and therefore we assume that the vaccine only affects an individual's susceptibility to the disease. Therefore the model is as follows. The epidemic is an SIR epidemic in a closed, homogeneously mixing population of size N. That is, apart from the initial infective who introduces the disease to the school, nobody is infected from outside the school and we do not model infectious contacts made by pupils with individuals outside of the school. The epidemic is initiated by a single infective with the extension to multiple initial infectives trivial. Whilst infectious, an individual makes infectious contacts at the points of a homogeneous Poisson process with rate λ, with the individual contacted chosen uniformly at random from the population. If the individual contacted by an infective is a susceptible individual of type k, they become infected with probability q_k, where $q_0 = 1$ and (q_1, q_2) are unknown parameters to be estimated. Therefore, for $k = 1, 2$, $1 - q_k$ denotes the protective benefit of being vaccinated k times. Infectious contacts with non-susceptible individuals have no affect on the recipient. That is, individuals can be infected at most once. The infectious periods of infected individuals are assumed to be independent and identically distributed according to a random variable I. Final size data contains no temporal information about the epidemic and is invariant to replacing (λ, I) by $(c\lambda, I/c)$ for any $c > 0$ (Ball and O'Neill (1999), van Boven *et al.* (2010), Neal (2012)). Therefore without loss of generality, we take I to have mean 1, so that λ denotes the basic reproduction of the epidemic. We assume $U(0, 1)$ priors for q_1 and q_2 and a Gamma(a, b) prior for λ. (Setting $a = 1$ and $b = 0$ gives an improper uniform prior for λ.)

In order to implement a non-centered parameterisation it is convenient to use a Sellke type construction, Sellke (1983) of the epidemic process, see for example Neal (2012). The construction differs slightly from Neal (2012) in that we have multiple types of individuals. For $j = 0, 1, 2$ and $i = 1, 2, \ldots$, let $s_{j,i}$ denote the total number of susceptibles of type j after the i^{th} infection with $s_{j,0} = n_j$. For $i = 1, 2, \ldots$, let

$\alpha_i = \sum_{j=0}^{2} s_{j,i} q_j / N$, the probability that following the i^{th} infection, an infectious contact will result in infection (a contact is with a susceptible and that the contact is successful). We augment the observed data $\mathbf{x}$ by $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_m)$, $\mathbf{I} = (I_1, I_2, \ldots, I_m)$ and $\mathbf{L} = (L_1, L_2, \ldots, L_m)$ defined as follows. Let $\boldsymbol{\omega}$ denote the order in which individuals are infected with $\omega_j = k$ if the j^{th} individual infected is of type k. Let I_j denote the infectious period of the j^{th} individual. Finally, the L_j's are independent and identically distributed according to $L_1 \sim \text{Exp}(1)$, and their role will be discussed in detail below.

We assume that $\boldsymbol{\omega}$ is consistent with the data, that is, x_k elements of $\boldsymbol{\omega}$ are equal to k ($k = 0, 1, 2$). Then

$$\pi(\boldsymbol{\omega}|\mathbf{q}) = \prod_{i=1}^{m} \frac{s_{\omega_i, i-1} q_{\omega_i}}{\sum_{k=0}^{2} s_{k,i-1} q_k}, \tag{3.1}$$

where $q_0 = 1$. Thus the order $\boldsymbol{\omega}$ explicitly depends on the parameters $\mathbf{q}$. Given that i infections have taken place $L_i/(\alpha_i \lambda)$ denotes the additional infectious pressure (units of infection) needed to ensure that the $(i+1)^{st}$ individual is infected. This is consistent with infectives making infectious contacts at rate λ with success probability α_i. Thus, given $\boldsymbol{\omega}$, $\mathbf{q}$, $\mathbf{L}$ and $\mathbf{I}$, the epidemic infects m individuals of types $\mathbf{x}$, if for all $1 \leq k \leq m-1$,

$$\frac{1}{\lambda} \sum_{i=1}^{k} \frac{L_i}{\alpha_i} \leq \sum_{i=1}^{k} I_i, \tag{3.2}$$

and

$$\frac{1}{\lambda} \sum_{i=1}^{m} \frac{L_i}{\alpha_i} > \sum_{i=1}^{m} I_i. \tag{3.3}$$

Let $h(\lambda, \mathbf{q}, \boldsymbol{\omega}, \mathbf{L}, \mathbf{I})$ denote the epidemic process generated by the first m infections. Then if $\boldsymbol{\omega}$ is consistent with the data and (3.2) and (3.3) are satisfied, we have that $h(\lambda, \mathbf{q}, \boldsymbol{\omega}, \mathbf{L}, \mathbf{I}) = \mathbf{x}$. We can construct an MCMC algorithm which moves around the joint space of $(\lambda, \mathbf{q}, \boldsymbol{\omega}, \mathbf{L}, \mathbf{I})$ but we collapse the algorithm by integrate out λ and L_m. That is, in the notation of Section 2, $\boldsymbol{\phi} = \mathbf{q}$, $\mathbf{w} = (\boldsymbol{\omega}, \mathbf{L}_{1:m-1}, \mathbf{I})$ and $\mathbf{v} = L_m$, where $L_{1:m-1} = (L_1, L_2, \ldots, L_{m-1})$. By integrating out L_m, a simple conditional distribution for λ, $\lambda|\boldsymbol{\omega}, \mathbf{q}, \mathbf{L}_{1:m-1}, \mathbf{I}, \mathbf{x}$, exists. From (3.2) and (3.3), we require that

$$H_m = \max_{1 \leq k \leq m-1} \frac{\sum_{i=1}^{k} L_i/\alpha_i}{\sum_{i=1}^{k} I_i} \leq \lambda < \frac{\sum_{i=1}^{m} L_i/\alpha_i}{\sum_{i=1}^{m} I_i}. \tag{3.4}$$

Thus $\lambda > H_m$ and $L_m > \alpha_m \{\lambda \sum_{i=1}^{m} I_i - \sum_{i=1}^{m-1} L_i/\alpha_i\} = J_m$, say, gives

$$\begin{aligned}
P(\mathbf{q}, \boldsymbol{\omega}, \mathbf{L}_{1:m-1}, \mathbf{I}|\mathbf{x}) &\propto \int_0^\infty \int_0^\infty \pi(\mathbf{x}|\lambda, \mathbf{q}, \boldsymbol{\omega}, \mathbf{L}, \mathbf{I}) \pi(\lambda) \pi(\mathbf{q}) \pi(\boldsymbol{\omega}|\mathbf{q}) \pi(\mathbf{L}) \pi(\mathbf{I}) \, dL_m \, d\lambda \\
&\propto \int_{H_m}^\infty \int_{J_m}^\infty \exp(-L_m) dL_m \frac{\lambda^{a-1}}{\Gamma(a)} \exp(-\lambda b) d\lambda \pi(\boldsymbol{\omega}|\mathbf{q}) \pi(\mathbf{L}_{1:m-1}) \pi(\mathbf{I}) \\
&= \exp\left(\alpha_m \sum_{i=1}^{m-1} L_i/\alpha_i\right) \pi(\boldsymbol{\omega}|\mathbf{q}) \pi(\mathbf{L}_{1:m-1}) \pi(\mathbf{I}) \\
&\quad \times \int_{H_m}^\infty \exp\left(-\lambda \alpha_m \sum_{i=1}^{m} I_i\right) \frac{\lambda^{a-1}}{\Gamma(a)} \exp(-\lambda b) d\lambda
\end{aligned} \tag{3.5}$$

