Bayesian Analysis of Severe Acute Respiratory Syndrome: The 2003 Hong Kong Epidemic

Phenyo E. Lekone*
Mathematics Department, University of Botswana, Private Bag 0022, Gaborone, Botswana

Received 30 October 2007, revised 3 March 2008, accepted 19 March 2008

Summary
This paper analyzes data arising from a Severe Acute Respiratory Syndrome (SARS) epidemic in Hong Kong in 2003 involving 1755 cases. A discrete time stochastic model that uses a back-projection approach is proposed. Markov Chain Monte Carlo (MCMC) methods are developed for estimation of model parameters. The algorithm is further extended to integrate numerically over unobserved variables of the model. Applying the method to SARS data from Hong Kong, a value of 3.88 with a posterior standard deviation of 0.09 was estimated for the basic reproduction number. An estimate of the transmission parameter at the beginning of the epidemic was also obtained as 0.149 with a posterior standard deviation of 0.003. A reduction in the transmission parameter during the course of the epidemic forced the effective reproduction number to cross the threshold value of one, seven days after control interventions were introduced. At the end of the epidemic, the effective reproduction number was as low as 0.001 suggesting that the epidemic was brought under control by the intervention measures introduced.

Key words: Back-projection; Bayesian inference; Control Intervention; SARS epidemics.

1 Introduction
Severe Acute Respiratory Syndrome, commonly known as SARS, is a new viral respiratory illness caused by a coronavirus, called SARS-associated coronavirus (SARS-CoV) (Leung et al., 2004). The disease was first identified in China’s southern province of Guangdong on 16 November, 2002 and then spread rapidly throughout the world via air travel with the last known case with a symptom onset date of 5 July identified in Taiwan (Chowell et al., 2003). It is reported that the epidemic infected 8098 cases (with 774 deaths), from a total of 28 countries (Chau and Yip, 2003a). A large proportion of the morbidity and the mortality was borne by Hong Kong with a total number of 1755 cases and 302 deaths occurring between 15 February 2003 and 31 May 2003 (Leung et al., 2004). Some of the measures introduced by the Hong Kong authorities to combat the disease include among others, public service announcements, addition of SARS to the list of notifiable diseases, a 2-week suspension of schools, introduction of health declarations for incoming residents and visitors, isolation of cases and home quarantining of close contacts and restrictions on their travel out of the country (Chau and Yip, 2003a).

Using discrete time data on an ongoing SARS epidemic in Hong Kong, Donnelly et al. (2003) estimated the mean and variance of the incubation time, time from admission to discharge and time from admission to death. These parameters were obtained using maximum likelihood estimation methods assuming a gamma distribution for each period with allowance for censoring due to incomplete observation. The estimate of the mean and variance of the incubation period was 6.37 days and 16.69 days, respectively. The estimated mean and variance of the incubation period was 6.37 days and 16.69 days, respectively.
572.9 days, respectively, and the estimated mean and variance of the time from admission to discharge was 23.5 days and 62.1 days, respectively. As pointed out by Leung et al. (2004), parameters above suffer from right censoring since they were obtained using data from an ongoing epidemic. To address the problem of censoring above, the authors re-estimated parameters above using complete data on all the 1755 cases. The data was completed by using laboratory verification tests on 83.6% of the cases. They estimated the mean incubation period to be 4.6(3.99) days (standard errors in brackets). They also estimated that the time from onset to recovery, and the time from onset to death each followed a gamma distribution with mean 26.5(14.0) and 23.7(14.9) days, respectively (standard errors in brackets). Using parameters estimated by Donnelly et al. (2003) and weekly data on an ongoing SARS epidemic in Hong Kong, Riley et al. (2003) estimated the basic reproduction number ($R_0$) from a discretized stochastic meta-population SEIR model as 2.7 with a 95% confidence interval of (2.2,3.7). It is evident that estimates obtained by Leung et al. (2004) are different from those obtained Donnelly et al. (2003) and as such the estimate of 2.7 above may be biased as it was obtained using estimates obtained with right censored data. It is imperative that a better estimate of the basic reproduction number would be obtained by using estimates obtained by Leung et al. (2004) are different from those obtained Donnelly et al. (2003) and as such the estimate of 2.7 above may be biased as it was obtained using estimates obtained with right censored data. It is imperative that a better estimate of the basic reproduction number would be obtained by using estimates obtained by Leung et al. (2004) are different from those obtained Donnelly et al. (2003) and as such the estimate of 2.7 above may be biased as it was obtained using estimates obtained with right censored data. It is imperative that a better estimate of the basic reproduction number would be obtained by using estimates obtained by Leung et al. (2004). By studying the exponential growth at the beginning of the epidemic and by estimating the generation time, Lipsitch estimated $R_0$ to be between 2.2 and 3.6. Using a likelihood-based approach on an ongoing epidemic, Wallinga and Tenius (2004) estimated $R_0$ to be between 2.2 and 3.7. McBryde et al. (2006) applied Bayesian inference techniques to estimate the basic reproduction number and other epidemiological determinants from data arising from a SARS epidemic in Shanzi province of China. They estimated $R_0$ to be 4.8 with a 95% credible interval of (2.2, 8.8).

