Appl. Statist. (2005)
54, *Part* 3, *pp.* 575–594

Discrete time modelling of disease incidence time series by using Markov chain Monte Carlo methods

Alexander Morton and Bärbel F. Finkenstädt

University of Warwick, Coventry, UK

[Received July 2003. Revised June 2004]

Summary. A stochastic discrete time version of the susceptible–infected–recovered model for infectious diseases is developed. Disease is transmitted within and between communities when infected and susceptible individuals interact. Markov chain Monte Carlo methods are used to make inference about these unobserved populations and the unknown parameters of interest. The algorithm is designed specifically for modelling time series of reported measles cases although it can be adapted for other infectious diseases with permanent immunity. The application to observed measles incidence series motivates extensions to incorporate age structure as well as spatial epidemic coupling between communities.

Keywords: Disease incidence time series; Markov chain Monte Carlo methods; Stochastic modelling of infectious diseases

1. Introduction

The pattern of measles epidemics in developed countries before vaccination is one of the best-documented population cycles in ecology. The availability of long disease notification series and the relatively simple natural history of the disease closely following the classical susceptible–infectious–recovered (SIR) paradigm have made measles an excellent test-bed for models of epidemics and have prompted an extensive theoretical and empirical literature on the population biology of the disease (Grenfell and Dobson, 1995). Although measles is now at a low level in the UK it still causes considerable mortality in young children in developing countries (McLean and Anderson 1988a, b) and understanding the dynamics of the disease is still important from a public health point of view (Black, 1984).

The theoretical framework that is most commonly used to explain the dynamics of viral and bacterial infections is based on the division of the human host population into compartments containing susceptible, infectious and recovered individuals. The SIR model is expressed as a system of differential equations where further developments to incorporate more biological realism such as spatial epidemic coupling between populations and heterogeneities within a population (due to age structure) have led to more complex mathematical models (Schenzle, 1984; Grenfell *et al.*, 1995; Feng and Thieme, 1995; Keeling and Grenfell, 1997), the parameters of which could not all be identified from the available time series data on case notification.

The aim of the study by Finkenstädt and Grenfell (2000) was therefore to forge a link between theoretical epidemic modelling and statistical approaches to the time series data. They developed a stochastic time series susceptible–infected–recovered or TSIR model which captures the essential SIR-type mechanism of an infectious disease. The challenge here lies in developing a

Address for correspondence: Bärbel F. Finkenstädt, Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK.
E-mail: B.F.Finkenstadt@warwick.ac.uk

 0035–9254/05/54575

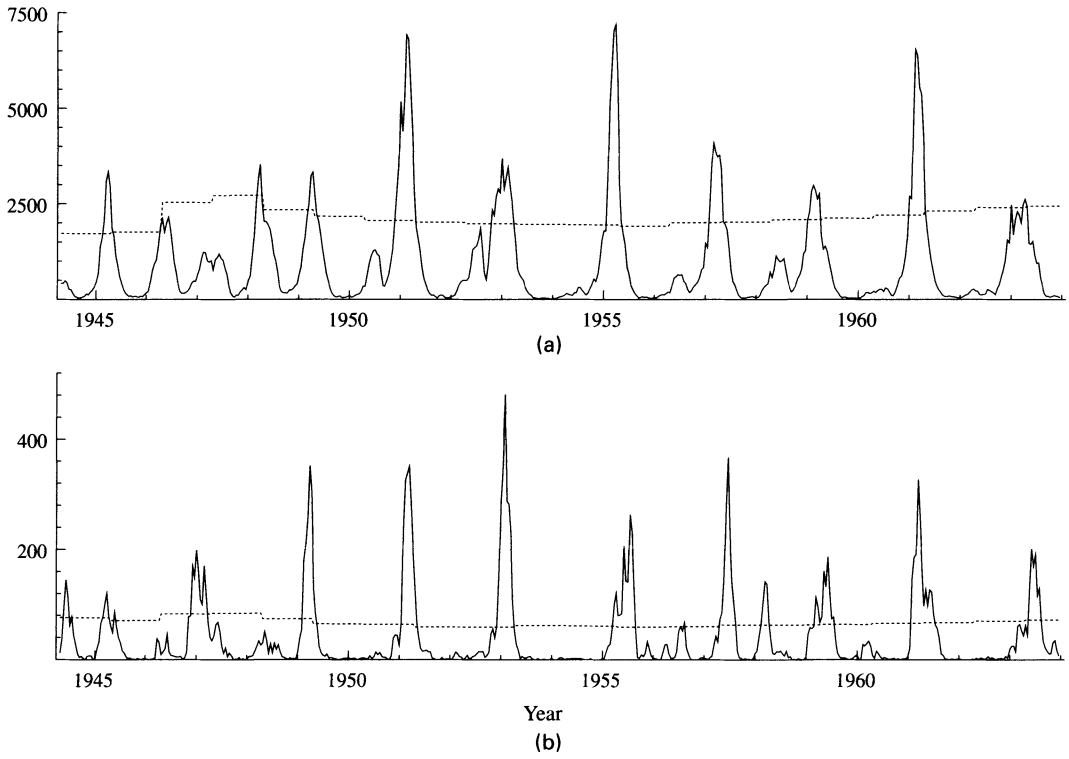

Fig. 1. Time series plots of reported measles cases (———) in each 2-week period for (a) London (population 3 200 000) and (b) Bournemouth (population 140 000) during the prevaccination era from 1944 to 1964 (the population sizes are the approximate mean sizes for this time period): ·····, number of births

model that can explain and thus predict the diverse range of different spatiotemporal dynamics that are observed in measles time series but at the same time is of a sufficiently modest structure to be identifiable from time series data. The richness in patterns is illustrated in Finkenstädt and Grenfell (1998) and Fig. 1 shows time series plots of reported measles cases in London and Bournemouth from 1944 to 1964 before mass vaccination began in Britain.

The predominant population cycle for Britain had a biennial pattern with alternating years of minor and major incidence of disease. During times of higher birth-rates, the accelerated replenishment into the susceptible class gave rise to annual cycles (Finkenstädt and Grenfell, 2000). This is observed at the time of the baby boom around 1947 as well as on the individual city level for places that are characterized by permanent high birth-rates such as Liverpool (Finkenstädt *et al.*, 1998).

The variation in the dynamic pattern with respect to community size was documented and studied in the now classic work of Bartlett (1957, 1960). Bartlett estimated a critical community size of about 250 000–300 000 inhabitants for measles. The critical community size is the size of population that is sufficiently large to maintain transmission in epidemic troughs. Large cities, above the critical community size, exhibit endemic dynamics that are characterized by smooth regular population cycles (see Fig. 1(a)). Centres below the critical community size exhibit series of epidemic outbreaks that are interrupted by extinction (Grenfell, 1992; Rhodes *et al.*, 1997) (see Fig. 1(b)). The incidences in most cities in England and Wales, therefore, are driven partly by extinction and reintroduction of the disease.

It is a challenge in mathematical epidemiology to understand the mechanisms that cause the disappearance and recurrence of epidemic outbreaks, and the establishment of a model that allows us to study this is of importance for theoretical epidemiology (Dietz, 1995). Recurrent epidemics have such intricate behaviour that they cannot be studied by deterministic models as the demographic stochasticity is an intrinsic part of the process (Nasell, 2002).

The TSIR model that was introduced in Finkenstädt and Grenfell (2000) can provide an approximation for measles time series in large communities where the disease is endemic and can forecast the effect of changes in the birth-rate. The model was further developed to account for extinction and recurrence of the disease in Finkenstädt *et al.* (2002) and the corresponding biological issues were discussed in detail in Bjørnstad *et al.* (2002) and Grenfell *et al.* (2002).

Because separate estimation methods are applied at each stage of the model hierarchy the approach that has been taken so far has been too inflexible to be of use for further development of a more realistic biological model. The aim of this paper is to develop a general and much more powerful Markov chain Monte Carlo (MCMC) approach which overcomes many of the previously encountered pitfalls. Firstly, it allows us to distinguish properly between the stochasticity of the reporting process (the measurement error) and the variability of the transmission process (the process noise). Secondly, we can add other complex, yet more biologically realistic, components into the model such as long-term variations in the rate of transmission, enabling a realistic estimation of the infected and susceptible populations as well as the rate of reporting. Finally, it is possible to draw comparisons between alternative hypotheses considering the modelling of transmission of the disease, both within and between populations.

Moreover, a proper development of the methodology for drawing inference provides a springboard for studying further topics concerning the dynamics of measles and other childhood diseases. These are undoubtedly beyond the scope of this paper. However, we outline here some of the objectives which can ultimately be achieved by using this approach.

