Statistics and Computing **15**: 315–327, 2005

A case study in non-centering for data augmentation: Stochastic epidemics

PETER NEAL* and GARETH ROBERTS†

**Mathematics Department, UMIST, P.O. Box 88, Manchester, M60 1QD, UK*
P.Neal-2@umist.ac.uk
†*Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, UK*
g.o.roberts@lancaster.ac.uk

Received May 2003 and accepted July 2005

In this paper, we introduce non-centered and partially non-centered MCMC algorithms for stochastic epidemic models. Centered algorithms previously considered in the literature perform adequately well for small data sets. However, due to the high dependence inherent in the models between the missing data and the parameters, the performance of the centered algorithms gets appreciably worse when larger data sets are considered. Therefore non-centered and partially non-centered algorithms are introduced and are shown to out perform the existing centered algorithms.

Keywords: stochastic epidemic models, bernoulli random graphs, non-centered and partially non-centered MCMC algorithms, data augmentation

1. Introduction

Non-centering parameterisations and their corresponding MCMC algorithms were introduced in Papaspiliopoulos *et al.* (2003). They provide effective alternatives to hierarchical or centered parameterisations particularly when observed data is relatively uninformative about missing data. Motivated by the need to develop compromises between centered and non-centered methods, that paper also introduces partially non-centered parameterisations, a continuum of parameterisations with centered and non-centred at the extremes. In the hierarchical Gaussian linear model case, exact theoretical analysis provides strong support for the use of partial non-centering. Furthermore, in much more realistic problems with similar hierarchical structure, empirical evidence also supports the use of the methodology (see, for example, the spatial GLMM methodology of Christensen *et al.* (2003)).

However non-centering methodology remains largely untested on more complex, structured missing data problems. The aim of this paper is to investigate non-centering methods on two such problems arising from inference for partially observed stochastic epidemics. These applications are particularly appropriate for two reasons. Firstly, practical use of these models on real data is inevitably complicated by considerable missing data. Secondly, datasets for epidemics can be very large (for example, the foot and mouth epidemic, see Keeling *et al.* (2001)) and MCMC methods are well-known to deteriorate for these models for large populations, as dependence between parameters and missing data grows. Furthermore, we see this fundamental work on parameterisations for epidemic models as being of considerable applied interest, and applications of this work in modelling the foot and mouth and other livestock epidemics and the spread of aquatic disease between UK fish farms, are currently the subject of ongoing work by the authors.

Many stochastic epidemic models are by now well-understood from a probabilistic view point (see, for example, Bailey (1975), Ball *et al.* (1997) and Andersson (1999)). They offer plausible and relatively parsimonious stochastic models for the macroscopic progress of an epidemic, which in turn lead to considerably more realistic models for epidemic final size distributions than their deterministic model counterparts. However, inference for these models is normally complicated by the fact that the data is only partially observed. Therefore it is often difficult to write down the likelihood for the observed data, necessitating the use of complicated data augmentation schemes. However, it is well-known that MCMC algorithms which involve the imputation of large sets of missing data can converge very slowly (see, for example, Meng and van Dyk (1997)).

This paper is concerned with epidemics where limited temporal data is available. Specifically, we assume that removal times

of infective individuals are observed. Thus the times when infections occur are not observed and need to be imputed. Typically this is the case where very limited knowledge about infected individuals is available. Often the data consists solely of the time at which the first (or last) signs of the disease upon the individual are noted. Where more informative information is available, the inference often becomes much more straightforward, see, for example, Neal and Roberts (2004).

We shall apply the *non-centered* methodology of Papaspiliopoulos *et al.* (2003) to two models with observed removal times: the *general stochastic epidemic* (GSE) (and its extensions) and the *Bernoulli random graph general stochastic epidemic* (BGSE). For both these models, we find that non-centering techniques assist the MCMC mixing with minimal extra computational costs. The BGSE example involves the use of a state-space expansion technique. However, even with the computational costs taken into account, the non-centered algorithms are shown to out perform the existing centered algorithms over a range of examples.

The terminology centered and non-centered parameterisations is borrowed from the hierarchical modelling literature, and a fuller description can be found in Papaspiliopoulos *et al.* (2003). The terms centered and non-centered parameterisations refer to the mechanisms by which the missing data (infection times) and the model parameters are updated, a detailed description of which will be given in Section 4. The connection between the relative performance of centered and non-centered methods and statistical information is well-understood in simpler models (Meng and van Dyk (1997), Papaspiliopoulos *et al.* (2003)). Here we are able to exploit the information gained from these simpler models, in order to design samplers that are robust to the information content in the specific data set observed. In particular, we introduce a novel partially non-centered parameterisation which is shown to be particularly effective in Sections 5 and 6. Furthermore, the partially non-centered parameterisation we introduce is very natural and was seen to out perform a partially non-centered parameterisation based on the approach taken in Papaspiliopoulos *et al.* (2003), Section 4.

The paper is structured as follows. We briefly outline non-centering methods in general in Section 2. Then in Section 3, the epidemic models are described, and the non-centering methods we shall be using on these models are introduced in Section 4. The GSE and its extensions are considered in Sections 5 and 6, respectively, and the benefits of the partially non-centered approach are demonstrated. In Section 7, we turn our attention to the BGSE where the non-centered algorithm is particularly effective. Finally in Section 8, we make some concluding remarks and outline some possible avenues for future research.

2. Centering and non-centering

In this short section we shall briefly describe and motivate non-centering in a general context. This description follows the work of Papaspiliopoulos *et al.* (2003).

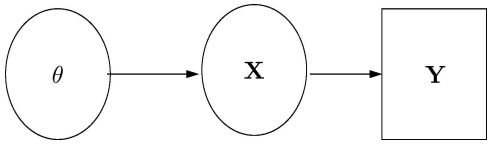

Fig. 1. *A hierarchical model and its centered parameterisation*

A centered parameterisation of a hierarchical model essentially parameterises in terms of the different layers of the hierarchy. For example, consider the hierarchy given by the graphical model in Fig. 1. Here we assume that θ is a parameter of interest and that $\mathbf{X}$ is a vector of missing data, with $\mathbf{Y}$ being the observed data.

The obvious centered parameterisation therefore is $(\theta, \mathbf{X})$. It has the advantages that it is statistically natural and the inherent conditional independence (e.g. $\theta \perp \mathbf{Y} \mid \mathbf{X}$) makes the implementation of appropriate MCMC algorithms simpler. On the other hand, if $\mathbf{Y}$ contains little information about $\mathbf{X}$, then the apriori dependence in the model implied by the link between θ and $\mathbf{X}$ is likely to lead to very poor convergence of MCMC. This problem is exacerbated when $\mathbf{X}$ is high-dimensional and the dependence between θ and $\mathbf{X}$ can then be particularly strong. In a limited set of simple examples such as the hierarchical Gaussian linear model, this effect can be analysed and quantified explicitly.

A natural alternative suggested by this discussion is to reparameterise $\tilde{\mathbf{X}} = h(X, \theta)$ in order to make θ and $\tilde{\mathbf{X}}$ independent. We call the parameterisation $(\theta, \tilde{\mathbf{X}})$ a non-centered parameterisation. Whilst this parameterisation gives independent components in the absence of data, informative data renders $\tilde{\mathbf{X}}$ and θ highly dependent. Thus its dependence properties are in some sense opposite to those of the centered parameterisation.

It should be noted that identification of the function h above can be rather complicated in many situations. Furthermore, there is clearly no unique non-centered parameterisation even for the simple hierarchical structure described by Fig. 1. Moreover, for more complex hierarchical structures (such as those with time-series structure as in the epidemic models of this paper) centering and non-centering can often only be described for certain parameters. However, the basic principle of non-centering remains the same—the breaking of an apriori dependence between parameter(s) and missing data.

For the purposes of MCMC algorithms, for instance for the hierarchical model in Fig. 1, weak data suggests the use of non-centered algorithms, whereas, strong data suggests using the centered specification. Therefore for most datasets, one would suspect that an improved parameterisation on both these extremes might be available as some kind of compromise between the two. This is the motivation behind partial non-centering. The idea is to define a continuum of parameterisations with the centered and non-centered as extremes. In the hierarchical Gaussian model, the benefits of partial non-centering methods can be investigated explicitly, and empirical evidence from other examples (see, for example, Christensen *et al.* (2003)) strongly support the use of the methodology.

3. The epidemic models

Both the GSE and the BGSE are examples of the most widely studied class of epidemic models, namely SIR epidemics (see, for example, Bailey (1975), Andersson (1999), O'Neill and Roberts (1999), and O'Neill and Becker (2001)). That is, at any point in time an individual is in one of three states; susceptible, infected or removed. The only transitions in state which we allow are; from susceptible to infected and; from infected to removed. The terms susceptible and infected are self-explanatory, in that, susceptible individuals are those who don't have the disease but are susceptible to infection, whilst infected individuals are infectious with the disease and are able to infect susceptible individuals. We assume that at the end of an individual's infectious period he becomes removed, either by death or recovery followed by immunity to further infection. Thus the individual may well remain a member of the community but is removed in the sense that he plays no further part in the epidemic process. We shall further assume that the population $\mathcal{N}$ of N individuals is closed and that there are initially a infectious individuals whilst the rest of the population is initially susceptible. Note that any individuals who are initially in the removed state play no part in the epidemic and can therefore be ignored. The assumption that the population is closed is reasonable, in that, many epidemics occur over relatively short time periods when compared with usual population demographics.