In this paper, a slight variation of the model by Riley et al. (2003) is formulated. The model is used in conjunction with gamma distributions estimated by Leung et al. (2004) and SARS data from Hong Kong that is well documented by Leung et al. (2004). Interest lies in estimating the transmission parameter, the basic reproduction number, and assessing whether control interventions were successful in controlling the epidemic or not. Markov chain Monte Carlo methods are used for numerical integration over the distribution of unobserved variables which in this article is the series of unobserved new cases. The modelling approach used is different from the one by McBryde et al. (2006) and uses a different data set.

The paper is organized as follows; Epidemiological features of SARS and the Hong Kong data are discussed in the next Section with the model presented in Section 3. The Bayesian inference algorithm using Markov Chain Monte Carlo methods is discussed in Section 4 with applications to simulated data and observed Hong Kong data discussed in Section 5. A discussion of the paper is given in Section 6.

### 2 SARS Epidemiology and Data

The disease is transmitted via close person-to-person contact with infected individuals (Chowell et al., 2003). In general, SARS begins with a high fever, with other symptoms including headache, an overall feeling of discomfort, and body aches. Some people also have mild respiratory symptoms at the outset while most patients develop pneumonia. An individual exposed to SARS becomes infectious after an incubation period of 2–7 days (Chowell et al., 2003). Once infectious, individuals remain so for 7–10 days before they recover, although about 4% or more die (Chowell et al., 2003). Circumstantial evidence suggests that, in most cases, infectiousness coincides with clinical symptoms (Leung et al., 2004; Riley et al., 2003).

Here we study data from the 2003 SARS outbreak in Hong Kong which is well documented by Leung et al. (2004). The data consist of three time series (see Figure 1) recorded from February 15 to
July 24, namely daily counts of SARS cases by date of symptom onset (1755 cases), daily counts of deaths from SARS (302 deaths) and daily recoveries from SARS (1410 recoveries).

It is also documented that the first case of SARS in Hong Kong showed symptoms on 15 February, 2003 and was subsequently hospitalized on 22 February (Leung et al., 2004). The last patient with SARS first experienced symptoms on 31 May and was hospitalized on 2 June (Leung et al., 2004). On June 23, the World Health Organization removed Hong Kong from its list of areas with recent local transmission of SARS (Leung et al., 2004). It is suggested by Chau and Yip (2003a) that control interventions were initially introduced on or around the 10th of March when there was a formal recognition of an outbreak in a hospital, with more measures being put in place thereafter. The epidemic lasted for about 128 days with control measures being introduced about 30 days after the start of the epidemic.