(a) The MCMC methodology will allow us to estimate the strength of coupling between epidemic centres from data and to test hypotheses about the nature of the coupling. In particular, we can quantify how much of the observed synchrony in the population dynamics of measles is explained by the communities pertaining to the same seasonality (i.e. the school year) as opposed to the correlation that results from the spatial epidemic coupling through migration of infected people.
(b) Because measles follows the SIR paradigm very closely many important questions are currently studied for measles. Much further work is needed to build statistical models that help in understanding the dynamics of other childhood diseases, notably whooping-cough and rubella which have a more complex natural history than measles. There is plenty of scope to adapt and apply the model that is studied here to other childhood diseases.
(c) Most models in epidemiology assume that the contact process between susceptible and infected individuals can be readily modelled by the assumption of mass action. However, we can allow and test for deviations from this assumption.
(d) The model's ability to predict the effect of variations in birth-rate gives the potential to examine the important question of childhood vaccination policies for developing countries. The effect of vaccination can be directly assessed by manipulating the susceptible population. Investigation of the model's performance in predicting the observed pattern of measles incidence following the onset of vaccination is an important future goal.

The paper is structured as follows. The next section introduces the TSIR model and derives the MCMC methodology to solve the associated inferential problem. The methods are illustrated

by means of simulated data in Section 3 and are applied to real incidence series in Section 4. The real data application motivates the need for some model extensions because the model suggested does not account for the differing transmission rates for school-age and preschool susceptible children. A simple solution is proposed to overcome this problem without having to resort to using a more complex age-structured model (Schenzle, 1984) that would not be identifiable unless age-structured data were available. Finally, we demonstrate that a further improvement can be achieved by allowing the spatial influx of infected children to vary according to the epidemic status of neighbouring communities.

2. Data, model and inference

2.1. Data

We focus on weekly notifications of measles in 60 towns and cities in England and Wales as available from `www.zoo.cam.ac.uk/zoostaff/grenfell/measles.htm`. These are official notifications, taken from the Registrar General's *Weekly Reports*—more details of the data set are given in Keeling and Grenfell (1997), Grenfell and Harwood (1997) and Grenfell and Bolker (1998). The clearest epidemic dynamics are before the onset of measles vaccination in 1967. We therefore analyse the prevaccination data set for 20 years from the start of 1944. Local annual birth-rates and population sizes are taken from the *Annual Reports* of the Registrar General.

2.2. Model

As in our previous work the timescale is chosen to represent the exposed and infectious period, which for measles is known to be about 2 weeks (Black, 1984). At this timescale the number of exposed and infectious individuals is dependent only on the previous number of infected and susceptible children and the resulting model has a simple Markov structure. We have experienced that this approximation works well for measles. More generally the use of smaller timescales is possible (at the expense of a more complex model and a greater computational effort) and must always be adapted to the natural history of the disease.

Let $Y_t \geqslant 0$ denote the reported number of cases at time t, $t = 1, \ldots, n$. We also observe the number of individuals B_t who are recruited into the susceptible class in the tth time period. For childhood diseases such as measles this is well approximated by the number of births delayed by the time interval that infants are protected through maternally derived immunity. Finally we take D_t to be the estimated number of individuals who leave the population, which is estimated from records of total population size.

Three unobserved processes are required for the model specification.

(a) I_t is the true number of infected individuals at time point t. I_{t+1} depends stochastically on the transmission of the disease between I_t and S_t (see below).
(b) S_t is the size of the susceptible population at time point t. Given S_1 all future values of this population can be constructed from the I_t, D_t and B_t, i.e. the numbers leaving and joining the susceptible class.
(c) θ_t is the influx of infected individuals. θ_t adds to I_t to give the effective size of the infected population. If the disease has faded out it can only be reintroduced at time $t+1$ if $\theta_t > 0$. This process particularly plays a role in modelling the stochastic dynamics that are observed in smaller epidemic populations.

These processes are interrelated in the TSIR framework as follows:

$$\lambda_{t+1} = g(I_t + \theta_t, S_t, \phi_t), \tag{1}$$

$$I_{t+1} \sim f(I_{t+1}|\lambda_{t+1}, K_{t+1}), \tag{2}$$

$$S_{t+1} = S_t + B_{t+1} - I_{t+1} - \nu D_{t+1}, \tag{3}$$

$$\theta_t \sim f(\theta_t|\Theta_t), \tag{4}$$

$$Y_t \sim f(Y_t|I_t, \rho) = \text{Bin}(I_t, \rho), \tag{5}$$

where $f(\cdot)$ is a generic term for a density function with parameters being functions of the arguments. We refer to g as the *transmission equation* and the dynamics of the system are ultimately dependent on its functional form and the values of the (here seasonally varying) parameter vector ϕ_t. The number of infected individuals I_t is a non-negative random variable with *transmission distribution* $f(\lambda_t, K_t)$ with expectation λ_t and where K_t is some other sequence influencing transmission depending on some parameter $\mathbf{k}$. In our study K_t is the shape parameter of the transmission distribution. The influx distribution $f(\Theta_t)$ is also discrete and non-negative with expectation Θ_t. We assume a binomial *reporting distribution* (5) where each infected individual is reported with probability ρ.

In this formulation mortality of children due to the disease is assumed to be negligible. The term νD_t denotes (following rounding) the number of susceptible individuals who leave the population at each time point owing to mortality or migration. Since this term is likely to be very small in comparison with S_t and because D_t provides only a rough estimate of the reduction in population we take the dynamics of the susceptible class in equation (3) to be deterministic. The *influx process* θ_t is introduced to represent the number of infected individuals in the community who caught the disease from outside the community. Where these individuals were resident before contracting the disease is taken to be irrelevant.

Following the work of Fine and Clarkson (1982), Finkenstädt and Grenfell (2000) used the transmission equation

$$\lambda_{t+1} = r_t I_t^{\alpha} S_t \tag{6}$$

where r_t is a factor of proportionality that is time varying with a period of 1 year. For measles the infected and susceptible populations are largely comprised of school-children. Therefore 26 bi-weekly transmission parameters are estimated to capture the strong seasonal variation in contact rates over the school year. Taking the mixing parameter $\alpha = 1$ would correspond to homogeneous mixing of the infected and susceptible individuals (the assumption of mass action). Because interaction between the populations tends to be concentrated in schools this assumption may not be appropriate. For further discussion of how these parameters can be interpreted see Bjørnstad *et al.* (2002) and Fine and Clarkson (1982). In Finkenstädt and Grenfell (2000) the parameters were estimated by using linear least squares, taking logarithms of each side of equation (6). In the following section we introduce a rigorous and new approach for drawing inference in the more general TSIR model.

2.3. Inference

To estimate all parameters and states simultaneously of the model given in expressions (1)–(5) we use MCMC sampling. This involves simulation from the joint posterior density by setting up a Markov chain whose stationary distribution is equal to this target posterior density (see, for example, Gilks *et al.* (1995) for a review on MCMC methods). To derive the MCMC approach we use the following probabilistic representation of the model that clearly shows its three-stage hierarchy: the reporting process

$$Y_t \sim f(I_t, \rho), \tag{7}$$

the transmission process

$$I_t \sim f(I_{t-1}, \Phi, \theta_{t-1}) \tag{8}$$

and the influx process

$$\theta_t \sim f(\Theta), \tag{9}$$

where Φ indicates the complete set of parameters which influence transmission as part of both the predictor λ_t and the transmission distribution itself. This includes an estimate of the size of the susceptible population (we shall take the initial value S_1) and ν, the mean proportion of individuals leaving the population who are susceptible. Likewise, although the distribution of the influx process θ_t may vary with t, Θ denotes only the unknown parameters which influence this distribution.

We denote $\mathbf{I} = (I_1 I_2 \ldots I_n)^{\mathrm{T}}$ and define the process vectors $\mathbf{B}$, $\mathbf{D}$, $\mathbf{Y}$, $\mathbf{S}$ and $\boldsymbol{\theta}$ similarly. Using Bayes theorem and expressions (7)–(9) we have that the posterior distribution of all unobserved states and parameters given the data $(\rho, \mathbf{I}, \Phi, \boldsymbol{\theta}, \Theta)|\mathbf{Y}$ is proportional to

$$f(Y_1|I_1, \rho)\left\{\prod_{t=2}^{n} f(Y_t|I_t, \rho)\, f(I_t|I_{t-1}, \Phi, \theta_{t-1})\, f(\theta_{t-1}|\Theta)\right\} f(I_1)\, f(\rho)\, f(\Phi)\, f(\Theta)$$

where $f(I_1)$, $f(\rho)$, $f(\Phi)$ and $f(\Theta)$ are the prior distributions. From this we can derive that our MCMC scheme requires sampling in turn from the following conditional distributions:

$$\pi(\rho|\cdot) = \pi(\rho|\mathbf{Y}, \mathbf{I}) \propto \left\{\prod_{t=1}^{n} f(Y_t|I_t, \rho)\right\} f(\rho), \tag{10}$$

$$\pi(\Phi|\cdot) = \pi(\Phi|\mathbf{I}, \boldsymbol{\theta}) \propto \left\{\prod_{t=2}^{n} f(I_t|I_{t-1}, \Phi, \theta_{t-1})\right\} f(\Phi), \tag{11}$$

$$\pi(\theta_t|\cdot) = \pi(\theta_t|I_t, I_{t+1}, \Phi, \Theta) \propto f(I_{t+1}|I_t, \Phi, \theta_t)\, f(\theta_t|\Theta), \qquad t < n, \tag{12}$$

$$\pi(\Theta|\cdot) = \pi(\Theta|\boldsymbol{\theta}) \propto \left\{\prod_{t=1}^{n-1} f(\theta_t|\Theta)\right\} f(\Theta), \tag{13}$$

$$\pi(I_1|\cdot) = \pi(I_1|Y_1, \rho, I_2, \Phi, \theta_1) \propto f(Y_1|I_1, \rho)\, f(I_2|I_1, \Phi, \theta_1)\, f(I_1), \tag{14}$$

$$\pi(I_t|\cdot) = \pi(I_t|Y_t, \rho, \mathbf{I}_t^*, \Phi, \tilde{\boldsymbol{\theta}}_t) \propto f(Y_t|I_t, \rho)\left\{\prod_{j=t}^{n} f(I_j|I_{j-1}, \Phi, \theta_{j-1})\right\}, \qquad t > 1, \tag{15}$$

where $\mathbf{I}_t^* = (I_{t-1}\, I_{t+1}\, I_{t+2} \ldots I_n)^{\mathrm{T}}$ and $\tilde{\boldsymbol{\theta}}_t = (\theta_{t-1}\, \theta_t \ldots \theta_{n-1})^{\mathrm{T}}$. Note that the conditional distribution of I_t involves all future values of $\mathbf{I}$ and $\boldsymbol{\theta}$ since from equation (3) I_t influences all S_j, $j > t$.