The infectious life history $(Q_i, \{W_{ij}; 1 \leq j \leq N\})$ of an infective, i say, comprises the length of individual i's infectious period, Q_i, and W_{ij} $(1 \leq j \leq N)$ the point of time, relative to individual i's infection, at which he makes an infectious contact with individual j.

For the GSE, the $\{Q_i; 1 \leq i \leq N\}$ are independent and identically distributed according to $Q \sim Exp(\gamma)$ and the $\{W_{ij}; 1 \leq i, j \leq N\}$ are independent and identically distributed according to $W \sim Exp(\beta)$. The epidemic's course can be constructed from $(Q_i, \{W_{ij}; 1 \leq j \leq N\})$ $(1 \leq i \leq N)$ as follows. Start with the initial infectives infectious at time 0. Then let I_i denote the time at which individual i becomes infected, then $R_i = I_i + Q_i$ denotes the time at which individual i becomes removed. If $W_{ij} < Q_i$, individual i makes an infectious contact with individual j at time $I_i + W_{ij}$. If individual j is susceptible at time $I_i + W_{ij}$, individual j becomes infected, otherwise nothing happens. We continue the above process until the epidemic has ceased, i.e. there are no infectives remaining in the population. In Section 6, we study an extension of the GSE where Q is an arbitrary, but specified non-negative random variable.

For our second example, we shall consider the BGSE. Here let $\mathcal{G}$ be a Bernoulli random graph on N vertices, that is, for each pair of vertices i and j there exists, independently of the remainder of the graph, an edge between vertices i and j with probability, p say. For $1 \leq i, j \leq N$, let $G_{ij} = 1$, if there exists an edge between vertices i and j in the graph $\mathcal{G}$ and $G_{ij} = 0$, otherwise. We employ the convention that $G_{ii} = 0$ $(1 \leq i \leq N)$. We then say that individuals i and j are acquaintances in the population $\mathcal{N}$ if there exists an edge between vertices i and j in the graph $\mathcal{G}$ (i.e. $G_{ij} = 1$). Therefore for the BGSE, the $\{Q_i; 1 \leq i \leq N\}$ are independent and identically distributed according to $Q \sim Exp(\gamma)$ and, the $\{W_{ij}; 1 \leq i, j \leq N\}$ are assumed to be independent with $W_{ij} \sim Exp(\beta)$, if $G_{ij} = 1$ and $W_{ij} = \infty$, otherwise.

Thus the GSE is a homogeneously mixing epidemic, while for the BGSE, an infective, i say, can only make infectious contacts with a subset of the population $\mathcal{N}$, namely his acquaintances $A_i = \{j; G_{ij} = 1\}$.

The model parameters of interest are β and γ for both the GSE and BGSE, and an additional parameter p for the BGSE. The temporal data are assumed to comprise just the removal times of each individual ultimately infected. For simplicity, we assume that the epidemic has ceased, so that the final size of the epidemic m is known. We also assume, for clarity in presentation of the results, that there is only one initial infective (i.e. $a = 1$). Let $0 = R_1 \leq R_2 \leq \cdots \leq R_m = T$ denote the observed removal times, and write $\mathbf{R} = (R_1, R_2, \ldots, R_m)$. Let $\mathbf{I} = (I_1, I_2, \ldots, I_m)$ denote the unobserved infection times, where for $1 \leq j \leq m$, the infection time I_j corresponds to the infection time of the individual removed at time R_j. Label the initial infective κ, so that $I_\kappa < I_j$ for all $j \neq \kappa$.

4. Centered and non-centered parameterisations for the GSE

Here, we shall introduce centered and non-centered parameterisations that we shall consider for the GSE. The prior dependence link to be broken by non-centering in our method will be that between $R_i - I_i$ and γ.

For the GSE, we have the observed data $\mathbf{R}$, the unobserved data $\mathbf{I}$ and the model parameters (β, γ). We can then implement the following (centered) MCMC algorithm (the details will be given in Section 5).

1. `Update the parameters` β `and` γ `from` $\pi(\beta \mid \gamma, \mathbf{I}, \mathbf{R})$ `and` $\pi(\gamma \mid \beta, \mathbf{I}, \mathbf{R})$ `with the Gibbs sampler, respectively.`
2. `Update one (or more) of` $\{I_i : 1 \leq i \leq m\}$ `using a Metropolis-Hastings step by proposing a replacement infection time` $R_i - I_i' \sim Exp(\gamma)$.

The centered parameterisation alternates between updating the model parameters (β, γ) and the missing data $\mathbf{I}$. By contrast, the non-centered parameterisation we introduce updates the model parameters and the missing data together. For $1 \leq i \leq m$, let $U_i = \gamma(R_i - I_i)$ (i.e. $I_i = R_i - \frac{1}{\gamma}U_i$). Note that apriori $U_i \sim Exp(1)$ $(1 \leq i \leq m)$. Then for the non-centered parameterisation we have a change in variables from $(\mathbf{R}, \mathbf{I}, \beta, \gamma)$ to $(\mathbf{R}, \mathbf{U}, \beta, \gamma)$. We can then implement the following (non-centered) MCMC algorithm.

1. `Update the parameters` β `from` $\pi(\beta|\gamma, \mathbf{I}, \mathbf{R})$ `using the Gibbs sampler.`

2. `Update` γ `using RWM (Random-walk Metropolis) with proposal distribution` $N(\gamma, \sigma_\gamma^2)$.
   (Note that updating γ involves updating the missing infection times **I**, since for $1 \le i \le m$, $I_i' = R_i - \frac{1}{\gamma'} U_i$.)
3. `Update one (or more) of` $\{U_i : 1 \le i \le m\}$ `using a Metropolis-Hastings step by proposing a replacement infection time` $U_i' \sim Exp(1)$.

Therefore the key difference between the two parameterisations is the following. For the centered algorithm the infection times **I** remain fixed when updating γ, whilst for the non-centered algorithm new infection times $\mathbf{I}'$ are proposed when updating γ. As mentioned in Section 2, it is shown in Papaspiliopoulos *et al.* (2003), using a simple hierarchical model as an example, that when there is strong (weak) dependence between the model parameters and the missing data then a non-centered (centered) parameterisation is preferable. For epidemic models the strength of dependence between the model parameters and the missing data is important in choosing which parameterisation to use. However, there are other serious considerations, in particular, it is important that the MCMC algorithm updates the missing data (**I** for the GSE and $(\mathbf{I}, \mathcal{G})$ for the BGSE) in an efficient manner. This suggests that a partially non-centered MCMC algorithm might be a sensible compromise, and this is particularly useful for the GSE. The strategy we employ for the partial non-centering parameterisation is as follows. At each iteration, we choose to center or non-center each infection period, independently, with probability μ, for some $0 \le \mu \le 1$. That is, when updating γ some of the infection times remain fixed whilst the other infection times change with γ.

For the BGSE, the centered parameterisation updates p and $\mathcal{G}$ separately. Whilst for the non-centered parameterisation the updating of p involves updating the graph $\mathcal{G}$ also.

5. General stochastic epidemic

5.1. *The partially non-centered parameterisations*

Let X_t and Y_t denote the number of susceptibles and infectives in the population at time t. Then

$$f(\mathbf{R}, \mathbf{I} \mid \beta, \gamma) \propto \prod_{i \neq \kappa} \{\beta Y_{I_i -}\} \exp\left(-\beta \int_{I_\kappa}^{T} X_t Y_t \, dt \right) \times \prod_{j=1}^{m} \{\gamma \exp(-\gamma(R_j - I_j))\}, \tag{5.1}$$

where the notation I_i- denotes the left hand limit, so for example, $Y_{I_i -}$ denotes $\lim_{s \uparrow I_i}(Y_s)$. Note that, at time, t say, there are Y_t infectives attempting to infect X_t susceptibles. Therefore $\int_{I_\kappa}^{T} X_t Y_t \, dt$ denotes the total number of person-to-person units of infectious pressure exerted during the course of the epidemic. Let $A = \int_{I_\kappa}^{T} X_t Y_t \, dt$, then we can rewrite $A = \sum_{j=1}^{m} \sum_{k=1}^{N} \{(R_j \wedge I_k) - (I_j \wedge I_k)\}$ with $I_j = \infty$ for $j = m+1, m+2, \ldots, N$. The likelihood (5.1) is in agreement with the first equation on page 101 of O'Neill and Becker (2001).

Due to the memoryless property of the exponential distribution, there is a natural alternative to the model setup described in Section 3. Instead of associating infection times with removal times, let $\mathbf{I} = (I_1, I_2, \ldots, I_m)$ denote the sequential order of times at which infections occur within the epidemic, so for $1 \le j < k \le m$, $I_j \le I_k$. The model setup thus has the likelihood function given by O'Neill and Roberts (1999), (3.1). We use the original model setup, since the association of each infection time with a removal time allows us to implement partially non-centered reparameterisations.