The integral on the righthand side of (3.5) is $P(Z > H_m)/(b + \alpha_m \sum_{i=1}^m I_i)^a$, where $Z \sim \text{Gamma}(a, b + \alpha_m \sum_{i=1}^m I_i)$. Thus if $a \in \mathbb{N}$, corresponding to an Erlang prior (Gamma distribution with integer shape parameter) on λ, it follows from (3.5) that

$$\begin{aligned} P(\mathbf{q}, \boldsymbol{\omega}, \mathbf{L}_{1:m-1}, \mathbf{I}|\mathbf{x}) &= \left\{ \sum_{k=0}^{a-1} \frac{(b + \alpha_m \sum_{i=1}^m I_i)^{k-a}}{k!} \right\} \pi(\boldsymbol{\omega}|\mathbf{q}) \pi(\mathbf{L}_{1:m-1}) \pi(\mathbf{I}) \\ &\quad \times \exp\left(-\alpha_m \left\{ H_m \sum_{i=1}^m I_i - \sum_{i=1}^{m-1} L_i/\alpha_i \right\} \right). \end{aligned} \tag{3.6}$$

We are now in position to describe a collapsed MCMC algorithm based upon (3.5) for obtaining samples from $\pi(\lambda, \mathbf{q}|\mathbf{x})$. The acceptance probability for each step is straightforward to compute using (3.5) and (3.1). Below we describe the steps per iteration with $(\lambda, \mathbf{q})$ stored at the end of each iteration.

MCMC algorithm

i) Update (q_1, q_2) using random walk Metropolis. We propose $q_k' \sim N(q_k, \sigma_q^2)$ $(k = 1, 2)$, with reflection at the boundaries 0 and 1 to ensure that $0 \leq q_1', q_2' \leq 1$.

ii) Update in turn the augmented data $\mathbf{w} = (\boldsymbol{\omega}, \mathbf{L}_{1:m-1}, \mathbf{I})$ as follows

   (a) Update $\boldsymbol{\omega}$ using an independence sampler with $\boldsymbol{\omega}'$ sampled uniformly at random from the set of possible orderings consistent with the data.

   (b) Update $\mathbf{L}_{1:m-1}$ by proposing to update a random subset $\mathcal{P}$ of the thresholds, where $|\mathcal{P}| = p$. If $i \in \mathcal{P}$, draw $L_i' \sim \text{Exp}(1)$, otherwise set $L_i' = L_i$.

   (c) Update $\mathbf{I}$ by proposing to update a random subset $\mathcal{R}$ of the infection periods, where $|\mathcal{R}| = r$. If $i \in \mathcal{R}$, draw $I_i' \sim I$, otherwise set $I_i' = I_i$.

iii) Draw $\lambda|\boldsymbol{\omega}, \mathbf{q}, \mathbf{L}_{1:m-1}, \mathbf{I}, \mathbf{x}$ from its conditional distribution, which is $\text{Gamma}(a, b + \alpha_m \sum_{i=1}^m I_i)$ conditioned to be greater than H_m. If $a = 1$, we can exploit the memoryless property of the exponential random variable to give

$$\lambda|\boldsymbol{\omega}, \mathbf{q}, \mathbf{L}_{1:m-1}, \mathbf{I}, \mathbf{x} \sim H_m + \text{Exp}\left(b + \alpha_m \sum_{i=1}^m I_i \right). \tag{3.7}$$

iv) Store $\boldsymbol{\theta} = (\lambda, \mathbf{q})$ as a sample from $\pi(\boldsymbol{\theta}|\mathbf{x})$.

We briefly comment upon the algorithm for the Honkajoki data. For updating $\mathbf{q}$, we want an acceptance rate of approximately 35% to optimise the random walk Metropolis. This is achieved by choosing $\sigma_q = 0.15$. More generally, σ_q can be chosen adaptively using an adaptive MCMC scheme, see, for example

Xiang and Neal (2014). The independence sampler for $\boldsymbol{\omega}$ has a high acceptance rate of 90%, which shows that the order of infection is largely irrelevant. If the data are such that the order of $\boldsymbol{\omega}$ is more important an updating scheme along the lines of that used in Section 4 could alternatively be used. For updating $\mathbf{L}_{1:m-1}$, we choose $p = 15$, which results in an acceptance rate of 33%. There is a compromise here as increasing p decreases the acceptance rate and maximise p times the acceptance rate gives close to optimal behaviour, see Xiang and Neal (2014). We choose $\pi(\lambda) \propto 1$ $(\lambda > 0)$ corresponding to an improper Gamma$(1, 0)$ prior. Finally, we consider three scenarios for I; $I \equiv 1$, $I \sim$ Gamma$(2, 2)$ and $I \sim$ Gamma$(\exp(\gamma), \exp(\gamma))$ with γ unknown. Except in the case where $I \equiv 1$, we updated $r = 15$ infection times together at each iteration. In the case of unknown γ we assigned an Exp(1) prior (this ensures that the shape parameter on the Gamma distribution is greater than or equal to 1) and added a random walk step to the MCMC algorithm to update γ with proposal standard deviation 1.

For all three scenarios the algorithm was run for 110000 iterations with the first 10000 iterations discarded as burn-in. The initial values for $\mathbf{q}$, $\boldsymbol{\omega}$ and $\mathbf{L}_{1:m-1}$ were drawn from the prior on $\mathbf{q}$, a random permutation of the m infections and Exp(1), respectively. The initial values for $\mathbf{I}$ are drawn from the appropriate distribution with $\gamma = 1$ in the third scenario. The burn-in is excessive with convergence appearing to be almost instantaneous, except for γ. Autocorrelation function plots show that for all the parameters, except γ, the correlation is decaying rapidly and is approximately 0 by lag-30. The estimated posterior means, standard deviations and effective sample sizes for $(\lambda, \mathbf{q}, \gamma)$ are given in Table 2. The poor mixing in γ is due to the lack of information in the final size data about the infectious period distribution, highlighted by the variability in γ and the similar parameter estimates for all three scenarios with $I \equiv 1$ corresponding to the limit as $\gamma \to \infty$ of $I \sim$ Gamma$(\exp(\gamma), \exp(\gamma))$.