The form of these conditional distributions may simplify depending on the model and the prior distributions that are chosen. For example, if the distribution describing the reporting process (7) is binomial as in expression (5) and the prior $f(\rho)$ is a beta distribution, then $\pi(\rho|\mathbf{Y}, \mathbf{I})$ will also have a beta distribution which we can sample directly. In the case that the conditional distribution cannot be sampled directly we use the Metropolis–Hastings algorithm. As an example, if we wish to sample from $\pi(x|\cdot)$ and the current state of the chain is x_1 we first choose a new value x_2 from some proposal density $q(x_1, x_2)$ and then accept x_2 with probability

$$\Delta(x_1, x_2) = \min\left\{\frac{\pi(x_2|\cdot)\, q(x_2, x_1)}{\pi(x_1|\cdot)\, q(x_1, x_2)}, 1\right\}.$$

The Metropolis–Hastings scheme is straightforward to implement and is suitable for sampling from any distribution. The MCMC algorithm has the following three steps.

Step 1: choose initial values for ρ, $\mathbf{I}$, $\mathbf{\Phi}$, $\boldsymbol{\theta}$ and $\mathbf{\Theta}$.
Step 2: update all these elements in turn by sampling from their conditional distributions as specified in expressions (10)–(15) by using the Metropolis–Hastings algorithm.
Step 3: once the Markov chain is deemed to have converged continue step 2 as many times as necessary to obtain the required number of realizations to approximate the marginal posterior distributions.

We have adopted the following guidelines.

(a) The starting values that are chosen for the state variables $\mathbf{I}, \boldsymbol{\theta}$ and S_1 must be realizable, i.e. $I_t \geqslant Y_t$, $S_t \geqslant I_{t+1}$ and $I_t + \theta_t = 0 \Rightarrow I_{t+1} = 0$. Any proposed move such that these rules are violated is rejected.
(b) To locate suitable starting values we use standard maximum likelihood estimation within the following scheme.
  (i) Taking $I_t = Y_t/\rho$ rearrange the transmission process in terms of Y_t and maximize the likelihood

$$\prod_{Y_{t-1}>0} f(Y_t | Y_{t-1}, \mathbf{\Phi}, \rho) \tag{16}$$

  with respect to $\mathbf{\Phi}$ and ρ. Although these parameter estimates are biased they provide good starting-points for estimation of the full model. Then form $\mathbf{I} = \mathbf{Y}/\hat{\rho}$ and $\mathbf{S}$ (rounding so that the series take integer values) using the preliminary parameter estimates.
  (ii) Keeping $\mathbf{I}$ and $\hat{\mathbf{\Phi}}$ fixed, sample from the conditional distributions of $\boldsymbol{\theta}$ and $\mathbf{\Theta}$ to obtain realistic starting values conditional on the other parameter estimates.
(c) We mostly use independent proper uniform priors for all the parameters in a similar way to Diggle *et al.* (1998). A consequence of this is that the resulting joint posterior distribution is proportional to the likelihood surface over the region specified.
(d) In practice we update each I_t and θ_t individually. Since they can take small integer values it is essential to use a discrete proposal distribution. Note that updating I_t alters all subsequent values of the susceptible series.
(e) We update $\mathbf{\Phi}$ and $\mathbf{\Theta}$ one component at a time. The choice of proposal distribution depends on the model and parameter in question. In the preliminary stages of setting up our algorithm we updated by using a normal random-walk Metropolis algorithm, which proved very robust. However, we recommend independence sampling following identification of a standard distribution which well approximates the conditional $\pi(x|\cdot)$ that we want to sample from. This is done by using a Newton–Raphson method to locate the mode of $\pi(x|\cdot)$ roughly and then using the second derivative at the mode to equate the form of $\pi(x|\cdot)$ to a distribution from which we can sample our proposal directly. The advantage of independence sampling is that a very high acceptance rate can be achieved without the chain being highly autocorrelated.

3. Inference using simulated data

In this section we investigate the performance of the MCMC algorithm by simulating two time series from the TSIR model representing *endemic* dynamics, where the disease never fades out, and *epidemic* behaviour, which is characterized by temporal extinction and recurrence of the disease. Endemic communities have a sufficiently large susceptible population that infection

can persist through epidemic troughs. The influx process of infected individuals therefore is not needed to explain the occurrence of new outbreaks and although it may exist it may be assumed to be negligibly small in comparison with the size of the local infected population (see Finkenstädt *et al.* (2002)). Indeed, the estimate of the rate of influx for the London measles series, using the method developed below, failed to converge owing to the insensitivity of the likelihood to the influx process.

The second time series is simulated to represent an epidemic community where the disease is subject to temporal extinction and can only recur if sparked off through one or more contacts of the local susceptible population with infectious individuals from other communities. In this case the process θ_t represents the spatial coupling of an epidemic community with other epidemic communities (possibly in the neighbourhood) which is an important issue in epidemic modelling and drawing inference for this kind of process from data on case notification has not been achieved yet. The specifications that are adopted here for the transmission function and distributions will be applied later to the observed measles time series.

3.1. Endemic series

Our first simulated series is designed to be comparable with the endemic dynamics that are observed for the London data, taking **B** to be the actual series of London birth-rates. The TSIR formulation that was chosen is

$$\lambda_{t+1} = r_t I_t^{\alpha} S_t, \tag{17}$$

$$I_t \sim \text{NegBin}(\lambda_t, k) = \frac{\Gamma(I_t + k)}{\Gamma(I_t + 1)\,\Gamma(k)} \left(\frac{\lambda_t}{k}\right)^{I_t} \left(1 + \frac{\lambda_t}{k}\right)^{-(I_t + k)}, \tag{18}$$

$$Y_t \sim \text{Bin}(I_t, \rho), \tag{19}$$

where the transmission parameter r_t has a period of 26 to capture the change in the rate of transmission over the school year. Therefore we wish to estimate ρ, $\mathbf{I}$ and $\mathbf{\Phi} = (\mathbf{r}^{\mathrm{T}}\, S_1\, k\, \alpha\, \nu)^{\mathrm{T}}$ where $\rho \in [0, 1]$ and $\mathbf{\Phi} \geqslant \mathbf{0}$. The majority of the computing time is used to compute the complicated conditional densities for each I_t. To prevent the chains for the parameters of the transmission equation (17) from becoming highly correlated it is useful to reparameterize to

$$\lambda_{t+1} = \frac{r_t^*}{\alpha} I_t^{\alpha} S_t,$$

where

$$r_t^* = \alpha r_t.$$

The proposal distributions for S_1, k, α, ν and each element of $\mathbf{r}$ were Gaussian, gamma, Gaussian, gamma and beta respectively, where the parameters were chosen at each iteration as described in point (e) of the previous section. The acceptance rate of the proposals was well above 90% for all the parameters, indicating that our algorithm is very efficient. To construct a suitable proposal for each I_t we use the fact that from distributions (18) and (19)

$$f(I_t | I_{t-1}, Y_t) \propto \frac{\Gamma(I_t + k)}{\Gamma(I_t - Y_t + 1)} \left\{\frac{(1-\rho)\lambda_t / k}{1 + \lambda_t / k}\right\}^{I_t} \Rightarrow f(Z_t) \propto \frac{\Gamma(Z_t + Y_t + k)}{\Gamma(Z_t + 1)} \left\{\frac{(1-\rho)\lambda_t / k}{1 + \lambda_t / k}\right\}^{Z_t},$$

where $Z_t = I_t - Y_t$, the full conditional for I_t also being a function of the subsequent values of $\mathbf{S}$ and therefore future values of $\mathbf{I}$. The distribution for Z_t is negative binomial, which we can sample from directly and then add Y_t to obtain a proposal for I_t. The average acceptance rate for these proposals was slightly below 90%.