We follow O'Neill and Roberts (1999) in assuming that *a-priori* β and γ are independent. The prior distributions for β and γ are $Gam(\nu_\beta, \lambda_\beta)$ and $Gam(\nu_\gamma, \lambda_\gamma)$, respectively. It is then straightforward to integrate β and γ out of (5.1), giving

$$f(\mathbf{R}, \mathbf{I}) \propto \left\{ \prod_{i \neq \kappa} Y_{I_i -} \right\} (\lambda_\beta + A)^{-(m+\nu_\beta - 1)} \times \left(\lambda_\gamma + \sum_{i=1}^{m} (R_i - I_i) \right)^{-(m+\nu_\gamma)}. \tag{5.2}$$

We then have the following (centered) MCMC algorithm based on (5.2).

1. `Update the parameter` γ `from` $\pi(\gamma \mid \beta, \mathbf{I}, \mathbf{R})$ `using the conditional distribution.`
2. `Update one of` $\{I_i : 1 \le i \le m\}$ `using a Metropolis-Hastings step by proposing a replacement infection time` $R_i - I_i' \sim Exp(\gamma)$.

β values can be generated from the resultant sample of $(\mathbf{R}, \mathbf{I}, \gamma)$ values since $\pi(\beta \mid \gamma, \mathbf{R}, \mathbf{I}) \sim Gam(m + \nu_\beta - 1, \lambda_\beta + A)$.

Using the terminology of Papaspiliopoulos *et al.* (2003), we introduce a γ-partially-non-centered MCMC algorithm (γpNC algorithm). We partition the set of individuals whom are ultimately infected into two groups $\mathcal{C}$ and $\mathcal{U}$. Let $\mathbf{I}^{\mathcal{C}}$ and $\mathbf{I}^{\mathcal{U}}$ denote the infection times of the individuals in groups $\mathcal{C}$ and $\mathcal{U}$, respectively. Let $U_i = \gamma(R_i - I_i)$ $(i \in \mathcal{U})$ and we propose a change in variables from $(\mathbf{R}, \mathbf{I}^{\mathcal{C}}, \mathbf{I}^{\mathcal{U}}, \beta, \gamma)$ to $(\mathbf{R}, \mathbf{I}^{\mathcal{C}}, \mathbf{U}^{\mathcal{U}}, \beta, \gamma)$. Note that if $\mathcal{U} = \emptyset$, then we have the centered model described above. Let $\omega = |\mathcal{U}|$, then the Jacobian for the transformation is $\gamma^{-\omega}$. Using (5.1) and integrating out the parameter β, gives

$$\begin{aligned} & g(\mathbf{R}, \mathbf{I}^{\mathcal{C}}, \mathbf{U}^{\mathcal{U}}, \gamma) \\ & \propto \left\{ \prod_{i \neq \kappa} Y_{I_i -} \right\} (\lambda_\beta + A)^{-(m+\nu_\beta - 1)} \gamma^{\nu_\gamma - 1} \exp(-\lambda_\gamma \gamma) \\ & \quad \times \prod_{i \in \mathcal{C}} \{\gamma \exp(-\gamma(R_i - I_i))\} \prod_{i \in \mathcal{U}} \exp(-U_i) \end{aligned} \tag{5.3}$$

where $I_k = R_k - \frac{1}{\gamma} U_k$ $(k \in \mathcal{U})$.

We now describe how the resulting γpNC algorithm was implemented using (5.3).

For $1 \leq i \leq m$ and $0 \leq \mu \leq 1$, let

$$Z_i = \begin{cases} 1 & \text{with probability } \mu, \\ 0 & \text{with probability } 1-\mu. \end{cases} \tag{5.4}$$

Then set $\mathcal{C} = \{i : Z_i = 1\}$ and $\mathcal{U} = \{i : Z_i = 0\}$.

1. `Update` $\mathbf{Z}$, `and hence` $\mathcal{C}$ `and` $\mathcal{U}$, `using (5.4).`
2. `Update` γ `using RWM with proposal distribution` $N(\gamma, \sigma_\gamma^2)$.
   Note that unless $\mathcal{U} = \emptyset$, it is extremely difficult to sample γ directly from the conditional distribution.
3. `Draw` j `uniformly at random from` $\{1, 2, \ldots, m\}$.
   `Then if` $j \in \mathcal{C}$ ($j \in \mathcal{U}$) `update` I_j (U_j) `using a Metropolis-Hastings step by proposing` $R_j - I_j' \sim Exp(\gamma)$ ($U_j' \sim Exp(1)$).

Then β values can be generated from the resultant sample of $(\mathbf{R}, \mathbf{I})$ values as before.

An alternative partially non-centered algorithm based on Papaspiliopoulos *et al.* (2003), is to partially non-center each infectious period. As mentioned in the introduction, this algorithm did not perform as well as the non-centered algorithm outlined above, hence the details are omitted.

5.2. *Results*

We compare the γpNC algorithm with both the centered (CE) algorithm outlined above and the O'Neill and Roberts (OR) algorithm (see, O'Neill and Roberts (1999)). Note that both the CE and OR algorithms are examples of fully (100%) centered algorithms, however, as observed above they differ in their construction and as consequence have different efficiencies. The comparison is done using a simulated data set 1 (population size 200 of whom 82 are ultimately infected; true model parameters $\beta = 0.001$ and $\gamma = 0.15$) and a smallpox data set which sees an outbreak of size 30 in a closed population of size 120 (see, for example, Bailey (1975), O'Neill and Roberts (1999) and O'Neill and Becker (2001)). For both data sets, we consider a range of partial non-centerings, in particular for $j = 0, 1, \ldots, 10$, we ran the γpNC algorithm with $\mu = 0.1j$. In each case the algorithm was run for a burn-in period of 10000 iterations followed by 10000 further iterations. The acceptance rate for γ was monitored over the further 10000 iterations. If the acceptance rate was less than 20% or greater than 40%, we adjusted σ_γ and ran the algorithm for a further 10000 iterations, again monitoring acceptance rates. We continue this process until we have an acceptance rate between 20 and 40%. We then fixed σ_γ since an acceptance rate between 20 and 40% is close to optimal (see, Roberts and Rosenthal (2001)). The algorithm was then run for a further 5000000 iterations to obtain a sample of size 500000, from the stationary distribution, taken after every 10 iterations. The output was thinned to allow storage of data from longer runs of the algorithm. In Table 1, we record for both data sets and each partial non-centering, the value of σ_γ and the acceptance rate over the 5000000 iterations used to obtain the sample. In all cases we set $\pi(\beta) \sim Exp(0.001)$ and $\pi(\gamma) \sim Exp(0.001)$, corresponding to weak, uninformative priors.

Table 1. *σ_γ and the acceptance rates for the γpNC algorithm for the smallpox and simulated data sets, respectively*

	Smallpox data set		Simulated data set	
% Centered	σ_γ	% Accepted	σ_γ	% Accepted
0	0.050	21.7	0.048	28.3
10	0.050	23.4	0.059	25.1
20	0.050	25.2	0.073	22.0
30	0.050	27.0	0.078	21.8
40	0.050	28.8	0.059	28.3
50	0.050	30.7	0.095	19.8
60	0.050	32.4	0.100	19.8
70	0.050	34.4	0.067	28.8
80	0.050	36.3	0.100	21.3
90	0.050	38.3	0.100	22.4
100	0.106	25.1	0.100	23.5

To compare the efficiency of the algorithms, we use the integrated autocorrelation function, see, for example, Geyer (1992). Let $\boldsymbol{\gamma} = (\ldots, \gamma_0, \gamma_1, \ldots)$ denote a sample of γ from the stationary distribution and let C_γ denote the integrated autocorrelation function for γ then $C_\gamma = 1 + 2\sum_{k=1}^{\infty} \rho_k$ where $\rho_k = corr_\pi(\gamma_0, \gamma_k)$. Let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_{500000})$ denote the γ values from one particular realisation of the algorithm. Therefore for all $k \geq 1$, let $t_k = 500000 - k$ and then we have the following unbiased estimate for ρ_k from the data

$$\hat{\rho}_k = \frac{1}{\hat{\sigma}_\gamma^2}\left\{\frac{1}{t_k}\sum_{i=1}^{t_k}\gamma_i\gamma_{i+k} - \left(\frac{1}{t_k}\sum_{i=1}^{t_k}\gamma_i\right)\left(\frac{1}{t_k}\sum_{i=1}^{t_k}\gamma_{i+k}\right)\right\}$$

where $\hat{\sigma}_\gamma^2$ denotes the estimated (from the data) variance for the posterior distribution of γ. Then for $S > 1$, we can estimate C_γ by $\hat{C}_\gamma = 1 + 2\sum_{k=1}^{S}\hat{\rho}_k$. The choice of S is crucially important. If S is too small then important correlation terms ρ_k might be ignored, and hence, $\hat{C}_\gamma$ would be a bad estimate of C_γ. On the other hand, if S is too large, then it can be hard to distinguish between actual correlation and Monte Carlo error and as consequence $\hat{C}_\gamma$ can be affected by Monte Carlo error. Therefore bearing in mind the above observations, and for consistency across estimates, we have set $S = 500$ throughout this section.

The results of the MCMC output are recorded for the smallpox dataset and the simulated data set in Tables 2 and 3, respectively. The tables include the posterior means of β and γ and calculations of $\hat{C}_\beta$ and $\hat{C}_\gamma$ for the γpNC, CE and OR algorithms.

We note that the results obtained for β are qualitatively similar to those obtained for γ. Therefore from now on we restrict attention to the integrated autocorrelation function for the γ parameter. Similar qualitative results were also obtained using variance of batch means, Geyer (1992).