Table 2 about here.

For comparison we ran the corresponding MCMC algorithms without λ and L_m integrated out for 110000 iterations with the first 10000 iterations discarded as burn-in. We initialised with $\lambda = 3$ and simulated $\mathbf{q}$, $\boldsymbol{\omega}$, $\mathbf{L}$ and $\mathbf{I}$ until we obtained an epidemic consistent with the data. The modifications to the above algorithm are that $\mathbf{L}$ is updated in place of $\mathbf{L}_{1:m-1}$ and λ drawn uniformly from

$$\left(\max_{1 \le k \le m-1} \left\{ \sum_{i=1}^{k} \frac{L_i}{\alpha_i} / \sum_{i=1}^{k} I_i \right\}, \sum_{i=1}^{m} \frac{L_i}{\alpha_i} / \sum_{i=1}^{m} I_i \right),$$

which ensures that (3.2) and (3.3) are satisfied. To optimise performance we set $\sigma_q = 0.05$, $p = 5$ and $r = 5$. The estimated posterior means and standard deviations are similar to those reported in Table 2 for the collapsed MCMC algorithm but the mixing is far worse. The estimated effective sample sizes for q_1, q_2, λ and γ for the non-collapsed MCMC algorithm are also given in Table 2. The effective sample sizes

with the exception of γ are in all cases at least 18 times higher for the corresponding collapsed MCMC algorithm and since both the collapsed and non-collapsed MCMC algorithms have virtually identical running times this represents a very significant improvement. The similar performance in the mixing of γ in the algorithms is due to the updating of γ only depending on $\mathbf{I}$ and thus is largely unaffected by the inclusion of collapsing.

The choice of an improper Gamma$(1, 0)$ prior in the above analysis makes implementing the MCMC algorithm particularly straightforward as the first term on the right hand side of (3.6) is simply $1/(\alpha_m \sum_{i=1}^{m} I_i)$ and (3.7) can be exploited to sample λ. For an Erlang prior distribution, Gamma(a, b) with $a \in \mathbb{N}$, on λ we can still use (3.6) and sample λ from a truncated Gamma distribution. This involves minor adjustments to the code with very similar results in terms of mixing but with the code taking between $5 - 10\%$ and $10 - 15\%$ longer to run with Gamma$(2, 2)$ and Gamma$(10, 10)$ priors, respectively, on λ.

Finally, in van Boven *et al.* (2010) there are other measles data sets which the above MCMC algorithm can be applied to. Moreover, there are data from a measles outbreak from a school in Duisburg, Germany with missing vaccination status for approximately 30% of the population. Details of how to extend the MCMC algorithm to incorporate missing information and the corresponding analysis are given in the supplementary material.

4 Spatial epidemic

In this Section we consider spatial $S \to I$ epidemic models where individuals start off susceptible but once they become infected they remain so forever. A spatial $S \to I$ epidemic model is appropriate for many agricultural diseases, for example, *citrus tristeza virus* (CTV) in a citrus orchard, Gibson (1997a), Gibson (1997b), and the spread of *Sugarcane yellow leaf virus* (SCYLV) on a sugarcane plantation, Daugrois *et al.* (2011) Brown *et al.* (2014). In the above papers the plants are located on a 2-dimensional rectangular lattice, $\mathcal{L}$, although the approach is not limited to this case. We continue by describing the infectious process in the spatial $S \to I$ epidemic model, and typical longitudinal data for such scenarios. The starting point for our research is the MCMC algorithm of Gibson (1997a) which captures the spatial spread of the disease, but not the rate of infection. We extend the MCMC algorithm of Gibson (1997a) to incorporate estimation of the infection rate parameter and as in Section 3 we integrate out the infection parameter. The result is an efficient MCMC algorithm which can be used for forward (and backward) prediction of cases. Moreover, we show that by ignoring the rate of transmission of the disease the MCMC algorithm of Gibson (1997a) can induce a slight bias into the estimation of the spatial parameter. We

highlight the differences and advantages of this approach to that taken in Brown *et al.* (2014) where estimation of both the spatial spread of the disease and the infection rate is performed. Finally, we employ the new MCMC algorithm to the CTV data, Gibson (1997a) and the spread of SCYLV, Daugrois *et al.* (2011), Brown *et al.* (2014).

Let $\mathbf{x}, \mathbf{y} \in \mathcal{L}$ denote the location of two individuals on a subset of $\mathbb{R}^2$ with typically $\mathcal{L}$ being a lattice. Once individual $\mathbf{x}$ is infected it makes infectious contacts with individual $\mathbf{y}$ at the points of an inhomogeneous Poisson point process with rate $k(t)F_\alpha(\mathbf{x}-\mathbf{y})$, where $k(t)$ represents an underlying infection rate (non-negative and possibly time varying) and $F_\alpha(\cdot)$ is a non-negative function (parameterised by α) which characterises the force of infection from $\mathbf{x}$ and $\mathbf{y}$. Throughout this paper, and in line with previous work, the force of infection between two individuals will depend upon their Euclidean distance and $F_\alpha(\cdot)$ will be a monotonically decreasing function of distance. If individual $\mathbf{y}$ is susceptible when individual $\mathbf{x}$ makes infectious contact it becomes infected (and immediately infectious), otherwise the infectious contact has no affect on individual $\mathbf{y}$ as it is already infectious. We follow Brown *et al.* (2014), by assuming that $k(t)$ is constant ($k(t) = \lambda$ for all t). (Note that Brown *et al.* (2014) uses β rather than λ for the infection rate.) The extension to piecewise constant infection rates with change points corresponding to the observation times is trivial. Finally, it is commonplace, see, Gibson (1997b) and Brown *et al.* (2014), to assume that there is an external background risk to infection, $k(t)r$. Thus if $\mathcal{I}$ denotes the set of location of infectives at times t, the infectious exposure that an individual $\mathbf{y}$ is subjected to at time t is $\lambda\{r+\sum_{\mathbf{x}\in\mathcal{I}} F_\alpha(\mathbf{x}-\mathbf{y})\}$.

The observed data are assumed to be snapshots of the $S \to I$ epidemic at a discrete set of time points. Let $t_0(=0) < t_1 < \ldots < t_m$ denote times at which the infectious status of all individuals are known. For $i = 0, 1, \ldots, m$, let $\mathcal{S}_i$ and $\mathcal{I}_i$ denote the set of susceptible and infected individuals, respectively at time point t_i. For $i = 1, 2, \ldots, m$, let $\mathcal{W}_i = \mathcal{I}_i \backslash \mathcal{I}_{i-1}$, the set of individuals infected between time points t_{i-1} and t_i with $n_i = |\mathcal{W}_i|$. In Gibson (1997a), $m = 1$ and the two time points correspond to dates a year apart. In Brown *et al.* (2014), $m = 6$ with $\mathbf{t} = (0, 6, 10, 14, 19, 23, 30)$ and time units of a week. (Note that in this case $\mathbf{t}$ should be $(0, 6, 11, 15, 19, 23, 30)$, see Daugrois *et al.* (2011).)