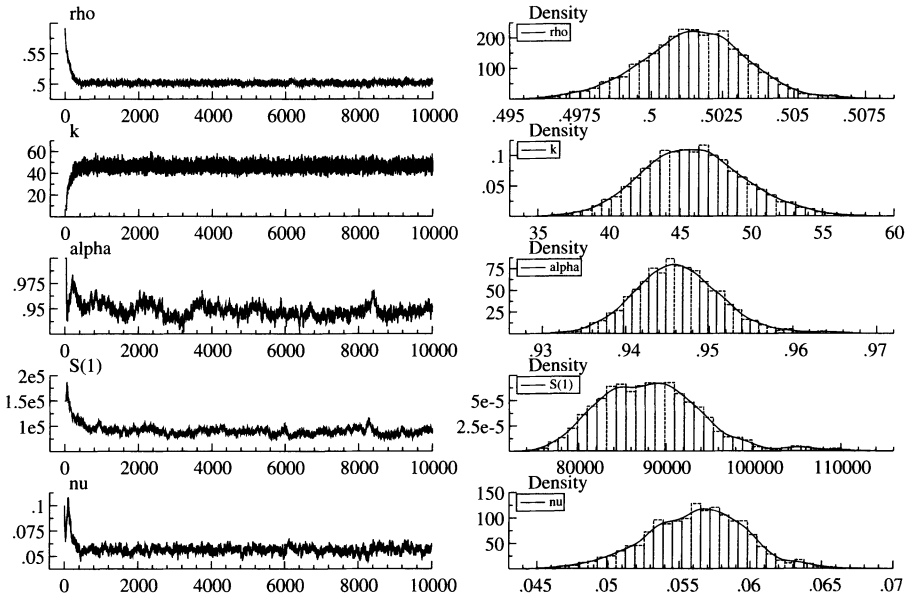

Fig. 2. Time series plots monitoring the MCMC output with histograms and kernel density estimates of the second 5000 samples for the first simulated series: the true parameter values are $\rho = 0.5$, $k = 50$, $\alpha = 0.95$, $S_1 = 100\,000$ and $\nu = 0.05$

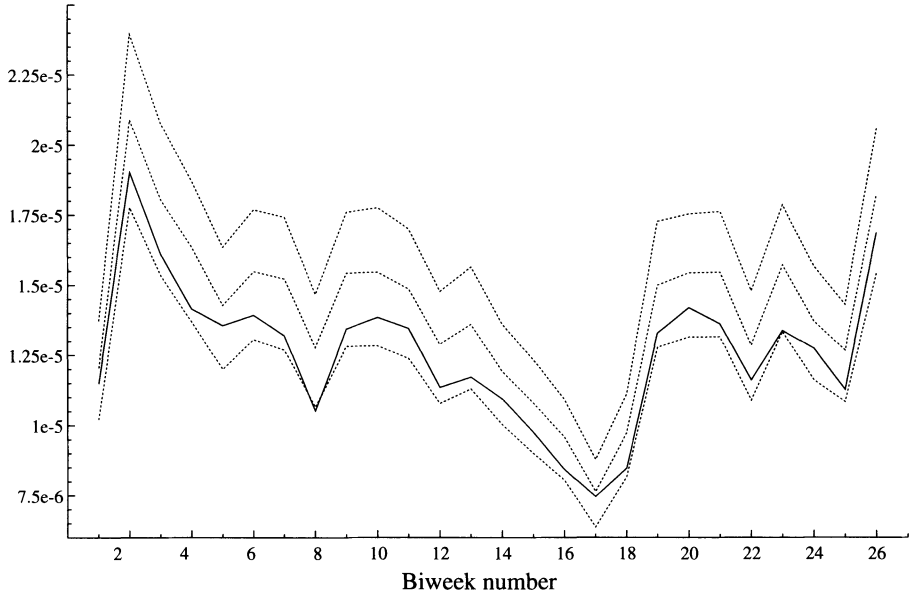

Fig. 3. Seasonal transmission parameter **r** (———) for comparison with the posterior means and 95% central credibility intervals (·····) for the first simulated series

The output for five of the parameters is shown in Fig. 2 and their true values are stated. We purposely chose very poor starting values but it appears that the chains converge fairly rapidly. Fig. 2 also shows estimates of the marginal posterior densities of these parameters obtained by sampling the second half of the chain. The kernel density estimates indicate that the marginal densities are all fairly symmetric. From Fig. 3 it appears that we can estimate the seasonal

variation of the transmission rate parameter accurately. The positive bias in estimation of the rate of transmission is balanced by negative bias in the estimate of the susceptible population with such a trade-off only being apparent when this population is so large that the model is partially insensitive to its exact value. Estimation for 10 further simulations showed a mean absolute cross-correlation between chains of close to 0.25 with the largest absolute mean cross-correlation of approximately -0.6 being between S_1 and $\bar{r}$. However, overestimation of ν, as apparent in Fig. 2, was not a general feature.

3.2. *Epidemic series*

For the second simulated series inference of the latent influx process must be incorporated. We chose the model

$$\lambda_{t+1} = r_t (I_t + \theta_t)^{\alpha} S_t, \tag{20}$$

$$I_t \sim \text{NegBin}(\lambda_t, K_t) = \frac{\Gamma(I_t + K_t)}{\Gamma(I_t + 1)\,\Gamma(K_t)} \left(\frac{\lambda_t}{K_t}\right)^{I_t} \left(1 + \frac{\lambda_t}{K_t}\right)^{-(I_t + K_t)}, \tag{21}$$

$$Y_t \sim \text{Bin}(I_t, \rho), \tag{22}$$

$$\theta_t \sim \text{Poisson}(\Theta), \tag{23}$$

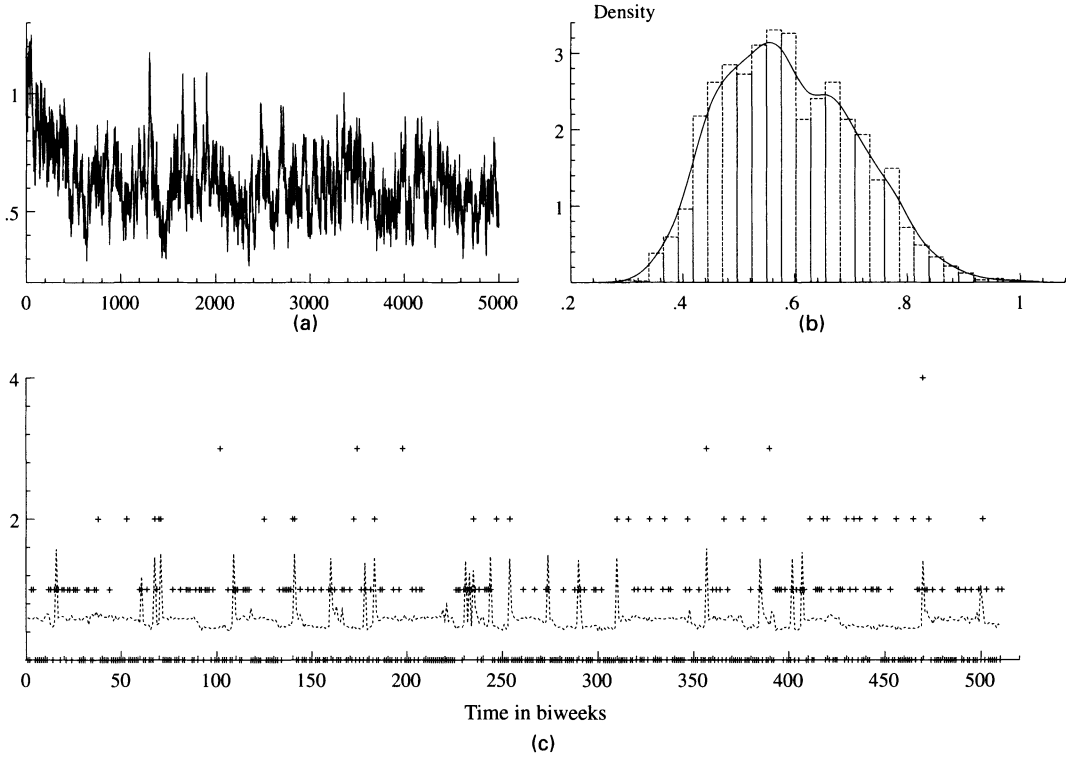

Fig. 4. Summary of MCMC output for the influx process for the second simulated series: (a) chain for Θ; (b) histogram and kernel density estimate (———) formed by sampling the second half of the chain for Θ; (c) influx of infected individuals θ_t in each time period with the posterior means (+, true θ_t; · · · · · ·, estimated θ_t)

where $K_t = k(I_{t-1} + \theta_{t-1})$ and r_t again has a period of 26. The parameters $\Phi = (\mathbf{r}^T\ S_1\ k\ \alpha\ \nu)^T$ were taken from preliminary estimation of the Bournemouth series whereas the reporting rate $\rho = 0.6$ and the influx rate $\Theta = 0.5$ were assigned arbitrarily. Although the conditional distributions are slightly different the MCMC algorithm is formulated as described for the first example. The cross-correlation between S_1 and $\bar{r}$ was -0.28 and further simulations showed that only when the susceptible population is very large, as in the first example, do the susceptible population and transmission rate parameters compensate for one another to the degree that estimated posterior distributions can be inconsistent with the true parameter values. Each θ_t was sampled by using the Metropolis–Hastings algorithm with a Poisson(Θ) proposal distribution. Provided that the prior is uniform the conditional distribution of Θ is gamma$(1 + \Sigma\,\theta_t, n - 1)$ and so can be sampled directly.