3 Model

In this section, a slight variation of the model by Riley et al. (2003) is considered. Riley et al. (2003) used a discretized stochastic metapopulation SEIR model with a gamma distribution (formulated as a sum of exponential distributions) used for the waiting time in each disease compartment. The model considered here is a single population version of the model of Riley et al. (2003) with a back-projection method (Brookmeyer and Gail, 1986), used instead of a gamma distribution represented as a sum of independent exponential random variables. The single population model is motivated by the complete SARS data discussed in the previous section and by the fact that it requires less detail to specify and less data for inference as compared to the model of Riley et al. (2003). A description of the model is as follows; assume that an individual is susceptible (S) until exposed to infection. Once infected an individual goes through an exposure stage (E) where they are not infectious followed by an infectious (I) stage where they are classified into two categories \( I^- \) and \( I^+ \). The classification is based on whether an infectious individual will proceed to die \( (R^-) \) with probability \( p \) or proceed to recover \( (R^+) \) with probability \( 1 - p \). Let \( T_E, T_+ \) and \( T_- \) represents the time spent in class \( E, I^+ \) and
I −, respectively. Let \( f_t = P(T_E = t) \), \( g_t^+ = P(T_+ = t) \) and \( g_t^- = P(T_− = t) \) be probability distributions for the time spent in each disease class. Let \( \mu_E \) and \( \sigma_E^2 \) represent the mean and variance of \( T_E \), respectively and define \( \mu_+^+, \sigma_+^2, \mu_− \) and \( \sigma_−^2 \), similarly. Consider a time interval \((t, t + h]\) where \( h \) represents the time interval at which measurements are taken, here \( h = 1 \) day. Let \( B(t), C(t), D^+(t) \) and \( D^−(t) \) denote the number of individuals exiting class \( S, E, I^+ \) and \( I^− \) during the time interval \((t, t + h]\), respectively.

Let \( \beta(t) \) be the time dependent infection rate and define \( P(t) = 1 - \exp \left[ -\frac{\beta(t)}{N} h(t) \right] \), where \( I(t) = I^+(t) + I^−(t) \) and \( N \) is the population size, as the probability that a susceptible individual will get infected in \((t, t + h]\). The rationale for the form of \( P(t) \) is given in Lekone and Finkenstädt (2006). Let \( R(t) \) be the number of cases who recover or die at time \( t \). Given initial conditions \( S(0) = s_0, E(0) = e \) and \( I(0) = a \), the discrete time stochastic model is then given by

\[
\begin{align*}
S(t + h) &= S(t) - B(t), \\
E(t + h) &= E(t) + B(t) - C(t), \\
I(t + h) &= I(t) + C(t) - D^+(t) - D^−(t), \\
S(t) + E(t) + I(t) + R(t) &= N
\end{align*}
\]

where

\[
B(t) \sim \text{Bin}(S(t), P(t)), \quad C(t) \sim \text{Pois}(\lambda_t), \quad D^+(t) \sim \text{Pois}(\eta_t^+), \quad D^−(t) \sim \text{Pois}(\eta_t^-)
\]

with

\[
\lambda_t = \sum_{i=0}^7 f_i B(t - i), \quad \eta_t^+ = p \sum_{i=0}^7 g_i^+ C(t - i), \quad \eta_t^- = (1 - p) \sum_{i=0}^7 g_i^- C(t - i).
\]

Notations Bin \((n, q)\) and Pois \((\mu)\) represent a binomial distribution with parameters \(n\) and \(q\) (success probability) and a Poisson distribution with mean \(\mu\), respectively. The back-projection approach (Brookmeyer and Gail, 1986) is used in the previous equation to construct the means for the Poisson distributions. This approach has been shown to be successful in epidemic modelling (Becker, Watson and Carlin, 1991; Becker und Xu, 1994) and particularly in reconstructing SARS epidemics (Chau and Yip, 2003a, b).