In order to construct a tractable likelihood we need to use data augmentation. We begin by following Gibson (1997a) by specifying the order in which infections occur in a given interval. Let $\mathbf{y}_{\omega(i,j)}$ denote the j^{th} individual infected in the i^{th} time interval. Let $Q_{i,j}$ denote the length of time from the $(j-1)^{st}$ infection in time interval i until the j^{th} infection, where the time of the 0^{th} infection is taken to be t_{i-1} and the time of the $(n_i+1)^{st}$ is taken to be after time t_i. Then if $\mathcal{I}_{i,j-1}$ and $\mathcal{S}_{i,j-1}$ are the sets of

infectives and susceptibles, respectively, after the $j-1^{st}$ infection in time interval i,

$$\begin{aligned} Q_{i,j} &\sim \text{Exp}\left(\lambda \sum_{\mathbf{y}\in\mathcal{S}_{i,j-1}} \left\{ r + \sum_{\mathbf{x}\in\mathcal{I}_{i,j-1}} F_\alpha(\mathbf{x}-\mathbf{y}) \right\}\right) \\ &= \left\{ \lambda \sum_{\mathbf{y}\in\mathcal{S}_{i,j-1}} \left\{ r + \sum_{\mathbf{x}\in\mathcal{I}_{i,j-1}} F_\alpha(\mathbf{x}-\mathbf{y}) \right\}\right\}^{-1} \text{Exp}(1) \\ &= \{\lambda h_{i,j}(\alpha, r, \boldsymbol{\omega})\}^{-1} L_{i,j}, \qquad \text{say,} \end{aligned} \tag{4.1}$$

where $\boldsymbol{\omega}$ and $\boldsymbol{\omega}^i$ denote the total set of infection orderings and the total set of infection orderings in time interval i, respectively. We will primarily use $L_{i,j}$ rather than $Q_{i,j}$ in the data augmentation since *a priori* $L_{i,j} \sim \text{Exp}(1)$, a non-centered parameterisation. Note that $L_{i,j}$ has the same distribution to L_i in Section 3 and plays the same role in defining the additional infectious pressure required for the next infection. Let $\mathbf{L}^i = (L_{i,1}, \ldots, L_{i,n_i+1})$ and $\mathbf{L} = (\mathbf{L}^1, \ldots, \mathbf{L}^m)$. Then exploiting the independence of the epidemic process between different time intervals,

$$\begin{aligned} \pi(\mathcal{I}, \boldsymbol{\omega}, \mathbf{L}|r, \alpha, \lambda) &= \prod_{i=1}^{m} \pi(\mathcal{I}_i|\boldsymbol{\omega}^i, \mathbf{L}^i, \alpha, \lambda) \\ &= \prod_{i=1}^{m} \left\{ 1_{\{U_i\}} \prod_{j=1}^{n_i} \left(\frac{r + \sum_{\mathbf{x}\in\mathcal{I}_{i,j-1}} F_\alpha(\mathbf{x} - \mathbf{y}_{\omega(\mathbf{i},\mathbf{j})})}{h_{i,j}(\alpha, r, \boldsymbol{\omega})} \right) \prod_{j=1}^{n_i+1} \exp(-L_{i,j}) \right\}, \end{aligned} \tag{4.2}$$

where U_i denotes the event that n_i infections take place in interval i with

$$1_{\{U_i\}} = 1_{\left\{\sum_{j=1}^{n_i} L_{i,j}/(\lambda h_{i,j}(\alpha,r,\boldsymbol{\omega})) \le t_i - t_{i-1} < \sum_{j=1}^{n_i+1} L_{i,j}/(\lambda h_{i,j}(\alpha,r,\boldsymbol{\omega}))\right\}}. \tag{4.3}$$

Note that the n_i+1^{st} infection after time t_i corresponds to the first infection after time t_{i+1}. We include both L_{i,n_i+1} and $L_{i+1,1}$ in the likelihood, as we explicitly use $L_{i+1,1}$ for the time of the first infection in interval $i+1$ and integrate out L_{i,n_i+1} to ensure that n_i infections occur in interval i. Moreover, the condition (4.3) mirrors the conditions for fixing the size of the measles epidemic, (3.2) and (3.3), in that we require λ to be large enough, so that n_i infections occur in interval i, but also to be small enough that no more infections take place. Thus mirroring Section 3 we seek to integrate out λ and $(L_{1,n_1+1}, \ldots, L_{m,n_m+1})$, the additional infectious pressure to the next infection.

From (4.2), we have that

$$\pi(\boldsymbol{\omega}, \mathbf{L}, r, \alpha, \lambda|\mathcal{I}) \propto \pi(\mathcal{I}, \boldsymbol{\omega}, \mathbf{L}|r, \alpha, \lambda)\pi(r)\pi(\alpha)\pi(\lambda). \tag{4.4}$$

We proceed by integrating out λ and $(L_{1,n_1+1}, \ldots, L_{m,n_m+1})$ before outlining the MCMC algorithm for obtaining samples from $\pi(r, \alpha, \lambda|\mathcal{I})$. We begin with the special case $m = 1$ (the situation considered in Gibson (1997a)) and an improper uniform prior on λ, $\pi(\lambda) \propto 1$ $(\lambda > 0)$. In this case $\pi(\lambda|r, \alpha, \mathcal{I})$ can be

sampled at each iteration as an add-on to the MCMC algorithm of Gibson (1997a). This corresponds to fully integrating out $\mathbf{L}$ in the updating of r, α and $\boldsymbol{\omega}$. Details of the generalisation to $m > 1$ and a Gamma prior on λ, where full integration out of $\mathbf{L}$ is not possible are given in the supplementary material.