The estimated posterior distributions of the parameters were again fairly consistent with their true values and so we focus on estimation for the influx process. Figs 4(a) and 4(b) show the chain and posterior distribution for the rate of influx Θ. The reason for the large variance of the posterior becomes apparent from Fig. 4(c). Here we plot the posterior mean of $\boldsymbol{\theta}$ for comparison with its actual value. During periods when the influx process has no discernible effect, when either Y_t is large or remains 0, the mean estimate of θ_t is close to Θ. Only when $Y_t = 0$ and $Y_{t+1} > 0$ is θ_t estimated to be significantly different from Θ, thus resulting in a number of 'spikes' as observed in the plot.

Although it is recognized that extinction and recurrence of diseases are complex stochastic phenomena that are difficult to model, one problem that is not considered in theoretical epidemiology is the complication that arises due to measurement error. As the data are under-reported a natural question to ask is whether a zero reported case incidence represents a true fade-out or a case that was not reported. At this level both random variables are of equal size and thus measurement error cannot just be considered as a negligible process. Our MCMC algorithm can deal with this in a natural probabilistic way. We now illustrate the ability of the algorithm to impute the processes $\mathbf{I}$ and $\boldsymbol{\theta}$ by means of sample MCMC output. In Table 1 we list the true values of $\mathbf{Y}$, $\mathbf{I}$ and $\boldsymbol{\theta}$ for the time points $t = 60$–72 and the posterior means and standard deviations of the corresponding I_t and θ_t. The example shows the following.

Table 1. Comparison of the true values of the unobserved processes I_t and θ_t for $t = 60, \ldots, 72$ with their posterior means from the MCMC output for the second simulated series†

t	Y_t	I_t	*Estimate of* I_t	θ_t	*Estimate of* θ_t
60	3	3	5.37 (1.86)	1	0.55 (0.76)
61	0	0	0.74 (0.97)	1	1.18 (0.88)
62	1	2	1.68 (1.00)	0	0.50 (0.73)
63	0	0	0.11 (0.42)	0	0.50 (0.72)
64	0	0	0.03 (0.23)	1	0.50 (0.72)
65	0	2	0.03 (0.20)	0	0.54 (0.72)
66	0	0	0.04 (0.24)	0	0.51 (0.73)
67	0	0	0.04 (0.26)	0	0.58 (0.78)
68	0	0	0.10 (0.41)	2	1.52 (0.77)
69	2	2	3.31 (1.49)	1	0.51 (0.72)
70	0	0	0.24 (0.62)	2	0.51 (0.73)
71	0	1	0.14 (0.47)	2	1.53 (0.88)
72	15	20	24.24 (3.57)	0	0.64 (0.81)

†The values in parentheses are posterior standard deviations.

(a) The standard errors are large in comparison with the posterior means because the samples can take only integer values.
(b) The posterior mean of each I_t is close to Y_t/ρ but is also strongly dependent on I_{t-1} and I_{t+1}.
(c) The algorithm is quite conclusive that the infection is reintroduced when $t=68$ and $t=71$ on average, taking $\theta_t \geqslant 1$.
(d) $I_{65}=2$ but since $Y_{65}=0$ there is no information to suggest that the disease has been reintroduced. Consequently the algorithm usually takes $I_{65}=0$ and is not conclusive regarding a migration event at $t=64$.
(e) Cases are reported at $t=60$ and $t=62$ but $Y_{61}=0$. The algorithm cannot distinguish whether the presence of infection in the community at $t=62$ is due to a migration event or persistence of the disease unreported at $t=61$.

4. Inference for observed data

We now use our MCMC algorithm to identify TSIR models which can approximate the dynamics of the real time series of measles cases, choosing London and Bournemouth as representative time series for an endemic and epidemic community respectively.

4.1. Endemic measles series: London

We ran the MCMC algorithm of Section 3.1 for the London data, generating chains of length 10000. The initial values of each chain were obtained by using the direct likelihood method that is explained in Section 2. These proved particularly appropriate as the chains moved only

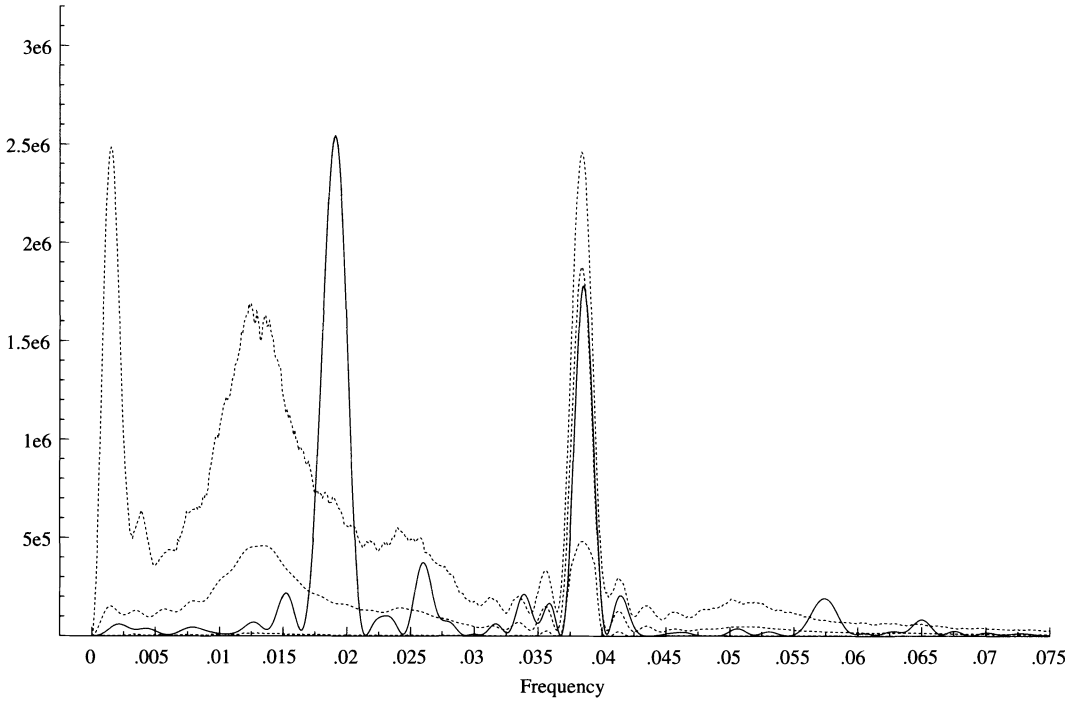

Fig. 5. Periodogram of the London measles series (———) with the mean and 95% envelopes (·····) from the periodograms of 1000 stochastic realizations generated from the estimated TSIR model without long-term variations in transmission

marginally from their starting-points. We sampled the second 5000 values from each chain to estimate the marginal posterior distributions. Use of the entire posterior distribution for inference incorporates the uncertainty in the estimation of both the stochastic and the systematic components of the model. However, it is convenient for now to summarize the marginal posterior distributions by their means.

We assess the ability of the models with the posterior means as parameter estimates to approximate the dynamics of the measles time series by simulation. In particular, we shall focus on how well the simulated trajectories and spectral densities match those of the observed data. The results are slightly disappointing as stochastic simulations appear to fluctuate more noisily than the data without exhibiting the same cyclical pattern. Evidence for this can be seen in Fig. 5 where we plot the periodogram of the square-root-transformed case counts together with the mean periodogram and 95% envelopes from 1000 stochastic realizations of the TSIR model. It appears that the simulations replicate the annual cyclicity of the data well but are less able to reproduce the stronger biennial cycle. A peak at zero frequency is indicative that the series is fading out. Results for other large cities were similarly poor with the model apparently not capable of detecting the biennial structure.

Problems occur because the structure of the susceptible population varies over time, which our TSIR formulation does not allow for. A period of high birth-rate results in several years where there is a large susceptible population because children may often have limited contact with infected children until they reach school age. This is dealt with in Finkenstädt and Grenfell (2000) by imposing a smoothly varying reporting rate that is evaluated from a local regression of cumulative cases on cumulative births designed so that $\mathbf{S}$ is stationary. Consequently, an increase in the birth-rate lowers the local reporting rate; as a result the system exaggerates the number of infected children, thus reducing the surplus of susceptible children. The transmission process therefore describes the interaction between *transformed* infected and susceptible populations.