Tables 2 and 3 do not take into account the relative speed in terms of cpu time of the various algorithms. In particular, the CE algorithm is the fastest whilst the γpNC algorithm is the slowest.

Table 2. *Summary of MCMC output for smallpox data set*

% Centered	Mean γ	$\hat{C}_\gamma$	Mean β	$\hat{C}_\beta$
0	0.1057	62.308	9.918×10^{-4}	17.994
10	0.1051	12.915	9.895×10^{-4}	5.073
20	0.1051	7.079	9.891×10^{-4}	4.178
30	0.1050	4.636	9.883×10^{-4}	3.896
40	0.1050	4.291	9.881×10^{-4}	4.088
50	0.1051	4.274	9.863×10^{-4}	4.902
60	0.1051	4.839	9.886×10^{-4}	5.670
70	0.1052	6.803	9.899×10^{-4}	8.020
80	0.1050	8.455	9.878×10^{-4}	9.570
90	0.1050	10.295	9.879×10^{-4}	10.137
100	0.1053	18.126	9.913×10^{-4}	17.337
CE	0.1053	16.769	9.894×10^{-4}	16.092
OR	0.1053	8.928	9.908×10^{-4}	8.215

Table 3. *Summary of MCMC output for simulated data set 1*

% Centered	Mean γ	$\hat{C}_\gamma$	Mean β	$\hat{C}_\beta$
0	0.1760	85.010	1.127×10^{-3}	46.402
10	0.1750	25.402	1.122×10^{-3}	16.369
20	0.1751	17.698	1.122×10^{-3}	15.686
30	0.1757	14.755	1.126×10^{-3}	15.870
40	0.1754	16.856	1.124×10^{-3}	21.712
50	0.1753	20.274	1.125×10^{-3}	24.975
60	0.1749	28.570	1.120×10^{-3}	34.638
70	0.1755	37.723	1.125×10^{-3}	42.327
80	0.1738	55.717	1.114×10^{-3}	59.595
90	0.1760	67.764	1.128×10^{-3}	71.317
100	0.1758	105.294	1.126×10^{-3}	103.238
CE	0.1758	99.684	1.130×10^{-3}	98.472
OR	0.1760	43.173	1.130×10^{-3}	43.667

For both data sets, the speed of the γpNC algorithms was found to vary by at most 3% over the various partial non-centerings considered. Therefore we have scaled the results for the γpNC algorithms by the mean speed of this algorithm. The efficiency of an algorithm is measured as the amount of cpu time required to do 100000 iterations multiplied by the integrated autocorrelation function for γ. The relative efficiency of the algorithm is then given by the algorithm's efficiency divided by the efficiency of the OR algorithm. Thus a relative efficiency below 1 (i.e. below the solid line in Figs. 2 and 3) corresponds to an improvement on the OR algorithm.

For the smallpox data set, the optimal γpNC algorithm considered ($\mu = 0.5$) and the OR algorithm are both equally efficient. Whereas for simulated data set 1, the optimal γpNC algorithm ($\mu = 0.3$) has a relative efficiency of approximately 0.75 which is a noticeable improvement on the OR algorithm. The results of Figs. 2 and 3 demonstrate a trend we see with other data sets, that is, the larger m (the size of the outbreak) is, the more non-centered the optimum γpNC algorithm is and the greater the gains in efficiency over the OR algorithm. This trend is also seen in Fig. 4 where we consider a simulated epidemic

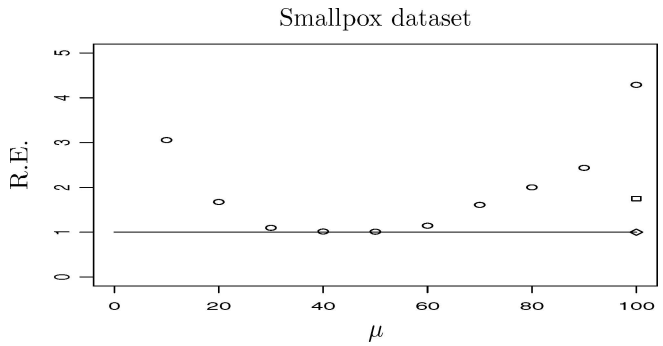

Fig. 2. *The relative efficiency (R.E.) of the γpNC (circles), CE (square) and OR (diamond) algorithms, respectively, for the smallpox data set*

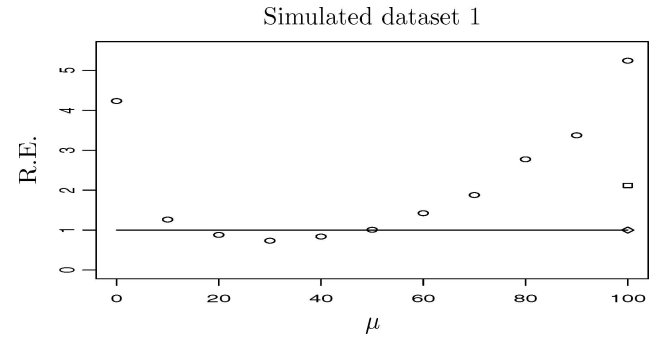

Fig. 3. *The relative efficiency (R.E.) of the γpNC (circles), CE (square) and OR (diamond) algorithms, respectively, for the simulated data set*

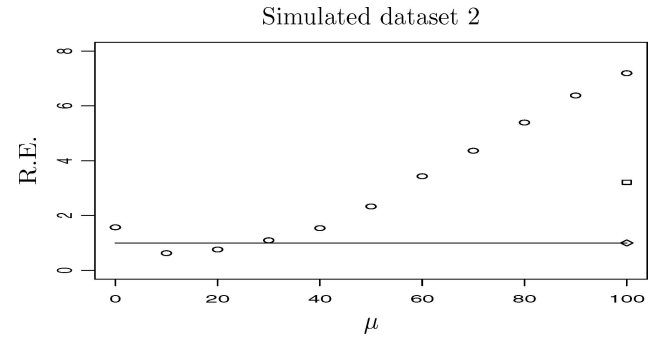

Fig. 4. *The relative efficiency (R.E.) of the γpNC (circles), CE (square) and OR (diamond) algorithms, respectively, for simulated data set 2*

of size 350 in a population of size 1000 (Simulated data set 2). Each of the algorithms was run as before for Simulated data set 2 to obtain samples of size 500000.

From Fig. 4 we have that the optimal choice of μ for the γpNC algorithm is $\mu = 0.1$. Furthermore this results in an improvement on the OR algorithm by almost a factor of 3, and has a relative efficiency of approximately 0.62. A further explanation of why the optimal choice of μ decreases with m is given in Section 6.3 below.

For the γpNC algorithm, we only update one of the U_i's or I_i's in each iteration. Alternatively we could do multiple updates

of the U_i's and I_i's in each iteration. This can be done by either repeating step 3 a number of times within each iteration or by block updating (proposing to update more than one of the U_i's and/or I_i's at once). The block updating was found to be very inefficient, due to the likelihood being discontinuous. Repeating step 3 a number of times within each iteration did improve the efficiency of the algorithm but not considerably. The same approach can be adopted in the CE and OR algorithms and as a result we get similar results in terms of relative efficiency to those obtained in Figs. 2–4.

6. Extensions of the general stochastic epidemic

6.1. *The partially non-centered parameterisations*

There are numerous extensions of the GSE, see for example, Ball (1986) and O'Neill and Becker (2001). One of the most natural and easy to implement extensions is to allow more general infectious periods. That is, let Q be an arbitrary but specified non-negative distribution. Let $g_Q(\cdot)$ denote the probability density function of Q and let γ denote the parameters governing Q. Then (5.1) can be rewritten for general infectious period Q,

$$f(\mathbf{R}, \mathbf{I} \mid \beta, \gamma) \propto \prod_{i \neq \kappa} \{\beta Y_{I_i-}\} \exp(-\beta A) \prod_{j=1}^{m} g_Q(R_j - I_j) \quad (6.1)$$

where Y_t denotes the number of infectives at time t, $Y_{I_i-} = \lim_{s \uparrow I_i}(Y_s)$ and $A = \sum_{j=1}^{m} \sum_{k=1}^{N} \{(R_j \wedge I_k) - (I_j \wedge I_k)\}$ with $I_k = \infty$ for $k = m+1, m+2, \ldots, N$. In particular, we could consider $Q \sim Gam(\alpha, \delta)$, that is, $g_Q(x) = \frac{\delta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\delta x)$ and $Q \sim Weib(\alpha, \delta)$, that is, $g_Q(x) = \alpha \delta x^{\alpha-1} \exp(-\delta x^\alpha)$. Note that for both the Gamma and Weibull distributions, if we set $\alpha = 1$ then we recover the exponential distribution, and hence, the GSE. Since similar results were obtained for both the Gamma and Weibull distributions, we shall restrict attention to the Gamma distribution.

The Gamma distribution is defined by two parameters, the shape parameter, α, and the scale parameter, δ. We shall focus attention on the case where the shape parameter α is assumed to be known although the methodology can be extended to the case where α is unknown and is therefore a parameter within the model.