Introduction of control interventions will usually reduce the transmission parameter \(\beta(t)\). Capturing the change in \(\beta(t)\) as the epidemic progresses is made difficult by the fact that there is usually little or no information to be used to quantify the changes in the transmission parameter. Riley et al. (2003) assumed for simplicity that \(\beta(t)\) took three different values on three equally spaced time periods over the whole observation period. Choosing the number of change points may be difficult and as such modelling \(\beta(t)\) as described above may be difficult to implement. Usually, the time at which control interventions are introduced is known. To avoid the problem of change points, it is assumed that the transmission parameter \(\beta(t)\) is constant up to the time point when the control measures are introduced and after that decays exponentially. This can be formulated as follows

\[
\beta(t) = \begin{cases} 
\beta, & t < t_0 \\
\beta e^{−q(t−t_0)}, & t \geq t_0
\end{cases}
\]

where \(t_0\) is the time point at which control measures are introduced, \(\beta\) is the initial transmission rate and \(q > 0\) is the rate at which \(\beta(t)\) decays for \(t > t_0\). The same form of \(\beta(t)\) has been used successfully by Lekone and Finkenstädt (2006) to model control interventions in Ebola epidemics and a similar form has been applied by Chowell et al. (2004). It should however be mentioned that this simple form of \(\beta(t)\) may not capture the heterogeneity in SARS epidemics brought about by super-spreader events (rare events when an individual may generate many more than the average number of secondary cases). Thus, \(\beta(t)\) is used here for simplicity. The basic reproduction number \(R_0\) is defined as the average number of secondary cases generated by a primary case over his/her infectious period when introduced into a large population of susceptible individuals (Diekmann, Heesterbeek and Metz, 1990). The constant \(R_0\) thus measures the initial growth rate of the epidemic and for the model above
it can be shown that (see Riley et al., 2003)

\[ R_0 = \beta [p \mu_ - + (1 - p) \mu_+] \tag{8} \]

The time dependent effective reproduction number \( R_0(t) = \beta(t) [p \mu_ - + (1 - p) \mu_+] \) is the number of secondary cases per infectious case at time \( t \) and the time point at which \( R_0(t) \) assumes values smaller than one indicates when control measures have become effective in controlling the epidemic.

4 Bayesian Inference

Let \( \tau \) denote the end of the observation period and let \( B = \{B(t)\}_{t=0}^{T} \) represent the time series of \( B(t) \) from the beginning to the end of the observation period and define \( C, D^+ \) and \( D^- \) similarly. The epidemic model specified in (1) to (6) together with the contact rate (7) has parameter vector \( \Theta = \{\beta, q\} \) which we would like to estimate from knowledge of initial conditions, population size and from observation of \( \{C, D^+, D^-\} \) and from knowledge of \( p \) and the parameters of the distributions of \( T_E, T_-, T_+ \). Consideration of these data is motivated by the observed SARS data discussed in the previous section.

The time series \( \{I(t)\} \) is in this case fully determined by iterating (3) from the initial condition \( s_0 \). In reality, the series \( B \) is not observed. If \( B \) was observed, then the series \( \{S(t)\} \) would be determined by iterating (1) from the initial condition \( s_0 \), thus allowing for the maximum likelihood (ML) estimate of \( \Theta \) to be obtained via maximizing the likelihood

\[ \prod_t \text{Bin}(B(t), S(t), P(t)) \tag{9} \]

where \( \text{Bin}(x, n, p) \) represents a binomial distribution with parameters \( n \) and \( p \) and evaluated at the point \( x \). The unobserved series \( B \) is imputed using MCMC methods which allow for numerical integration over the distribution of the unobserved process. Proposals for \( B \) are generated and accepted according to a Metropolis Hastings step similar to the one described in Lekone and Finkenstädt (2006). Let \( \Omega \) be the set containing means and variances of distributions of \( T_E, T_-, T_+ \). By using the fact that \( B(t), C(t), D^+(t) \) and \( D^-(t) \) are conditionally independent gives the augmented likelihood for the data as

\[ L(B, C, D^+, D^-, \Omega \mid \Theta) = \prod_{t=0}^{T} \text{Bin}(B(t), S(t), P(t)) \text{Po}(C(t), \lambda_c) \text{Po}(D^-(t), \eta^-) \text{Po}(D^+(t), \eta^+) \tag{10} \]

where \( \text{Po}(x, \mu) \) represents a Poisson distribution with mean \( \mu \) evaluated at the point \( x \).