Of course the actual consequence of a change in the structure of the susceptible population is to change the rate of transmission. Because the results of Finkenstädt and Grenfell (2000) were so promising we shall transfer their smoothly varying reporting rate into the dynamics of the transmission process. Given a reporting rate sequence ρ_t, $t = 1, \ldots, n$, we denote $I_t^* = Y_t/\rho_t$, $t = 1, \ldots, n$, and the corresponding susceptible series $\mathbf{S}^*$. If we take the transmission equation in terms of these transformed series to be equation (6) then we require the form of the transmission

Table 2. Summary of posterior estimates for the preferred models for London and Bournemouth†

Parameter	*Results for London series*		*Results for Bournemouth series*	
	Mean	*95% credibility interval*	*Mean*	*95% credibility interval*
α	0.950	(0.940, 0.960)	0.856	(0.829, 0.886)
k	29.90	(26.15, 33.85)	0.126	(0.111, 0.143)
$\bar{R}$ mean($\bar{S}$)	1.501	(1.265, 1.814)	1.794	(1.379, 2.294)
$\bar{S}$	93922	(87328, 117540)	3354	(2942, 3752)
ρ	0.4671	(0.4645, 0.4699)	0.613	(0.604, 0.622)
ν	0.0040	(0, 0.0098)	0.032	(0.004, 0.068)
e	0.132	(0.035, 0.230)	0.203	(0.060, 0.371)
$\bar{\theta}$	—	—	1.466	(1.003, 1.984)

†Note that the shape parameter k is defined differently for these two models.

for the 'true' infected series, $\mathbf{I} = \mathbf{Y}/\bar{\rho}$, and the resulting susceptible series $\mathbf{S}$. It is easy to show that (see Appendix A)

$$\lambda_{t+1} = \left\{ r_t \rho_t^e \left(1 + \frac{C_t}{S_t} \right) \right\} I_t^\alpha S_t = R_t I_t^\alpha S_t, \tag{24}$$

where $e = 1 - \alpha$, $C_1 = 0$ and

$$C_t = \sum_{j=2}^{t} \frac{Y_j(\rho_j - \bar{\rho})}{\bar{\rho}\rho_j}, \qquad t \geqslant 2.$$

The approach that we now take therefore is to use a transmission equation of the above form, first evaluating the sequence ρ_t, $t = 1, \ldots, n$, as shown in Finkenstädt and Grenfell (2000) and taking e as an additional parameter to estimate. The advantage of this method is that our approximation is at least as good as those obtained in the previous references but we can now obtain estimates of a constant ρ, ν and realistic infected and susceptible series.

Incorporating the long-term variations in transmission involved only minor modifications of our MCMC algorithms. The initial values for the chains were now obtained by maximizing likelihood (16) with respect to transmission function (24). We note that this maximized log-likelihood increases from -3047 for the original model to -2983 with the majority of the increase due to the final factor of R_t. Table 2 shows posterior means and 95% central credibility intervals for the parameters. We plot the product of the slowly varying and seasonally varying rates of transmission in Fig. 6. The annual pattern is similar to that plotted in Fig. 3 and is dependent

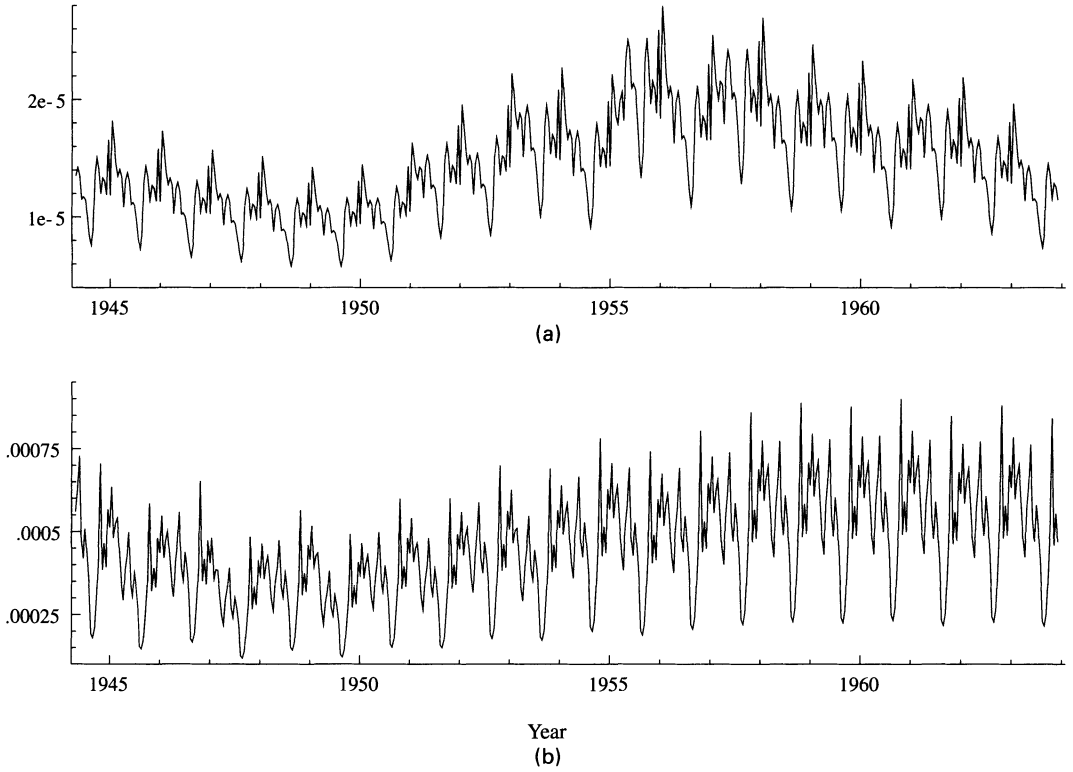

Fig. 6. Estimated rates of transmission R_t, $t = 1, \ldots, n$, for the two measles series (a) London and (b) Bournemouth

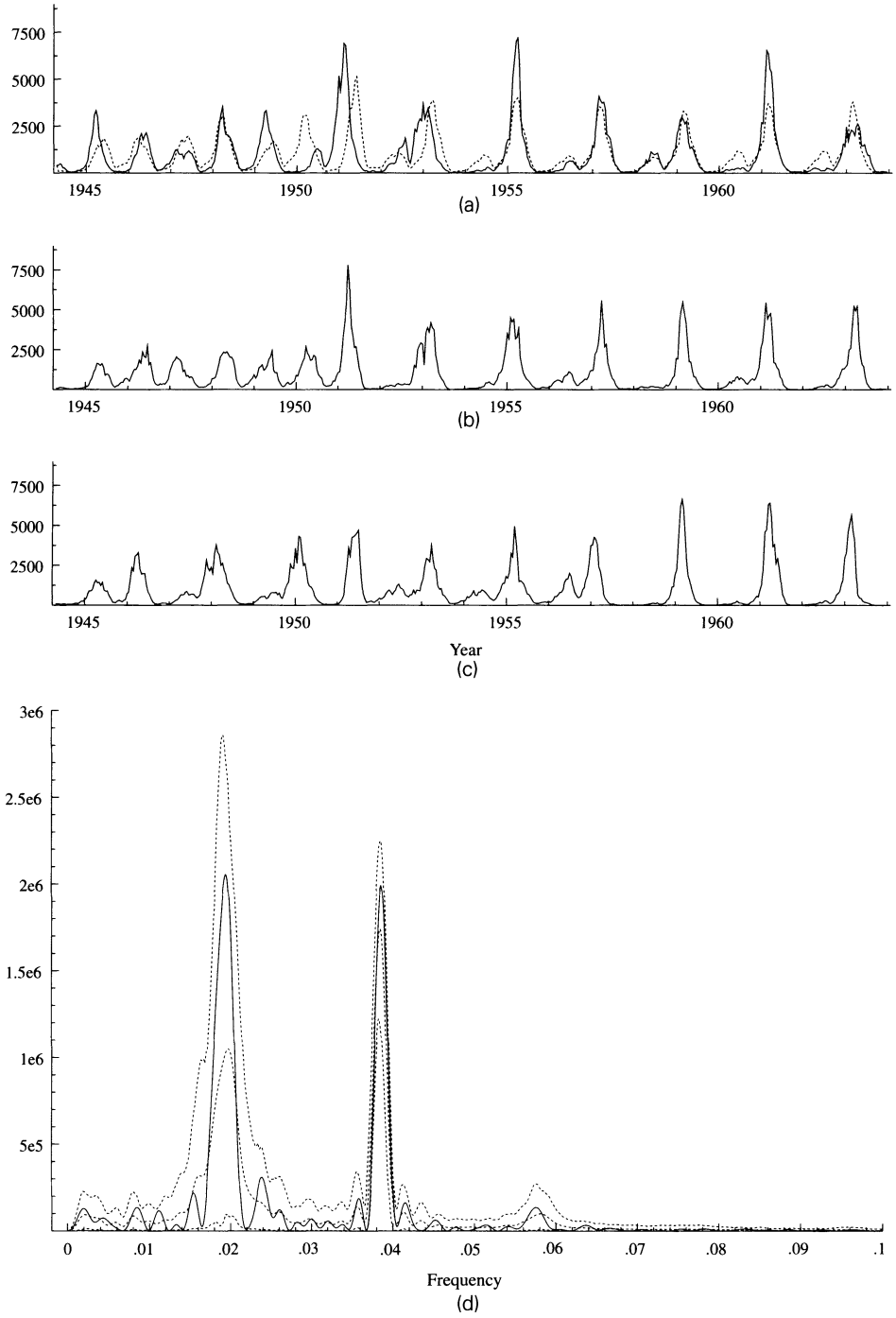

Fig. 7. (a) Bi-weekly incidence of measles in London (———) together with a forecast from $t = 0$ by using the estimated TSIR model with long-term variations in transmission (· · · · · · ·), (b), (c) stochastic realizations from the approximating TSIR model and (d) periodogram of the measles incidence series (———) with the mean and 95% envelopes from the periodograms of 1000 stochastic realizations generated from this model (· · · · · · ·)

on school terms as discussed by Finkenstädt and Grenfell (2000) with the prominent dip during bi-weeks 16–18, corresponding to the long school summer holidays, being particularly evident. Referring to the plot of the birth-rate in Fig. 1 we see that the transmission rate is lower following periods of high birth-rate as expected. Figs 7(a)–7(c) show that stochastic realizations from the approximating TSIR model now very much resemble the observed series and the simulations appear to reproduce the transition from biennial to annual epidemics. The expected trajectory from $t = 0$ shows that the model appears to predict the future pattern of the major and minor epidemics. We summarize 1000 stochastic realizations from the model by the mean of their periodograms in Fig. 7(d). It is clear that the simulations do now often exhibit biennial cyclicity.