Consider the case $m = 1$ and $\pi(\lambda) \propto 1$ ($\lambda > 0$). Then letting $\check{\mathbf{L}}^1 = (L_{1,1}, \ldots, L_{1,n_1})$, we have that

$$\begin{aligned}
\pi(\boldsymbol{\omega}, \check{\mathbf{L}}^1, r, \alpha|\mathcal{I}) &\propto \int_\lambda \int_{L_{1,n_1+1}} \pi(\mathcal{I}, \boldsymbol{\omega}, \mathbf{L}|r, \alpha, \lambda)\pi(r)\pi(\alpha)\pi(\lambda)\, dL_{1,n_1+1}\, d\lambda \\
&= \prod_{j=1}^{n_1} \left(\frac{r + \sum_{\mathbf{x} \in \mathcal{I}_{i,j-1}} F_\alpha(\mathbf{x} - \mathbf{y}_{\omega(\mathbf{i},\mathbf{j})})}{h_{i,j}(\alpha, r, \boldsymbol{\omega})} \right) \prod_{j=1}^{n_1} \exp(-L_{i,j})\pi(r)\pi(\alpha) \\
&\quad \times \int_\lambda \int_{L_{1,n_1+1}} \exp(-L_{1,n_1+1}) 1_{\{U_1\}}\, dL_{1,n_1+1}\, d\lambda \\
&= K \int_\lambda \int_{L_{1,n_1+1}} \exp(-L_{1,n_1+1}) 1_{\{U_1\}}\, dL_{1,n_1+1}\, d\lambda, \qquad \text{say.} \qquad (4.5)
\end{aligned}$$

Given λ, for $1_{\{U_1\}}$ to be equal to 1, we require that

$$L_{1,(n_1+1)} > \lambda h_{1,n+1}(\alpha, r, \boldsymbol{\omega}) \left(t_1 - \frac{1}{\lambda} \sum_{j=1}^{n_1} \frac{1}{h_{1,j}(\alpha, r, \boldsymbol{\omega})} L_{1,j} \right) > 0. \qquad (4.6)$$

Therefore it follows from (4.5) with

$$\mathcal{C} = \left[\lambda h_{1,n+1}(\alpha, r, \boldsymbol{\omega}) \left(t_1 - \frac{1}{\lambda} \sum_{j=1}^{n_1} \frac{1}{h_{i,j}(\alpha, r, \boldsymbol{\omega})} L_{1,j} \right), \infty \right)$$

that

$$\begin{aligned}
& \pi(\boldsymbol{\omega}, \check{\mathbf{L}}^1, r, \alpha|\mathcal{I}) \\
\propto\ & K \int_\lambda 1_{\{\lambda > \sum_{j=1}^{n_1} \frac{1}{h_{1,j}(\alpha, r, \boldsymbol{\omega})} L_{1,j}/t_1\}} \int_{L_{1,n_1+1} \in \mathcal{C}} \exp(-L_{1,n_1+1})\, dL_{1,n_1+1}\, d\lambda \\
=\ & K \int_\lambda 1_{\{\lambda > \sum_{j=1}^{n_1} \frac{1}{h_{1,j}(\alpha, r, \boldsymbol{\omega})} L_{1,j}/t_1\}} \exp(-\lambda h_{1,t+1}(\alpha, r, \boldsymbol{\omega}) t_1) \\
& \times \exp\left(h_{1,n_1+1} \sum_{j=1}^{n_1} \frac{1}{h_{1,j}(\alpha, r, \boldsymbol{\omega})} L_{1,j} \right) d\lambda \\
=\ & K \exp\left(h_{1,n_1+1}(\alpha, r, \boldsymbol{\omega}) \sum_{j=1}^{n_1} \frac{1}{h_{1,j}(\alpha, r, \boldsymbol{\omega})} L_{1,j} \right) \\
& \times \left[-\frac{1}{h_{1,n_1+1}(\alpha, r, \boldsymbol{\omega}) t} \exp\left(-\lambda h_{1,n_1+1}(\alpha, r, \boldsymbol{\omega}) t_1 \right) \right]_{\sum_{j=1}^{n_1} \frac{1}{h_{1,j}(\alpha, r, \boldsymbol{\omega})} L_{1,j}/t_1}^{\infty} \\
=\ & \prod_{j=1}^{n_1} \left(\frac{r + \sum_{\mathbf{x} \in \mathcal{I}_{1,j-1}} F_\alpha(\mathbf{x} - \mathbf{y}_{\omega(\mathbf{1},\mathbf{j})})}{h_{1,j}(\alpha, r, \boldsymbol{\omega})} \right) \prod_{j=1}^{n_1} \exp(-L_{1,j})\pi(r)\pi(\alpha) \frac{1}{h_{1,n_1+1}(\alpha, r, \boldsymbol{\omega}) t_1}.
\end{aligned} \qquad (4.7)$$

Note that it follows from the third line of (4.7) that

$$\lambda|\boldsymbol{\omega}, \check{\mathbf{L}}^1, r, \alpha, \mathcal{I} \sim \sum_{j=1}^{n_1} \frac{L_{1,j}}{h_{1,j}(\alpha, r, \boldsymbol{\omega}) t_1} + \text{Exp}(h_{1,n_1+1}(\alpha, r, \boldsymbol{\omega}) t_1). \qquad (4.8)$$

Furthermore in (4.7), the only term involving $\check{\mathbf{L}}^1$ is $\prod_{j=1}^{n_1}\exp(-L_{1,j})$. Therefore integrating out $\check{\mathbf{L}}^1$ yields,

$$\pi(\boldsymbol{\omega}, r, \alpha|\mathcal{I}) \propto \prod_{j=1}^{n_1}\left(\frac{r+\sum_{\mathbf{x}\in\mathcal{I}_{1,j-1}}F_\alpha(\mathbf{x}-\mathbf{y}_{\omega(\mathbf{1},\mathbf{j})})}{h_{1,j}(\alpha,r,\boldsymbol{\omega})}\right)\pi(r)\pi(\alpha) \times\frac{1}{h_{1,n_1+1}(\alpha,r,\boldsymbol{\omega})}. \tag{4.9}$$

We observe that (4.9) differs slightly to $\pi(\boldsymbol{\omega}, r, \alpha|\mathcal{I})$ given in Gibson (1997a) including the final term $1/h_{1,n_1+1}(\alpha, r, \boldsymbol{\omega})$. The difference is due to the slightly different scenarios considered. In Gibson (1997a), the posterior is derived based upon, from time 0 the next n_1 infections being the set of individuals $\mathcal{W}_1$, regardless of the time taken. Since we are explicitly taking into account time, as well as requiring the next n_1 infections to be the set of individuals $\mathcal{W}_1$, we also require that no more infections take place. Ignoring this leads to a slight bias in the estimation of α and r. The larger $h_{1,n_1+1}(\alpha, r, \boldsymbol{\omega})$ is, the smaller the range of λ values (combined with $\mathbf{L}^1$) which are consistent with exactly n_1 infections taking place in the time interval.