As in Section 5.1, we shall assume that *apriori* β and δ are independent. The prior distributions (for conjugacy) for β and δ are $Gam(\nu_\beta, \lambda_\beta)$ and $Gam(\nu_\delta, \lambda_\delta)$, respectively. It is then straightforward to integrate β and δ from (6.1), and construct a centered MCMC algorithm (CE algorithm) along similar lines to Section 5.1. The procedure alternates between updating δ from its conditional distribution and updating an infection time by proposing $R_j - I'_j \sim Q$.

We introduce a δ-partially-non-centered MCMC algorithm (δpNC algorithm) for the Gamma distribution. The procedure is basically the same as for the γpNC algorithm for the GSE. Partition the set of individuals into two groups $\mathcal{C}$ and $\mathcal{U}$, respectively. For $Q \sim Gam(\alpha, \delta)$, let $U_i = \delta(R_i - I_i)$ $(i \in \mathcal{U})$. We then propose a change of variable from $(\mathbf{R}, \mathbf{I}^{\mathcal{C}}, \mathbf{I}^{\mathcal{U}}, \beta, \alpha, \delta)$ to $(\mathbf{R}, \mathbf{I}^{\mathcal{C}}, \mathbf{U}^{\mathcal{U}}, \beta, \alpha, \delta)$. Then using (6.1) and integrating out the parameter β, gives

$$\begin{aligned}\hat{f}(\mathbf{R}, \ \mathbf{I}^{\mathcal{C}}, \mathbf{U}^{\mathcal{U}}, \delta \mid \alpha) &\propto \left\{ \prod_{i \neq \kappa} Y_{I_i-} \right\} (\lambda_\beta + A)^{-(m+\nu_\beta-1)} \delta^{\nu_\delta - 1} \exp(-\lambda_\delta \delta) \\ &\times \prod_{i \in \mathcal{C}} \left\{ \frac{\delta^\alpha}{\Gamma(\alpha)} (R_i - I_i)^{\alpha-1} \exp(-\delta(R_i - I_i)) \right\} \\ &\times \prod_{i \in \mathcal{U}} \left\{ \frac{1}{\Gamma(\alpha)} U_i^{\alpha-1} e^{-U_i} \right\}.\end{aligned}$$

We are now in a position to describe the δpNC algorithm. For $1 \leq i \leq m$ and $0 \leq \mu \leq 1$, let

$$Z_i = \begin{cases} 1 & \text{with probability } \mu, \\ 0 & \text{with probability } 1 - \mu. \end{cases} \quad (6.2)$$

Then set $\mathcal{C} = \{i : Z_i = 1\}$ and $\mathcal{U} = \{i : Z_i = 0\}$.

1. `Update` $\mathbf{Z}$, `and hence` $\mathcal{C}$ `and` $\mathcal{U}$ `using (6.2).`
2. `Update` δ `using RWM with proposal distribution` $N(\delta, \sigma_\delta^2)$.
3. `Draw` j `uniformly at random from` $\{1, 2, \ldots, m\}$. `Then if` $j \in \mathcal{C}$ $(j \in \mathcal{U})$ `update` I_j (U_j) `using a Metropolis-Hastings step by proposing` $R_j - I'_j \sim Gam(\alpha, \delta)$ $(U'_j \sim Gam(\alpha, 1))$.

Then β values can be generated from the resultant sample of $(\mathbf{R}, \mathbf{I})$ values as before.

6.2. *Results*

We compare the δpNC algorithm with the CE algorithm. The comparison is done using a simulation study and a comparison of the integrated autocorrelation function for the various algorithms. The parameters used for the Gamma distribution simulation study are presented in Table 4.

For each of the data sets, we consider a range of partial non-centerings, in particular for $j = 0, 1, \ldots, 10$, we ran the δpNC algorithm with $\mu = 0.1j$. Samples of size 500000 were obtained in exactly the way as in Section 5.1. We choose uninformative priors for β and δ, in particular, we set $\pi(\beta) \sim$

Table 4. *Simulation study for Gamma distribution*

Simulated data set	Gamma 1	Gamma 2	Gamma 3	Gamma 4
True α	0.2	0.5	2	5
True δ	0.1	0.25	1	2.5
True β	0.007	0.007	0.007	0.007
n	100	100	100	100
m	61	57	49	61

$Gam(0.001, 0.001)$ and $\pi(\delta) \sim Gam(0.001, 0.001)$. To compare the efficiency of the algorithms we restrict attention to the integrated autocorrelation function for δ, C_δ. As before we estimate C_δ by $\hat{C}_\delta = 1 + 2\sum_{k=1}^{500} \hat{\rho}_k$. We get similar qualitative results if we were to consider the β parameter instead. The results of the simulation study are reported in Table 5.

In Table 5, the posterior means for the δ parameter are generally slightly higher than the parameter values use in the simulations. This is due to the posterior distribution of δ being right skewed. However, in all cases, it can be seen from posterior density plots for δ that the posterior mode of δ is very close to the true parameter value.

Table 5 is very informative about the relative merits of the centered and non-centered parameterisations in an epidemic context. Note that all the simulated epidemics were of similar sizes within populations of size 100. This was done so that we could compare the merits of the algorithms for different values of α. The results tell us that as α increases, that the optimal partial non-centered algorithm becomes increasingly non-centered. We have that for $\alpha = 0.2, 0.5, 2.0$ and 5.0, the corresponding optimal non-centered algorithms have $\mu = 0.7, 0.5, 0.3$ and 0.1, respectively. The table also demonstrates the vast difference between the best and worst partial non-centered algorithms for a particular data set. Note that if anything the table undersells these differences since the estimates of $\hat{C}_\delta > 100$ are conservative (for example, in the worst case, data set Gamma 1 and the fully non-centered algorithm ($\mu = 0.0$), $\hat{\rho}_{501} = 0.173$ which is far from being insignificant).

However, Table 5 does not take into account the relative speed, in terms of cpu time, of the two algorithms with the CE algorithm being just over two times faster than the δpNC algorithm. The comparison of relative efficiency of the various partial non-centerings are done in a similar manner to Section 5.2. The relative efficiency of the algorithm is given by dividing the algorithm's efficiency by the efficiency of the CE algorithm. Thus a relative efficiency below 1 corresponds to an improvement on the CE algorithm, which is the natural centered algorithm for general infectious periods. The relative efficiencies of the various partial non-centerings for the simulation study data sets are given in Fig. 5.

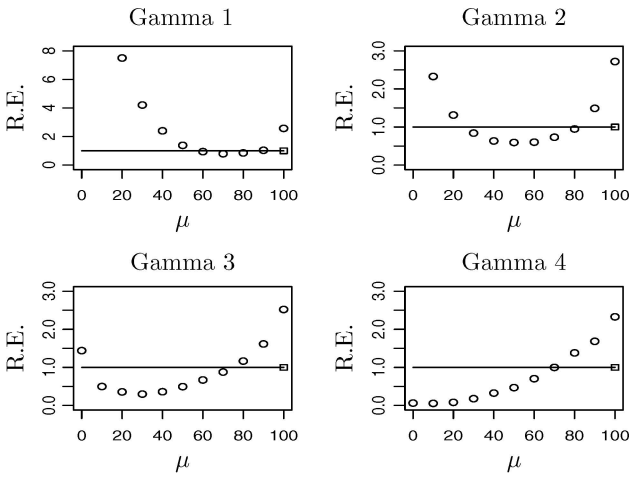

Fig. 5. *The relative efficiency (R.E.) of the δpNC (circles) and CE (square) algorithms, respectively, for the Gamma simulation study*

Figure 5 shows the great gain in efficiency that can be delivered by the partial non-centered algorithms. In the best case, (data set Gamma 4 and $\mu = 0.1$), the gain is a factor of approximately 18. Note, by contrast, that in the worst cases the relative efficiencies for certain values of μ are off the scale given in the plots. In all cases, the optimal partial non-centered algorithm out performs the CE algorithm. These results clearly hold for the GSE if we were to compare the γpNC algorithm with the CE algorithm rather the OR algorithm which is a special case (see Figs. 2–4). In other words, the GSE example given in Section 5 undersells the benefits of the partially non-centered

Table 5. *Estimates of the posterior mean of δ and the integrated autocorrelation function of δ, $\hat{C}_\delta$ for the Gamma simulation study*

	Gamma 1		Gamma 2		Gamma 3		Gamma 4	
% Centered	Mean δ	$\hat{C}_\delta$	Mean δ	$\hat{C}_\delta$	Mean δ	$\hat{C}_\delta$	Mean δ	$\hat{C}_\delta$
0	0.1226	406.816	0.3064	157.182	0.9963	36.508	2.591	9.171
10	0.1244	129.828	0.3029	37.685	0.9963	12.616	2.592	7.930
20	0.1218	74.767	0.3016	21.283	0.9966	9.041	2.592	11.774
30	0.1233	41.929	0.3030	13.597	0.9962	7.579	2.586	26.098
40	0.1234	23.928	0.3029	10.270	0.9949	9.126	2.584	47.448
50	0.1234	13.776	0.3024	9.617	0.9964	12.547	2.583	68.359
60	0.1233	9.388	0.3036	9.755	0.9959	16.986	2.592	102.394
70	0.1230	7.915	0.3013	11.845	0.9943	22.219	2.592	145.620
80	0.1235	8.498	0.3022	15.355	0.9962	29.573	2.584	201.295
90	0.1240	10.344	0.3024	24.151	0.9996	40.968	2.639	245.667
100	0.1231	25.644	0.3028	44.056	0.9979	63.916	2.572	320.914
CE	0.1234	21.935	0.3028	35.640	0.9978	55.796	2.551	339.490

parameterisation, since we compare the γpNC algorithm with the special OR algorithm rather than the generic CE algorithm. Furthermore, the partial non-centered approach could prove very useful from a practical point of view. The distribution of the length of infectious periods of many diseases, such as measles (see, Neal and Roberts (2004)), can be approximated by a Gamma distribution with $\alpha > 1$, and as we have seen the partial non-centered parameterisation is particularly good in such situations.