4.2. Epidemic measles series: Bournemouth

We ran the MCMC algorithm for the Bournemouth data, now using a transmission equation of the form (24) but with θ_t introduced as in equation (20). In Fig. 8 we compare the periodogram of the root-transformed data with the mean periodogram from 1000 simulated series. It is obvious that the simulations do not exhibit the biennial cycles that are present in the data.

The deficiencies of our model can be attributed to the constant rate influx process (23). It is more realistic to assume that the influx of infected individuals into Bournemouth reflects the state of the epidemic in the surrounding area. Therefore we take

$$\theta_t \sim \text{Poisson}(c_t \Theta)$$

where the sequence c_t is designed to account for this variation and Θ is a constant to be estimated. Finkenstädt and Grenfell (1998) considered various geographical covariates which could

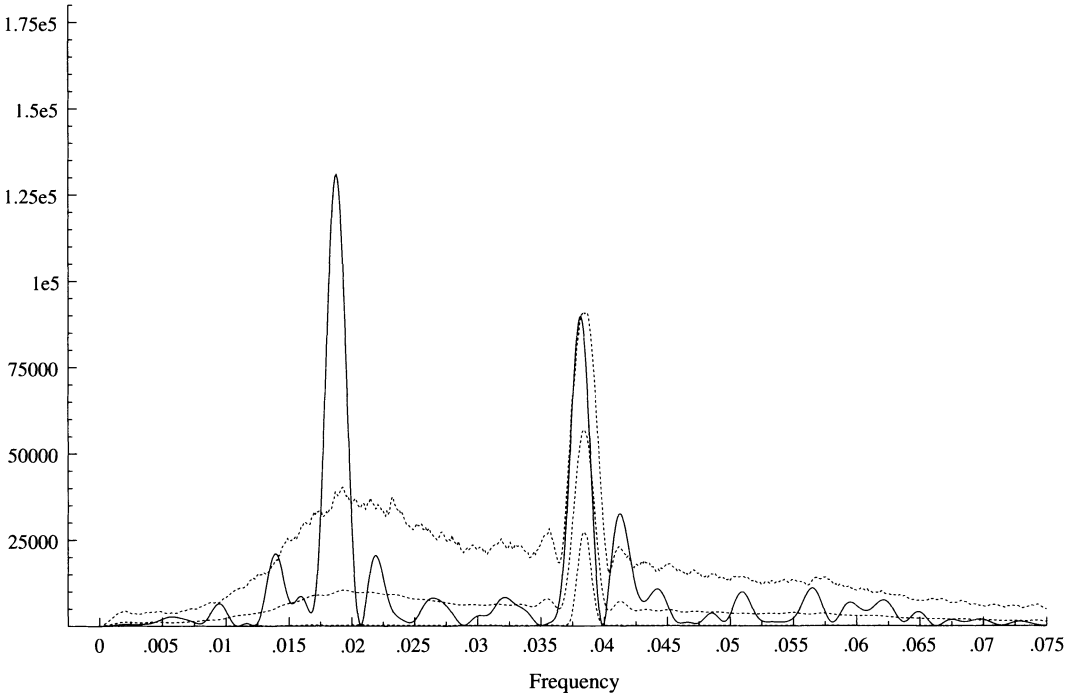

Fig. 8. Periodogram of the Bournemouth measles incidence series (———) with the mean and 95% envelopes from the periodograms of 1000 stochastic realizations generated from the approximating TSIR model with constant influx rate (·····)

be used to predict epidemic behaviour from the incidence data in other communities. Here we shall incorporate the interaction simply by taking c_t to be the number of reported cases in London, the closest large city to Bournemouth. Following very minor modifications to our MCMC algorithm we obtained parameter estimates for the TSIR model with a time-varying rate of influx. Table 2 shows the posterior means with 95% central credibility intervals for the parameters and the transmission rates are as shown in Fig. 6. We summarize 1000 stochastic realizations from the model by the mean of their periodograms in Fig. 9(d) and plot two of these realizations for comparison with the observed data in Figs 9(a)–9(c). It is now conceivable that the measles series could have been generated by the approximating model, demonstrating the importance of a realistic influx pattern. Accurate long-term prediction for the Bournemouth series is unrealistic because the epidemic clockwork in small communities is more sensitive to demographic stochasticity and the stochastic influx of infection. However, the total number of cases during each 2-year period is more accurately predicted than the precise timing of the epidemics.

5. Discussion

We have considered a general model for the practical analysis of observed disease incidence time series for diseases which conform to the SIR paradigm. The algorithm that was developed not only permits inference about the transmission of disease but also inference regarding the rate of under-reporting and the influx of infection. The potential for realistic imputation and parameter estimation was demonstrated by using simulated data.

The use of MCMC methods is critical for model inference. The performance of our algorithm is very satisfactory. Although we only summarize the results for two communities we have estimated TSIR models for more than 30 measles series and experienced few difficulties in convergence of the chains, even when using poor starting values. We have assessed convergence from visual and time series analysis of the sample output although more formal methods are increasingly being developed (see Cowles and Carlin (1996)). In most cases, 10000 iterations of the chain were deemed sufficient to obtain a representative sample from the target distribution. The 10000 iterations took approximately 40 min on a 2.4 GHz laptop personal computer using the OX language of Doornik (1999) although the algorithm runs more quickly for smaller communities which have longer periods of zero incidence.

The paper provides an approach to drawing inference for the TSIR models but there are undoubtedly alternative ways to construct the MCMC algorithm and to analyse the results. A fully Bayesian analysis involving specification of informative priors is easily incorporated in the current formulation. We should mention that the specification of uniform priors had little effect on the convergence of the MCMC scheme and resulting inference provided that they were non-zero for all realistic parameter values. The motivation for simply using the posterior means to summarize posterior information was to discriminate clearly between the competing models in Section 4. The distributions that were used to approximate the reporting, transmission and influx processes are also subject to the modeller's discretion. The transmission process that was used here is very similar to that in Finkenstädt and Grenfell (2000) although we have used more flexible transmission distributions. It is straightforward to compare different mathematical formulations of the transmission process for endemic series by calculating the Akaike information criterion following maximization of the likelihood (16) for non-zero Y_{t-1}. An investigation of different long-term and short-term transmission processes and the development of methods for model comparison for the full model are in progress.

The application to time series of reported measles cases in Section 4 revealed the importance