For $m \geq 1$, we can update $\boldsymbol{\omega}$ efficiently using a scheme based upon Gibson (1997a). Specifically, we propose to switch $(\omega_{i,j}, \omega_{i,j+1})$ in a systematic manner. The key observation is that the changes in the likelihood involved with switching $(\omega_{i,j}, \omega_{i,j+1})$, only depends upon who has been infected prior to the j^{th} infection in time interval i and not the order in which they were infected. Furthermore, the order of infection of individuals $(\omega_{i,j}, \omega_{i,j+1})$ has no effect on any subsequent infections. Therefore for all i, we can consider the switching the orders of infected pairs

$$(\omega_{i,1}, \omega_{i,2}), (\omega_{i,3}, \omega_{i,4}), \ldots, (\omega_{i,2m_i-1}, \omega_{i,2m_i}), \tag{4.10}$$

in parallel, where m_i is the largest integer such that $2m_i \leq n_i$. Also for all i, the switches

$$(\omega_{i,2}, \omega_{i,3}), (\omega_{i,4}, \omega_{i,5}), \ldots, (\omega_{i,2k_i}, \omega_{i,2k_i+1}) \tag{4.11}$$

can be considered in parallel, where k_i is the largest integer such that $2k_i + 1 \leq n_i$.

We are now in position to outline an iteration of the MCMC algorithm for $m = 1$ with (r, α, λ) stored at the end of each iteration. The extension to $m > 1$ is given in the supplementary material.

MCMC algorithm

i) Update (r, α) using random walk Metropolis with a bivariate Gaussian proposal. Use (4.9) to compute the acceptance probability.

ii) Update $\boldsymbol{\omega}$.

(a) For $l = 1, 2, \ldots, m_1$, propose to switch the order of infection for $(\omega_{1,2l-1}, \omega_{1,2l})$, sequentially or preferably in parallel.

(b) For $l = 1, 2, \ldots, k_1$, propose to switch the order of infection for $(\omega_{1,2l}, \omega_{1,2l+1})$, sequentially or preferably in parallel.

iii) Sample λ. Simply draw a new set $\check{\mathbf{L}}^1$ and then sample $\lambda|\boldsymbol{\omega}, \check{\mathbf{L}}^1, r, \alpha, \mathcal{I}$ from its conditional distribution given by (4.8).

iv) Store (λ, r, α) as a sample from $\pi(\boldsymbol{\theta}|\mathbf{x})$.

Finally, before implementing the MCMC algorithm, we compare our approach to Brown *et al.* (2014), where estimation of $\mu = \lambda r$, α and $\lambda (= \beta)$ is performed. In Brown *et al.* (2014), the infection times $\boldsymbol{\tau}$ of the plants are imputed rather than the order of infection and the inter-infection time intervals. Note that $\boldsymbol{\tau}$ can easily be obtained from λ, $\boldsymbol{\omega}$ and $\check{\mathbf{L}}$ and visa-versa. The main disadvantage of the MCMC algorithm of Brown *et al.* (2014) is that updating an element of $\boldsymbol{\tau}$ has an effect on many components of the likelihood (all subsequent infection times), whereas the update $\boldsymbol{\omega}$ can, as noted above, be done efficiently and quickly by switching the orders of infection.

We now employ the MCMC algorithm to estimate the parameters for the CTV data and the SCYLV data. The CTV data introduced in Marcus *et al.* (1984) and analysed in Gibson (1997a) is used to illustrate the methodology with $m = 1$, whilst the main emphasis is on the longitudinal analysis of the SCYLV data experiment in Daugrois *et al.* (2011), analysed in Brown *et al.* (2014). The CTV data consist of the locations of infected trees in a citrus orchard at two time points, 1 year apart. There are 131 infected trees discovered at the first time point, 1981 with a further 45 trees infected by the second time point, 1982. In Gibson (1997a), there is assumed to be no background infection ($r = 0$) and two choices of $F_\alpha(\mathbf{x} - \mathbf{y})$ are considered; (a) $\exp(-\alpha|\mathbf{y} - \mathbf{x}|)$ and (b) $|\mathbf{y} - \mathbf{x}|^{-2\alpha}$ with α estimated on the basis of the 45 infections between the two observed time points. For both cases we estimated α and λ based on 50,000 iterations following a burn-in of 2,000 iterations. The posterior means (standard deviations) of α and λ are, for case (a), 0.269 (0.0678) and 0.168 (0.113), respectively, and for case (b), 1.32 (0.158) and 3.05 (2.54), respectively. The estimates for α are consistent with those presented in Gibson (1997a) where a discrete set of α values rather than the continuous model presented here are used. It should be noted that the non-collapsed MCMC algorithm was not a practical alternative to the collapsed algorithm as the key condition (4.3) was often violated, and proposed moves rejected, for all but very small moves in $\boldsymbol{\theta} = (\lambda, r, \alpha)$ and $\mathbf{L}$, even with longer runs of 10^6 iterations.

The estimation of λ combined with α allows for the prediction of the number and location of new cases

in the forthcoming year. However, we use knowledge of λ to estimate the time of the introductory case of CTV into the orchard. We follow Gibson (1997a) by modelling the first 176 infectious cases, assuming that there is a single introductory case, whose location is unknown but is one of the 131 locations infected before the end of 1981. The parameter α is estimated based on the full data, whereas the 45 infections occurring between the two observation points allow us to estimate λ. Then for $i = 2, 3, \ldots, 131$, we can simulate the length of time, X_i^0, between the $(i-1)^{st}$ and i^{th} infection with $X_i^0 \sim \text{Exp}(\lambda \sum_{x \in \mathcal{I}_{0,i-1}} \sum_{y \in \mathcal{S}_{0,i-1}} F_\alpha(\mathbf{x} - \mathbf{y}))$. Thus $T_0 = \sum_{i=2}^{131} X_i^0$ represents an estimate of the time of the introductory case prior to the first time point (1981). Using $F_\alpha(\mathbf{x} - \mathbf{y}) = \exp(-\alpha|\mathbf{y} - \mathbf{x}|)$, the posterior means (standard deviations) of α and λ are 0.1071 (0.0141) and 0.0192 (0.0061), respectively. The estimate of α is again consistent with corresponding analysis in Gibson (1997a). However, unlike Gibson (1997a), we can estimate T_0 which is found to have a posterior mean of 15.3 years with a standard deviation of 4.5 years.