Multiplying (10) by the prior \( \pi(\Theta) \) gives, up to a constant of proportionality, the posterior distribution

\[ \pi(\Theta, B \mid C, D^+, D^-, \Omega) \propto L(B, C, D^+, D^-, \Omega \mid \Theta) \pi(\Theta) \tag{11} \]

that we wish to sample from. An MCMC algorithm that samples in turn from the conditional distributions \( \pi(B \mid C, D^+, D^-, \Omega, \Theta) \) and \( \pi(\Theta \mid B, C, D^+, D^-, \Omega) \) produces draws from the desired \( \pi(\Theta, B \mid C, D^+, D^-, \Omega) \). Preliminary analysis showed that \( a \text{ posteriori} \), the parameters \( \beta \) and \( q \) were correlated and this problem was overcome by updating them as a block rather than as single-site updates.

5 Applications to Data

5.1 Application to simulated data

The performance of the algorithm is tested by applying it to a simulated time series from the model (1)–(7) with parameter values \( s_0 = 6700000 \), \( e = 1 \), \( a = 0 \), \( \beta = 0.2 \), \( q = 0.1 \), \( r = 40 \), \( T_E \sim \Gamma(5, 16) \).
$T_+ \sim \Gamma(27, 200), T_- \sim \Gamma(24, 225), p = 0.2$ and $h = 1$. The notation $\Gamma(b, c)$ represents a gamma distribution with mean $b$ and variance $c$. A discrete approximation was used for the gamma distributions. These parameter values are, as far as possible, tuned to the data as they are motivated by earlier studies on SARS and by what is known about its natural history (Chowell et al., 2003; Leung et al., 2004). The simulated epidemic resulted in a final size (total number of susceptible individuals who get infected during the course of the epidemic) of 731 with 155 of the cases being fatal and 557 having recovered by the end of the epidemic after $t = 122$ days.

The MCMC algorithm is applied to the simulated data pretending that $B$ is unobserved. Initially, the performance of the back-projection method in reconstructing the series is inspected. In Figure 2 the simulated series $C,D^+$ and $D^-$ are shown along with their means estimated via the back-projection method. Figure 2 suggests that the back-projection method produces sensible means in all the cases and hence a Poisson distribution with this mean will be a reasonable way of rejecting/accepting

![Figure 2](image-url)

Figure 2  Actual number of new infectious cases, new deaths and new removals for the simulated data along with their mean values obtained using the back-projection method.

Informative gamma priors are included in the second column for comparison. A cross in each graph indicates the ML-estimate. The true parameter values are 0.2 for $b$, 0.1 for $q$ and 5.28 for $R_0$.

![Figure 3](image-url)

Figure 3  Posterior distributions for $b$, $q$ and $R_0$ for the simulated data (first column). A solid and a dotted line (first column) indicates a posterior distribution obtained using a gamma and a uniform prior distribution, respectively. Informative gamma priors are included in the second column for comparison. A cross in each graph indicates the ML-estimate. The true parameter values are 0.2 for $\beta$, 0.1 for $q$ and 5.28 for $R_0$. 

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim  
www.biometrical-journal.com
proposals for \( B \). Inference is then conducted using two parameter sets for the prior distribution of \( b \) and \( q \). At first, a non-informative prior set \( \{b \sim U(0,1), q \sim U(0,1)\} \) is used to run the MCMC algorithm followed by an informative prior set \( \{b \sim \Gamma(0.2,0.002), q \sim \Gamma(0.1,0.0005)\} \). The notation \( U(a,b) \) represents a uniform distribution with parameters \( a \) and \( b \). In each case, after discarding a burn-in period of 10000, a sample of size 1000 taken every 1000 iterations of the chain was used to obtain posterior distributions. Convergence was verified using convergence diagnostic measures and the chains were deemed to have converged.