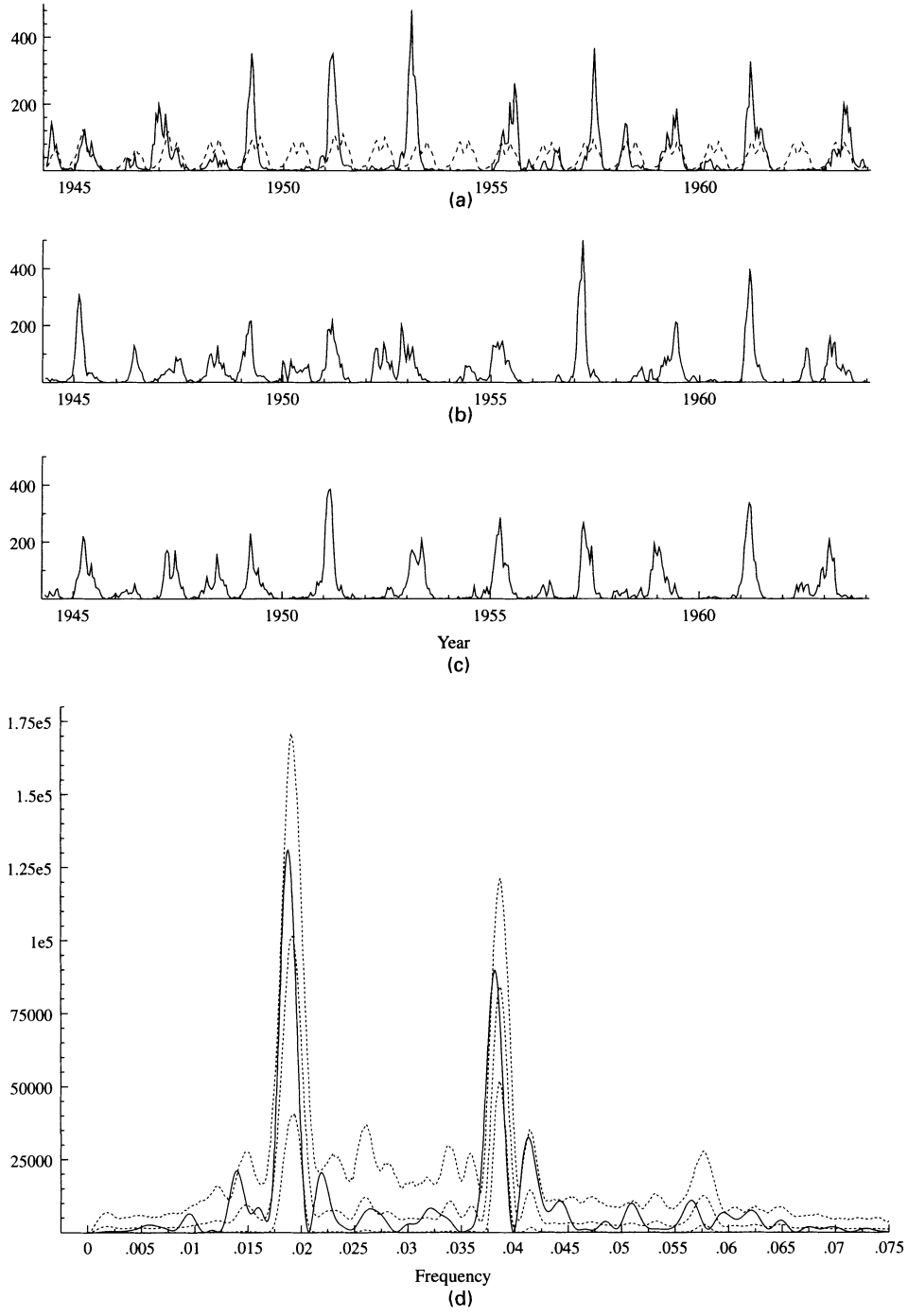

Fig. 9. (a) Bi-weekly incidence of measles (———) in Bournemouth together with a forecast from $t = 1$ by using the estimated TSIR model with time-varying influx rates (-------), (b), (c) stochastic realizations from the approximating TSIR model and (d) periodogram of the measles incidence series (———) with the mean and 95% envelopes from the periodograms of 1000 stochastic realizations generated from this model (·····)

of accounting for changes in the age structure of the susceptible population. The smoothly varying rates of transmission that were used for this are a complicated function of the previous numbers of cases and births in the community and it is straightforward to optimize the rates by varying the bandwidth of the local regression. Ideally we would want to assess whether a model that explicitly accounts for age structure leads to better predictability of childhood diseases. This has been suggested in theoretical models (Schenzle, 1984) but has not been tested with data. However, because age-structured data for childhood diseases are not usually available it is equally important to study how we can allow for age structure in a simpler model that does not require age-structured data. The approach that was taken here provides a step in this direction.

In summary, we have provided a stochastic model

(a) that uses mechanistic rules from epidemic theory but is of sufficiently modest structure that the parameters can be estimated from time series data on reported cases and births,
(b) that allows for measurement error due to under-reporting of cases and for the random extinction and recurrence of disease and
(c) which produces remarkably realistic time series simulations.

This approach therefore also attempts to bridge a gap between mechanistic modelling from theoretical epidemiology, which is often too complex to be confronted with data, and parameter estimation from disease incidence time series.

Appendix A: Deriving the long-term variation in rates of transmission

Let $I_t = Y_t/\bar{\rho}$ and $I_t^* = Y_t/\rho_t$, $t = 1, \ldots, n$, and denote the corresponding susceptible series $\mathbf{S}$ and $\mathbf{S}^*$ respectively. Taking $S_1 = S_1^*$ from equation (3) it is easy to see that

$$S_t^* = S_t + \sum_{j=2}^{t} (I_j - I_j^*) = S_t + \sum_{j=2}^{t} \frac{Y_j(\rho_j - \bar{\rho})}{\bar{\rho}\rho_j}. \tag{25}$$

Since the transmission equation (6) implies that

$$E(I_{t+1}^*) = \lambda_{t+1}^* = r_t^* (I_t^*)^\alpha S_t^*, \qquad t > 1,$$

on substituting $I_t^* = (\bar{\rho}/\rho_t^*) I_t$, $\forall t$, we obtain

$$E(I_{t+1}) = \lambda_{t+1} = \rho_t^{1-\alpha} \bar{\rho}^{\alpha-1} r_t^* I_t^\alpha S_t^*, \qquad t > 1.$$

Taking $r_t = \bar{\rho}^{\alpha-1} r_t^*$ and substituting equation (25) we obtain equation (24) as required.

References

Bartlett, M. S. (1957) Measles periodicity and community size. *J. R. Statist. Soc.* A, **120**, 48–60.

Bartlett, M. S. (1960) The critical community size for measles in the United States. *J. R. Statist. Soc.* A, **123**, 37–44.

Bjørnstad, O. N., Finkenstädt, B. F. and Grenfell, B. T. (2002) Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecol. Monogr.*, **72**, 169–184.

Black, F. L. (1984) Measles. In *Viral Infections of Humans: Epidemiology and Control* (ed. A. S. Evans), pp. 397–418. New York: Plenum.

Cowles, M. K. and Carlin, B. P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Am. Statist. Ass.*, **91**, 883–904.

Dietz, K. (1995) Some problems in the theory of infectious disease transmission and control. In *Epidemic Models: Their Structure and Relation to Data* (ed. D. Mollison). Cambridge: Cambridge University Press.

Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998) Model-based geostatistics (with discussion). *Appl. Statist.*, **47**, 299–350.

Doornik, J. A. (1999) *An Object-orientated Matrix Programming Language, Ox 3.0*. London: Timberlake Consultants.

Feng, Z. and Thieme, H. R. (1995) Recurrent outbreaks of childhood diseases revisited: the impact of isolation. *Math. Biosci.*, **128**, 93–130.
Fine, P. E. M. and Clarkson, J. A. (1982) Measles in England and Wales: I, An analysis of factors underlying seasonal patterns. *Int. J. Epidem.*, **11**, 5–14.
Finkenstädt, B. F., Bjørnstad, O. N. and Grenfell, B. T. (2002) A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks. *Biostatistics*, **3**, 493–510.
Finkenstädt, B. F. and Grenfell, B. T. (1998) Empirical determinants of measles metapopulation dynamics in England and Wales. *Proc. R. Soc. Lond.* B, **265**, 211–220.
Finkenstädt, B. F. and Grenfell, B. T. (2000) Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Statist.*, **49**, 187–205.
Finkenstädt, B. F., Keeling, M. J. and Grenfell, B. T. (1998) Patterns of density dependence in measles dynamics. *Proc. R. Soc. Lond.* B, **265**, 753–762.
Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds) (1995) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
Grenfell, B. T. (1992) Chance and chaos in measles dynamics. *J. R. Statist. Soc.* B, **54**, 383–398.
Grenfell, B. T., Bjørnstad, O. N. and Finkenstädt, B. F. (2002) Dynamics of measles epidemics: scaling noise, determinism and predictability with the TSIR model. *Ecol. Monogr.*, **72**, 185–202.
Grenfell, B. T. and Bolker, B. M. (1998) Cities and villages: infection hierarchies in measles epidemics. *Ecol. Lett.*, **1**, 63–70.
Grenfell, B. T., Bolker, B. M. and Kleczkowski, A. (1995) Seasonality, demography and the dynamics of measles in developed countries. In *Epidemic Models: Their Structure and Relation to Data* (ed. D. Mollison). Cambridge: Cambridge University Press.
Grenfell, B. T. and Dobson, A. P. (1995) *Ecology of Infectious Diseases in Natural Populations*. Cambridge: Cambridge University Press.
Grenfell, B. T. and Harwood, J. (1997) (Meta)population dynamics of infectious diseases. *Trends Ecol. Evol.*, **12**, 395–399.
Keeling, M. J. and Grenfell, B. T. (1997) Disease extinction and community size: modeling the persistence of measles. *Science*, **275**, 65–67.
McLean, A. R. and Anderson, R. M. (1988a) Measles in developing countries: Part I, Epidemiological parameters and patterns. *Epidem. Infectn*, **100**, 111–133.
McLean, A. R. and Anderson, R. M. (1988b) Measles in developing countries: Part II, The predicted impact of mass vaccination. *Epidem. Infectn*, **100**, 419–442.
Nasell, I. (2002) Measles outbreaks are not chaotic. In *Mathematical Approaches for Emerging and Reemerging Infectious Diseases* (ed. C. Castillio-Chavez), pp. 85–115. New York: Springer.
Rhodes, C. J., Jensen, H. J. and Anderson, R. M. (1997) On the critical behaviour of simple epidemics. *Proc. R. Soc. Lond.* B, **264**, 1639–1646.
Schenzle, D. (1984) An age-structured model of pre- and post-vaccination measles transmission. *IMA J. Math. Appl. Med. Biol.*, **1**, 169–191.