In Brown *et al.* (2014) trial B from Daugrois *et al.* (2011) is analysed. Specifically, Daugrois *et al.* (2011) details four trials of the spread of SCYLV in trial plantations which are initially disease-free. The SCYLV is transmitted from plant to plant via aphids. The infectious status of all 1742 plants in trial B were recorded at weeks 0,6,11,15,19,23 and 30. As noted above this differs slightly from Brown *et al.* (2014) but this difference has very little effect on parameter estimates. We follow Brown *et al.* (2014) in taking $F_\alpha(\mathbf{x}-\mathbf{y}) = \frac{1}{2\pi\alpha^2} \exp(-|\mathbf{x}-\mathbf{y}|^2/(2\alpha^2))$, a Gaussian transmission kernel proposed in Brown *et al.* (2014) to capture the diffusive movement of aphids. In Brown *et al.* (2014) μ is used to denote the background risk of infection and this corresponds to $r \times \lambda$ in our parameterisation. We estimated the model parameters based on both the full data and a single time interval, the process observed at weeks 0 and 30 only. We choose improper, uniform priors for the parameters in contrast to the weak informative priors chosen in Brown *et al.* (2014). The algorithms were run for 110,000 iterations with the first 10,000 iterations discarded as burn-in. The proposal standard deviations for the random walk updates of r and α are 0.001 and 0.02, respectively, for the full data and 0.01 and 0.2, respectively for the single time interval. For the full data, we repeated the process of proposing to update two randomly chosen L values ten times per iteration. The time taken to run the algorithm was similar in both cases taking just under a day on a 8 core processor, which even without parallelisation is comparable with results given in Brown *et al.* (2014). Based on the full data we obtain 0.0025 (0.00036), 1.19 (0.100) and 0.115 (0.0061) for the posterior means and standard deviations of the parameters $\mu(= r\lambda)$, α and λ, respectively. These parameter estimates are comparable with those presented in Brown *et al.* (2014). For the single time interval data, we obtain 0.0059 (0.0008), 0.777 (0.075) and 0.0624 (0.0074) for the posterior means and standard deviations of the parameters $\mu(= r\lambda)$, α and λ, respectively. Therefore there is a considerable disparity in the parameter

estimates based on the full data set as opposed to taking it as a single time interval with a clear loss of information from not taking into account the temporal spread of the disease.

There are a couple of observations to make about the analysis of the data. Firstly, in the switching step for updating $\boldsymbol{\omega}$ the proportion of moves accepted was over 99% for all positions over both data sets. This corresponds to their being little information in the data concerning the order of infection and is consistent with the plots of the density of infection times in the first and last intervals presented in Brown *et al.* (2014), Figure 3. We applied the MCMC algorithm to a small simulated data set presented in Gibson (1997a), Figure 3 with one initial infective and nine subsequent infections. Even in this case with clear chains of infection the acceptance rates for all positions in updating $\boldsymbol{\omega}$ were over 75%, increasing to 92% for the final position. This all suggests that, if a reasonable initial choice of $\boldsymbol{\omega}$ is made, decent parameter estimates can be obtained even if $\boldsymbol{\omega}$ is kept fixed (not updated). We proposed a fixed configuration for $\boldsymbol{\omega}$ chosen using $r = 0.1$ and $\alpha = 1.0$ and then estimated the parameters μ, α and λ using a MCMC run of 110,000 iterations with the first 10,000 iterations discarded as burn-in and $\boldsymbol{\omega}$ kept fixed. The algorithm ran approximately four times as fast without the update of the $\boldsymbol{\omega}$ and we obtained 0.0026 (0.00035), 1.17 (0.091) and 0.116 (0.0060) for the posterior means and standard deviations of the parameters $\mu(= r\lambda)$, α and λ, respectively. Thus we obtain a reasonable approximation of the posterior distribution in this case. Similar results were obtained starting from $r = 0.001$ and $\alpha = 2.0$, details along with plots are presented in the supplementary material. Secondly, the algorithm performs very well for a single time interval with mixing worsening as the number of time intervals increases. The main reason for the algorithm performing particularly well for a single time interval is the ability to full integrate out $\check{\mathbf{L}}$ as evidenced by the considerably larger proposal standard deviations used for a single time interval as opposed to that used for the full data.

5 Conclusions

In this paper we have shown how non-centred parameterisations can be exploited to construct and then collapse MCMC algorithms for epidemic models. Whilst the epidemic models and corresponding data sets in the two examples appear quite different there are common features which are exploited. The key feature is that in both cases an infectious individual, whilst infectious, makes infectious contacts at the points of a homogeneous Poisson point process. This is often the case and the approach taken here could be relaxed to inhomogeneous Poisson point processes, see, for example, Jewell *et al.* (2009) for an epidemic example. Continuous time models which do not assume a Poisson point process for the infectious process are rare, see Streftaris and Gibson (2012) for an exception. The superposition and decomposition of

Poisson processes gives great flexibility in modelling the infectious process and in particular, leads to the exponential tolerances, $\mathbf{L}$ exploited in both examples in this paper. Therefore the data augmentations and collapsing developed in this paper should be applicable to a wider class of epidemic models than the two presented here.

More generally, it is the Markov property of the Poisson process which is particularly useful here. In Neal (2014) a simple reparameterisation of Markov processes is exploited to allow coupled simulations of a Markov process. Specifically, if $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ represent rate parameters of a Markov process, then we can reparameterise by setting $\lambda = \theta_p$ and $\boldsymbol{\phi} = (\theta_1/\theta_p, \theta_2/\theta_p, \ldots, \theta_{p-1}/\theta_p)$. Thus the components of $\boldsymbol{\phi}$ are the relative rates of parameters $\boldsymbol{\theta}$ to $\lambda = \theta_p$. For a Markov process it is $\boldsymbol{\phi}$ and the state of the system which determine the probability that a given transition takes place, whereas λ determines the inter-arrival time between events. This is similar to the epidemic examples studied here, in that, the infection rate λ determines the additional infectious pressure required for the next infection and the other parameters and state of the population that determine who gets infected.

A simple example of a Markov process is the SIS epidemic model. Suppose that we have a population of size n and let $Y(t)$ denote the total number of infectives at time t. Since individuals can only be susceptible or infectious, $Y(t)$ defines the state of the system with $n - Y(t)$ being the number the number of susceptibles. Let infectious individuals make infectious contacts at the points of a homogeneous Poisson point process with rate β and recover at rate γ (exponentially distributed infectious periods with mean $1/\gamma$). Let $\lambda = \gamma$ and $\phi = \beta/\gamma$. Then at time t, the time until the next event (infection or recovery) is $\text{Exp}(\beta Y(t)(n - Y(t))/n + \gamma Y(t)) = \text{Exp}(\phi Y(t)(n - Y(t))/n + Y(t))/\lambda$ and the probability that the next event is an infection is

$$\frac{\beta Y(t)(n - Y(t))/n}{\beta Y(t)(n - Y(t))/n + \gamma Y(t)} = \frac{\phi(n - Y(t))/n}{\phi(n - Y(t))/n + 1}. \tag{5.1}$$

Finally, in this paper we have exploited a Sellke threshold and the order of infections to construct efficient MCMC algorithms for epidemic models. This approach could also be employed for temporal epidemic data, for example, in situations where the removal times but not the infection times are observed (Neal and Roberts (2005),Xiang and Neal (2014)). Specifically, we could augment the data with $\boldsymbol{\omega}$ and $\mathbf{L}$, which will define the infection times of individuals, and hence $\mathbf{I}$ will follow. More detailed inference on the parameters of the distribution of I are then possible and we can remove the restriction $E[I] = 1$ as the distribution is now identifiable. This alternative approach for temporal data could be a promising approach for epidemics in progress as an effective way of updating the number of unobserved infections as existing algorithms (O'Neill and Roberts (1999),Jewell *et al.* (2009)) add or remove one unobserved

infection time at a time leading to slow exploration of the space of the number of unobserved infectives.