Similar observations are made for the Weibull distribution, in that, as α increases the optimal partially non-centered algorithm becomes increasingly non-centered.

6.3. *Remarks*

We proceed by giving an explanation for why the non-centered parameterisation is to be preferred as m and α increase for Gamma distributed infectious periods. Suppose that α is fixed (as above) and that we observe an epidemic of size m. Assign the improper, uninformative prior, $\pi(\delta) \sim Gam(0, 0)$ to the parameter δ. Then for a given value of δ, the sum of the infectious periods are apriori Gamma distributed

$$\sum_{j=1}^{m}(R_j - I_j) \mid \alpha, \delta \sim Gam(m\alpha, \delta).$$

Also the posterior distribution of δ is

$$\pi(\delta \mid \mathbf{R}, \mathbf{I}, \alpha) \sim Gam\left(m\alpha, \sum_{j=1}^{m}(R_j - I_j)\right).$$

Thus for large m and/or α, the parameter δ and the sum of the infectious periods are apriori heavily correlated. Therefore if we consider δ and $\sum_{j=1}^{m}(R_j - I_j)$ as the two parameters of interest, then the centered algorithms alternates between updating each of these two parameters. It is well known that for the two-state Gibbs sampler that convergence of the algorithm is linked to the correlation between the parameters, see Amit (1991). This suggests that the centered algorithm performs badly when $m\alpha$ is large. This is accentuated by the fact that we only update one of the infection periods during each iteration. Clearly, our situation is far more complicated than a two state Gibbs sampler due to the interactions between the different removal and infection times. However, the results of Sections 5.2 and 6.2 support the above line of argument.

The non-centered parameterisation gets around the above problem of high correlation by breaking down the dependence between the infection times and δ. Note that apriori $\sum_{j=1}^{m} U_j \mid \alpha, \delta \sim Gam(m\alpha, 1)$, and so, apriori $\sum_{j=1}^{m} U_j$ is independent of δ. However, this does not take account of the intricate relationship between the different infection times $\mathbf{I}$ and the parameter δ.

This is obviously not the full story since the updating of the infection times $\mathbf{I}$ in such a way as to explore the full target space is needed. This is best done using a partial non-centered algorithm, since the centered algorithm updates only one infection time per iteration and the fully non-centered algorithm whilst updating all the infection times when updating δ maintains the ratio of length between the different infectious periods. Therefore, for $i \neq j$, it can take many iterations to alter the ratio $\frac{R_j - I_j}{R_i - I_i}$. Thus the best partial non-centered algorithm from this perspective is probably $\mu = 0.5$.

Therefore there is an intricate trade-off between many considerations when deciding on the algorithm to use. The results suggest that except for very small $m\alpha$, the δpNC algorithm is preferable to the CE algorithm. The choice of μ is then dependent on m and α with the optimal choice of μ decreasing with m and α. From a practical point of view, short runs of length 10000 say, with different values of μ give a fair idea of a good choice for μ. All the results suggest that the choice of μ is fairly robust, in that, any choice of μ close to the optimal choice of μ will produce a close to optimal δpNC algorithm.

The above observations hold for the Weibull distribution as well, and will probably hold for far more general choices of Q. However, a generalisation of the above argument does not readily present itself.

6.4. *Further extensions*

There are a number of further extensions that one could consider such as the inclusion of a latent period. The most natural extension of the above work is to consider unknown shape parameter α. In the case where Q is Gamma distributed, it is possible to non-center the infectious period with respect to the shape parameter as well as the scale parameter (see, Papaspiliopoulos *et al.* (2003), for details). One would expect to observe similar results to those given above, that is, when α is large then the non-centered algorithm for δ is good. However, the situation is complicated by questions such as whether to center or non-center with respect to the parameter α, the interaction between the parameters α and δ and how this affects the choice of algorithm and whether to update α and δ separately or jointly. There does not seem to be straightforward answers to these questions and this is a subject for future work.

7. Bernoulli random graph general stochastic epidemic

7.1. *A non-centering reparameterisation*

We now turn our attention to the BGSE. For convenience, we label the infected individuals $1, 2, \ldots, m$ such that for $1 \leq j \leq m$, individual j is removed at time R_j. The susceptible individuals are labelled $m + 1, m + 2, \ldots, N$.

The graph $\mathcal{G}$ can be constructed as follows. For $1 \leq i < j \leq N$, let $U_{ij} \sim U(0, 1)$. Then an edge exists between vertices i and j in the graph $\mathcal{G}$ if and only if $U_{ij} \leq p$. Note that $G_{ij} = 1_{\{U_{ij} \leq p\}}$. For completeness, set $U_{ji} = U_{ij}$ $(1 \leq i < j \leq N)$, $U_{ii} = 1$ $(1 \leq i \leq N)$ and $\mathbf{U} = \{U_{ij} : 1 \leq i, j \leq N\}$. Let $\mathcal{P}$ denote the infection path, i.e. who infects who. Then $\mathcal{P}$ is a random directed tree upon the graph $\mathcal{G}$ whose root is the vertex corresponding to the initial infective, κ. We assume that $\mathcal{P}$ is unknown, but we

choose $\mathcal{P}$ uniformly at random from all the possible infection paths given $(\mathcal{G}, \mathbf{I}, \mathbf{R})$. Note that apriori all the possible infection paths are equally likely.

Let $L = \pi(\mathbf{I}, \mathbf{R} \mid \beta, \gamma, p, \mathbf{U}, \mathcal{P})$ and $A = \sum_{j=1}^{m} \sum_{k=1}^{N} G_{jk}\{(I_k \wedge R_j) - (I_k \wedge I_j)\}$ with the convention that for $k \geq m+1$, $I_k = \infty$. Thus as before A denotes the total number of person-to-person units of infectious pressure exerted during the course of the epidemic. Then since $\mathcal{G}$, and hence $\mathcal{P}$, can be constructed from $(\mathbf{U}, p)$, it follows from Britton and O'Neill (2002) after some straightforward algebra, that

$$L = \beta^{m-1} \exp(-\beta A)\gamma^m \exp\left(-\gamma \sum_{j=1}^{m} (R_j - I_j) \right). \quad (7.1)$$

In fact, for inference we don't need the full graph $\mathcal{G}$ but the subgraph $\mathcal{F}$, $\{G_{ij} : 1 \leq i \leq m, i < j \leq N\}$, since the likelihood does not depend on whether or not links exist between individuals who remain susceptible throughout the epidemic. Similarly, we only need $\mathbf{U}^* = \{U_{ij} : 1 \leq i \leq m, i < j \leq N\}$ rather than $\mathbf{U}$.

We follow Britton and O'Neill (2002) in assigning independent priors to the individual parameters, in particular, $\pi(\beta) \sim Gam(\nu_\beta, \lambda_\beta)$, $\pi(\gamma) \sim Gam(\nu_\gamma, \lambda_\gamma)$ and $\pi(p) \sim Beta(d_1, d_2)$. Thus, by Bayes' Theorem, we have that

$$\pi(p, \mathcal{F}, \mathcal{P} \mid \mathbf{I}, \mathbf{R}) \propto L\pi(\beta)\pi(\gamma)\pi(\mathcal{F} \mid p)\pi(p), \quad (7.2)$$

where, since $\mathcal{F}$ is a subgraph of $\mathcal{G}$, a Bernoulli random graph,

$$\pi(\mathcal{F} \mid p) = p^{|\mathcal{F}|}(1-p)^{K-|\mathcal{F}|}$$

and $K = \binom{m}{2} + m(N-m)$. Furthermore, we can integrate out β and γ giving,

$$\pi(p, \mathcal{F}, \mathcal{P} \mid \mathbf{I}, \mathbf{R}) \propto (\lambda_\beta + A)^{-(m+\nu_\beta-1)} \left(\lambda_\gamma + \sum_{j=1}^{m} (R_j - I_j) \right)^{-(m+\nu_\gamma)} \pi(\mathcal{F} \mid p)\pi(p). \quad (7.3)$$

Therefore we have the following algorithm based on (7.3), which is an adaption of the Britton and O'Neill (2002) algorithm (ABO algorithm). The main difference from the Britton and O'Neill (2002) algorithm, is that, we use the subgraph $\mathcal{F}$ instead of the full graph $\mathcal{G}$. The adaption of the algorithm has been done to improve mixing.

1. Update p, β and γ, respectively, from the appropriate conditional distributions.
2. Update $\{G_{ij}; 1 \leq i \leq m, i < j \leq N\}$ one at a time using a Gibbs sampler (see, Britton and O'Neill (2002)) and then pick $\mathcal{P}$ uniformly at random from all the possible paths.