Acknowledgements

This work was supported by EPSRC grant, EP/J008443/1. We would like to thank Patrick Brown and Jean Heinrich Daugrois for providing access to and guidance on the sugarcane yellow leaf virus data set. We would like to thank two anonymous referees' and an associate editor for their comments which have assisted in revising the paper.

Supporting information. Additional supporting information may be found in the online version of this article at the publishers website.

References

Ball, F. and O'Neill, P. (1999) The distribution of general final state random variables for stochastic epidemic models. *J. Appl. Probab.*, **36**, 473–491.

van Boven, M., Kretzschmar, M., Wallinga, J., O'Neill, P.D., Wichmann, O. and Hahné, S. (2010) Estimation of measles vaccine efficacy and critical vaccination coverage in a highly vaccinated population. Journal of the Royal Society Interface, **7**, 1537–1544.

Brown, P.E, Chimard, F., Remorov, A., Rosenthal, J.S. and Wang, X. (2014) Statistical Inference and Computational Efficiency for Spatial Infectious-Disease Models with Plantation Data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **63**, 467–482.

Daugrois, J.H., Edon-Jock, C., Bonoto, S., Vaillant, J. and Rott, P. (2011) Spread of Sugarcane yellow leaf virus in initially disease-free sugarcane is linked to rainfall and host resistance in the humid tropical environment of Guadeloupe. *European Journal of Plant Pathology* **129**, 71-80.

Gibson, G.J. (1997a) Markov Chain Monte Carlo Methods for Fitting Spatiotemporal Stochastic Models in Plant Epidemiology. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **46**, 215–233.

Gibson, G.J. (1997b) Investigating mechanisms of spatiotemporal epidemic spread using stochastic models. *American Phytopathological Society* **87**, 139–146.

Halloran, M. E., Longini, I. M. and Struchiner, C. J. (2010) *Design and analysis of vaccine studies.* Springer, New York. Springer.

Jewell, C.P., Kypraios, T., Neal, P.J. and Roberts, G.O. (2009) Bayesian Analysis for Emerging Infectious Diseases. *Bayesian Analysis* **4**, 465–496.

Liu, J.S. (1994) The collased Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *J. Amer. Statist. Assoc.* **89**, 958–966.

Marcus, R., Svetlana, F., Talpaz, H., Salomon, R. and Bar-Joseph, M. (1984) On the spatial distribution of citrus tristeza virus disease. *Phytoparasitica* **12**, 45–52.

Neal, P. (2012) Efficient likelihood-free Bayesian Computation for household epidemics. *Statist. Comput.*, **22**, 1239–1256.

Neal, P. (2014) Simulation based sequential Monte Carlo methods for discretely observed Markov processes. arXiv:1404.4185

Neal, P.J. and Huang, C.L.T. (2015) Forward simulation MCMC with applications to stochastic epidemic models. *Scand. J. Statist.* **42**, 378–396.

Neal, P.J. and Roberts, G.O. (2005) A case study in non-centering for data augmentation: Stochastic epidemics. *Statist. Comput.* **15**, 315–327.

Neal, P.J., Roberts, G.O. and Yuen, W.K. (2012) Optimal Scaling of Random Walk Metropolisalgorithms with discontinuous target densities. *Ann. Appl. Probab.* **22**, 1880–1927

O'Neill, P.D. and Roberts, G.O. (1999). Bayesian inference for partially observed stochastic epidemics. *J. Roy. Statist. Soc. Ser. A* **162**, 121–129.

Papaspiliopoulos, O., Roberts, G.O. and Sköld, M. (2003) Non-centered parameterisations for hierarchical models and data augmentation. *Bayesian Statistics 7* (J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West, eds.) Oxford University Press, 307–326.

Paunio, M., Peltola, H., Valle, M., Davidkin, I., Virtanen, M. and Heinonen, O. (1998) Explosive school-based measles outbreak: Intense exposure may have resulted in high risk, even among revaccinees *American Journal of Epidemiology*, **148**, 1103–1110.

Sellke, T. (1983) On the asymptotic distribution of the size of a stochastic epidemic. *J. Appl. Probab.* **20**, 390–394.

Smith, A.F.M. and Roberts, G.O. (1993) Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **55**, 3–23.

Streftaris, G. and Gibson, G.J. (2012) Non-exponential tolerance to infection in epidemic systems - modeling, inference and assessment. *Biostatistics* **13**, 580–593.

Tanaka, M. M., Francis, A. R., Luciani, F. and Sisson, S. A. (2006) Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics* **173**, 1511–1520.

Tavaré, S., Balding, D.J., Griffiths, R.C. and Donnelly, P. (1997) Inferring coalescence times from DNA sequence data. *Genetics*, **145**, 505–518.

Xiang, F. and Neal, P. (2014) Efficient MCMC for temporal epidemics via parameter reduction. *Comput. Statist. Data Anal.* **80**, 240–250.

Corresponding author's address:

Fylde College

Lancaster University

Lancaster

LA1 4YF

United Kingdom

email address: p.neal@lancaster.ac.uk

Type, k	0	1	2
Total number of infected individuals, x_k	18	11	6
Total number of individuals, n_k	79	189	149

Table 1. Honkajoki measles data set.

Scenario	Parameter	λ	q_1	q_2	γ
1	Mean	2.75	0.308	0.225	-
$I \equiv 1$	St. Dev.	0.947	0.121	0.107	-
Collapsed	Eff. sample size	8285	8453	9376	-
Non-collapsed	Eff. sample size	166	332	512	-
2	Mean	2.94	0.284	0.209	-
$I \sim \text{Gamma}(2,2)$	St. Dev.	1.097	0.114	0.100	-
Collapsed	Eff. sample size	13712	8991	9615	-
Non-collapsed	Eff. sample size	141	324	436	-
3	Mean	2.96	0.284	0.205	1.055
$I \sim \text{Gamma}(\exp(\gamma), \exp(\gamma))$	St. Dev.	1.286	0.113	0.096	1.044
Collapsed	Eff. sample size	10547	9156	10220	216
Non-collapsed	Eff. sample size	158	332	536	146

Table 2. Estimated posterior means, standard deviations and effective sample size (collapsed and non-collapsed) λ, $\mathbf{q}$ and γ, for the three scenarios 1) $I \equiv 1$; 2) $I \sim \text{Gamma}(2,2)$ and 3) $I \sim \text{Gamma}(\exp(\gamma), \exp(\gamma))$.