   Note that for steps 3 to 5, we integrate out β and γ and use (7.3).
3. Update the components of $\mathbf{I}$ one at a time, in a uniformly random order using a Metropolis-Hastings step (see, Britton and O'Neill (2002)).
4. Update κ, by proposing a new initial infective and calculating the acceptance ratio for the move (see, Britton and O'Neill (2002)).
5. Draw c uniformly from the interval $[r^{-1}, r]$, where $r > 1$ is constant. The value c is then used to rescale the current set of infection times as outlined in Britton and O'Neill (2002). However, since β and γ are integrated out of the likelihood, we ignore the rescaling of β and γ.

The above centered algorithm updates p conditional upon the graph $\mathcal{F}$ (and $\mathbf{I}, \mathbf{R}$) and then updates $\mathcal{F}$ conditional upon p (and $\mathbf{I}, \mathbf{R}$). The reparameterised algorithm below utilises the construction of the graph $\mathcal{F}$ given at the beginning of this section using $\mathbf{U}^*$ and p. The resulting non-centered algorithm updates the graph $\mathcal{F}$ when updating p. The $\mathbf{U}^*$ are updated separately and this also updates $\mathcal{F}$. Note that $\prod_{k \neq \kappa}\{\sum_{j=1}^{m} 1_{\{I_j < I_k \leq R_j\}} G_{jk}\}$ is the number of possible infection paths $\mathcal{P}$ given $(\mathbf{U}^*, p, \mathbf{I}, \mathbf{R})$. Therefore by Bayes' Theorem the posterior density of interest, $\pi(p, \mathbf{U}^* \mid \mathbf{I}, \mathbf{R})$, is given by

$$\pi(p, \mathbf{U}^* \mid \mathbf{I}, \mathbf{R}) \propto \frac{\prod_{k \neq \kappa}\{\sum_{j=1}^{m} 1_{\{I_j < I_k \leq R_j\}} G_{jk}\} p^{d_1-1}(1-p)^{d_2-1}}{(\lambda_\beta + A)^{m+\nu_\beta-1}(\lambda_\gamma + \sum_{j=1}^{m}(R_j - I_j))^{m+\nu_\gamma}}. \quad (7.4)$$

We are now ready to outline our non-centered MCMC algorithm (NC algorithm) which breaks the apriori dependence between p and the random graph.

1. Update p using logarithmic RWM with proposal distribution $\log p' \sim N(\log p, \sigma_p^2)$.
2. Update the components of $\mathbf{U}^*$ one at a time using a Metropolis-Hastings step by proposing $U_i' \sim U(0,1)$.
3. Update the components of $\mathbf{I}$ one at a time, in a uniformly random order using RWM with $I_k' \sim N(I_k, \sigma_I^2)$.
4. Update κ as outlined below.

The procedure we adopt for updating κ is similar to that of Britton and O'Neill (2002). Suppose that individual i is the initial infective, so $\kappa = i$. Let $C_i = \{k : I_i < I_k \leq R_i \text{ and } G_{ik} = 1\}$ and let $\psi_i = |C_i|$. We select an individual, j say, uniformly at random from C_i. Now we propose to make $\kappa = j$ by swapping the infection times of individuals i and j, so that individual i becomes infected at time $I_i' = I_j$ and individual j becomes infected at time $I_j' = I_i$. Let $C_j' = \{k : I_j' < I_k' \leq R_i \text{ and } G_{jk} = 1\}$ and let $\psi_j' = |C_j'|$. Then we accept the move $\kappa = j$ with probability,

$$\min\left\{ 1, \frac{\prod_{k \neq \kappa'}\{\sum_{l=1}^{m} 1_{\{I_l' < I_k' \leq R_l\}} G_{lk}\}}{\prod_{k \neq \kappa}\{\sum_{l=1}^{m} 1_{\{I_l < I_k \leq R_l\}} G_{lk}\}} \left(\frac{\lambda_\beta + A}{\lambda_\beta + A'} \right)^{m+\nu_\beta-1} \frac{\psi_i}{\psi_j'} \right\}.$$

β and γ values can be generated from the resultant sample of $(p, \mathbf{U}^*, \mathbf{I})$ values according to their appropriate conditional distribution.

Table 6. *Detection times of cases of Gastroenteritis, Britton and O'Neill (2002), Example 3*

Day	0	1	2	3	4	5	6	7
Cases	1	0	4	2	3	3	10	5

The NC algorithm involves non-centering the graph $\mathcal{G}$, or, in particular, the subgraph $\mathcal{F}$. As with the GSE, we could introduce a partial non-centering for the graph $\mathcal{F}$, updating some of the edges with p while keeping other edges fixed. However, results over a number of data sets showed that either the ABO algorithm or NC algorithm out performed any of the partially non-centered algorithms. An explanation of why this should be the case is given in Section 7.3 below. Therefore for clarity in exposition, we have described the non-centered algorithm rather than the more complicated partially non-centered algorithm.

7.2. *Results*

We use Britton and O'Neill (2002), Example 3 to compare the ABO and NC algorithms. This example concerns an outbreak of Gastroenteritis in a hospital ward in South Carolina, January (1996), as reported in Cáceres *et al.* (1998). We follow Britton and O'Neill (2002) in restricting attention to the transmission of the disease through the nursing staff only. As noted by Britton and O'Neill (2002), the application of the BGSE to this particular data set is somewhat questionable but since we wish to compare our methodology with Britton and O'Neill (2002), we will use this data set and tolerate the less realistic assumptions. The removal times correspond to the date upon which the onset of symptoms were detected and the epidemic comprised 28 infectious cases among a population of size 89. The removal times are given in Table 6.

We shall focus on the parameter p. This is indicative of how well the MCMC algorithm mixes with respect to β and γ as well. We shall follow Britton and O'Neill (2002) and use weak, uninformative priors, namely, $\nu_\beta = \lambda_\beta = \nu_\gamma = \lambda_\gamma = 0.001$ and $d_1 = d_2 = 1$. For each of the algorithms, we obtained a sample of size 100000 taken after every 10 iterations with a burn-in of 10000 iterations.

From Fig. 6, we can see that for this data set the NC algorithm mixes better than the ABO algorithm. The time series plots suggest that there is a lot of uncertainty concerning p. We note the estimated posterior mean (standard deviation) for p are 0.55 (0.27) using the ABO algorithm and 0.56 (0.26) using the NC algorithm. Further, the time series plot for the NC algorithm suggests that the NC algorithm mixes better for higher rather than lower values of p. This was also the case in other examples we considered. We ran the NC algorithm to obtain samples of size 100000, firstly with $\log p' \sim N(\log p, \sigma_p^2)$ (logarithmic RWM) and then with $p' \sim N(p, \sigma_p^2)$ (standard RWM). Then in Fig. 7 for each batch of 100 observations we plot the batch mean against the proportion of accepted moves. We can see that

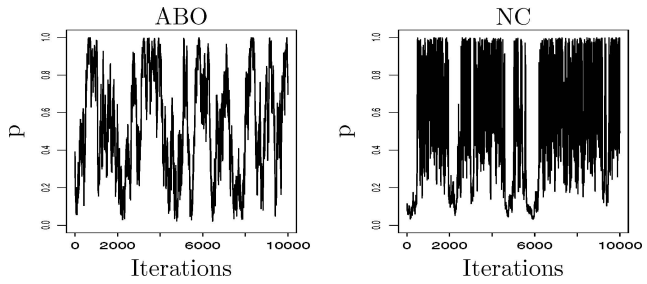

Fig. 6. *Time series plots for the first 10000 iterations of p for Britton and O'Neill (2002), Example 3, with output from the ABO and NC algorithms, respectively*

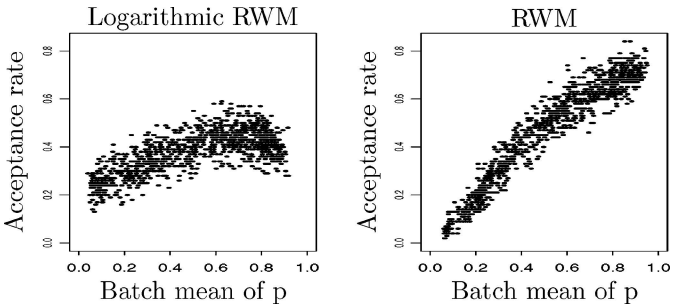

Fig. 7. *Scatterplot of batch mean of p against proportion of accepted moves for Britton and O'Neill (2002), Example 3, using logarithmic RWM and standard RWM, respectively*

using logarithmic RWM is a big improvement on using standard RWM.

The choice of prior for $p, \pi(p)$, is very important since the NC algorithm performs better for large values of p than for small values of p. We therefore ran both algorithms with $\pi(p) \sim Beta(3, 1)$ and $\pi(p) \sim Beta(1, 3)$. These prior distributions correspond to the prior belief that p is large or small, respectively. From the MCMC output, the autocorrelation function (acf) can be estimated, this is summarised in the acf plots in Fig. 8 for the different choice of priors for p.

The results of Fig. 8 support the observation made earlier that the real gains in mixing from using the NC algorithm are when p is large. Even when the underlying graph is fairly sparse (i.e. p is small) the NC algorithm performs no worse than the ABO algorithm which appears to perform equally well for all choices of $\pi(p)$.

We have not conducted a formal comparison of the computing costs of the ABO and NC algorithms since the difference in running time between the algorithms over a range of examples was negligible.

7.3. *Remarks*

The following two questions need further explanation, why does the NC algorithm mix better for large p rather than small p and why is the non-centered algorithm preferable to the partially

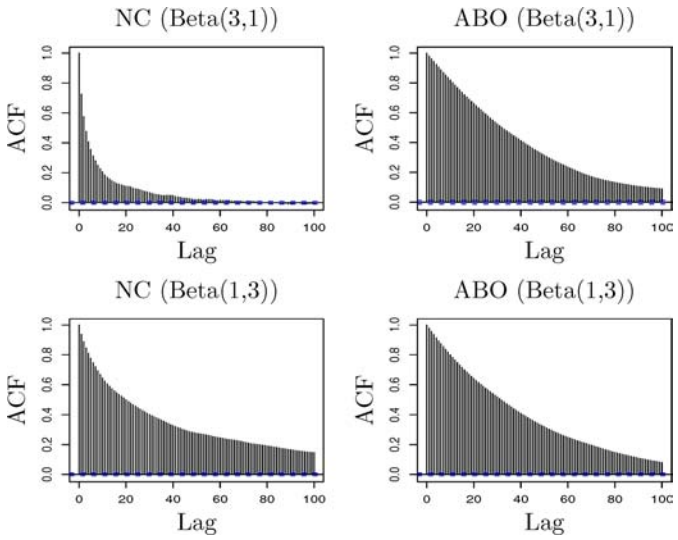

Fig. 8. *Acf plots for p, for both the ABO and NC algorithms with* $\pi(p) \sim Beta(3, 1)$ *and* $\pi(p) \sim Beta(1, 3)$*, respectively*

non-centered algorithm in apparent contradiction to the results of Sections 5 and 6? Clearly, there is high correlation between p and the number of possible infection paths since the higher p is, the more possible infection paths there are. The NC algorithm performs better when there is a large number of possible infection paths for the following reasons. If there is a large number of possible infection paths, a small proposed change in p which either adds or removes a few edges (and so, creates or deletes a few infection paths) will have little effect on the likelihood given in (7.4), and so, most moves will be accepted. On the other hand, if there are only a few possible infection paths then the addition or removal of even a few edges can have a dramatic effect on likelihood given in (7.4), such as removing a crucial link in the infection chain, and so, considerably more moves are rejected.

The construction of a partially non-centered algorithm is fairly straightforward. We considered the case, where in each iteration each edge is non-centered with probability $1 - \mu$ and centered with probability μ. However, as mentioned in Section 7.1, we found that such an algorithm performed poorly. In particular, it was found that for the above data set all values of $\mu > 0$ had worse mixing properties than the fully non-centered (NC) algorithm and except for values of μ close to 0, the mixing was considerably worse.

So why for the BGSE is the non-centered algorithm preferable to the partially non-centered algorithm, whereas for the GSE the partially non-centered algorithm is the best? There is no straightforward answer, but an understanding of why this might be the case can be obtained by considering the two models. For a BGSE epidemic of size m within in a population of size N, there is $\binom{m}{2} + m(N - m)$ edges to be imputed. (Thus the number of edges to be imputed is typically $O(m^2)$ since we are only usually interested in major outbreaks of the disease, i.e. $N = O(m)$.) This leads to a very strong prior dependence between p and the imputed variables, in particular, if we set $d_1 = d_2 = 0$ (corresponding to an uninformative, improper prior on p) then $\pi(p \mid \mathcal{F}) \sim Beta(|\mathcal{F}|, \binom{m}{2} + m(N - m) - |\mathcal{F}|)$. Thus for the BGSE, the overriding consideration is to breakdown this dependence between p and the subgraph $\mathcal{F}$ which is done by the non-centering approach. As we have mentioned in Section 6.3, the non-centered algorithm for the GSE and its extension breaks the prior dependence between **I** and the parameters governing the infectious period distribution. However, in this case the number of imputed infection times, **I**, is equal to m. Therefore the prior dependence is not as strong as for the BGSE, and so, other considerations become important in the choice of algorithm as was noted in Section 6.3.

8. Discussion

We have given two examples of models where partially (or completely) non-centered parameterisations can lead to considerably more rapidly convergent Markov chains than the conventional centered data-augmentation parameterisations. For the GSE and its extensions, the improved mixing comes at the cost of marginally increased computing costs, whereas in the BGSE, there are no extra computational costs associated with the use of non-centered methodology, despite the need for state-space expansion techniques.

The non-centered and partially non-centered algorithm become increasingly useful as the size of the epidemic m increases. The epidemics considered in this paper are generally very small with $m < 100$ (the exception being Simulated data set 2 in Section 5.2). Thus the results of this paper are of great practical interest since most epidemics that are of interest have outbreaks of size 100 or larger.

For both GSE and BGSE, our results find that non-centering improves algorithm performance, usually substantially. The two examples reach different conclusions regarding the the extent of the non-centering required for the optimal algorithm. However, some explanation for these differences is given in Section 7.3. For the BGSE the prior link between p and the graph is particularly strong so that the data is comparatively extremely weak, and fully non-centered methods are preferred.

These models are very simple although are surprisingly challenging for modern MCMC methods. However, the methodology described in this paper could be extended routinely to models where more complex imputed data needs to be included, for instance with latent periods, more general infectious periods, or random effects on infectious periods. For such models, we would expect non-centered methods to be even more advantageous over centered alternatives, since data are less likely to be able to break down strong prior dependencies. We have demonstrated this to a certain extent by considering different infectious periods in Section 6. It would be particularly interesting to study the case proposed in Section 6.4 where $Q \sim Gam(\alpha, \delta)$ and both α and δ are unknown.

The BGSE example raises interesting issues since the data are more informative for small values of p than for large ones.

As a result, the non-centered method has a tendency to mix poorly when visiting small p values. In this context, it is easy to construct hybrid MCMC strategies to solve this problem.

In conclusion therefore non-centered and partially non-centered parameterisations have much to offer problems of inference from stochastic epidemics, largely due to the sparsity of data and the consequent need to do extensive data-augmentation.

Broad conclusions about non-centering more generally can of course only tentatively be drawn. However the results of this paper reinforce those of related work on other model types (for example, in Christensen *et al.* (2003)), and is consistent with the limited theory available for the methodology (see, Papaspiliopoulos *et al.* (2003)). Thus our tentative conclusions would be that provided non-centering is practically implementable in computing time comparable to that of the basic centred method, then they will usually lead to algorithms with improved mixing. Fully non-centered methods are likely to be useful when there is a very strong prior link between missing data and parameters or where it is difficult to construct sensible partial non-centering methods.

Acknowledgments

The authors would like to thank Valerie Viallefont for her contribution in the early development of the work presented in this paper. The authors would like to thank the referees for their comments which have greatly improved the presentation of the paper. This research was supported by the UK Engineering and Physical Sciences Research Council, under research grant number GR/M62723.

References

Amit Y. 1991. On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions. J. Multivariate Analysis 38: 82–99.

Andersson H. 1999. Epidemic models and social networks. Math. Scientist 24: 128–147.

Bailey N.T.J. 1975. The Mathematical Theory of Infectious Diseases and its Applications, 2nd edn, Griffin, London

Ball F.G. 1986. A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. Adv. Appl. Prob. 18: 289–310.

Ball F.G., Mollison D., and Scalia-Tomba G. 1997. Epidemics with two levels of mixing. Ann. Appl. Prob. 7: 46–89.

Britton T. and O'Neill P.D. 2002. Bayesian inference for stochastic epidemics in populations with random social structure. Scan. J. Statist. 29: 375–390.

Cáceres V.M., Kim D.K., Bresee J.S., Horan J., Noel J.S., Ando T., Steed C.J., Weems J.J., Monroe S.S., and Gibson J.J. 1998. A viral gastroenteritis outbreak associated with person-to-person spread among hospital staff. Infection Control and Hospital Epidemiology 19(3): 162–167.

Christensen O., Roberts G.O., and Sköld M. 2003. Robust MCMC methods for spatial GLMMs. Submitted for publication.

Gelman A., Roberts G.O., and Gilks W.R. 1996. Efficient Metropolis jumping rules. Bayesian Statist. 5: 599–608.

Geyer C.J. 1992. Practical markov chain monte carlo (with discussion). Stat. Science 7: 473–511.

Keeling M.J., Woolhouse M.E.J., Shaw D.J., Matthews L., Chase-Topping M., Haydon D.T., Cornell S.J., Kappey J., Wilesmith J., and Grenfell B.T. 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochasic dispersal in a heterogeneous landscape. Science 294: 813–817.

Meng X.-L. and van Dyk D. 1997. The EM algorithm—an old folk song sung to a fast new tune (with discussion). J. R. Statist. Soc. B 59: 511–567.

Neal P.J. and Roberts G.O. 2004. Statistical inference and model selection for the 1861 hagelloch measles epidemic. Biostat. 5: 249–261.

O'Neill P.D. and Becker N.G. 2001. Inference for an epidemic when susceptibility varies. Biostat. 2: 99–108.

O'Neill P.D. and Roberts G.O. 1999. Bayesian inference for partially observed stochastic epidemics. J. R. Statist. Soc. A 162: 121–129.

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

Roberts G.O. and Rosenthal J.S. 2001. Optimal scaling for various Metropolis-Hastings algorithms. Statist. Science 16: 351–367.