Figure 3 shows the estimated posterior distributions of the parameters obtained for different prior cases. Posterior means and standard deviations are reported in Table 1. In all the two cases the Markov chain appeared to have converged after the burn-in period. The posterior means are well in agreement with the ML-estimate obtained from the complete data. Table 1 suggests that the estimation method is not very sensitive to the choice of prior information, as different choices result in parameter estimates that are in agreement with the ML-estimate. The acceptance rate for the sampler for \( B \) was around 70%.

In order to assess model fit, the final size of the epidemic was chosen as the discrepancy measure. The posterior predictive distribution of final sizes obtained from the Markov chain run did not show any evidence of lack of fit (posterior \( p \)-value = 0.491). A good reference for posterior predictive distributions and posterior \( p \)-values is Gelman, Meng and Stern (1996).

### 5.2 Application to observed SARS data

The algorithm described above was then applied to the real SARS data set where the unobserved process \( B \) was imputed. It is assumed that there was initially one exposed individual and that the very first case was exposed for five days using the approximate mean incubation period according to Leung et al. (2004). This assumption implies that the outbreak started on February 10. It can be deduced from Chau and Yip (2003a) that control interventions were introduced on or about 30 days from the start of the epidemic implying that \( r^* = 30 \). From the data and the estimation of Leung et al. (2004), it follows that

| Parameter | Estimate |
|-----------|----------|
| \( \beta \) | 0.149(0.003) |
| \( q \)   | 0.0901(0.0001) |
| \( R_0 \) | 3.88(0.09) |
\[ \mu_E = 4.6, \sigma_E^2 = 15.9, \mu_s = 26.5, \sigma_s^2 = 196, \mu_c = 23.7, \sigma_c^2 = 222, \ p = 0.172 \text{ and } N = 6.7 \text{ million.} \]

An estimate for \( p \) was obtained as the total number of deaths divided by the sum of deaths and recoveries. Values for \( R_0 \) obtained from other studies (Wallinga and Tenius, 2004; McBryde et al., 2006) together with Eq. (8) suggest that an informative value for \( \beta \) would be around 0.15. Gamma priors with parameters \((0.15, 0.0015), (0.1, 0.005)\) for \( f, g \) were therefore chosen to be informative.

After discarding a burn-in period of 100,000, a sample of size 1000 taken every 1000 iterations of the chain was used to obtain posterior distributions and convergence monitored via autocorrelation functions. Results obtained are summarized in Table 2 and in Figure 4. Results obtained with a vague prior on \( \beta \) and \( q \) were similar to the ones reported in Table 2 and are therefore not reported.

Results from Table 2 show that the estimated value of \( R_0 \) is around 3.88 with a posterior standard deviation of 0.09. This is larger than the value of 2.7 estimated by Riley et al. (2003) but smaller than the value of 4.8 estimated by McBryde et al. (2006). The estimate is of the same magnitude as the value of 3.6 estimated by Wallinga and Tenius (2004). The fact that a more precise value of \( R_0 \) is obtained in this study than was previously obtained from other studies could possibly be explained by the following reasons. Firstly, parametric forms of \( f, g^+ \) and \( g^- \) already estimated by other scholars are used and secondly unlike in the case of Riley et al. (2003) where only one time series \( \{ C(t) \} \) was used, this time series along with two other time series are used, thus improving the data availability. Moreover, the use of the back-projection for rejection sampling purposes gives the model more detail as this method has been shown to successfully reconstruct SARS epidemics. An estimate for the transmission parameter \( \beta \) was obtained as 0.149 with a posterior standard deviation of 0.003. This estimate is consistent with a daily infectivity at the beginning of the epidemic of 0.15 obtained by McBryde et al. (2006). The model predicts that the effective reproduction number crosses the threshold value of one, seven days after control interventions were introduced. At the end of the epidemic, the effective reproduction number is as low as 0.001 suggesting that the epidemic was brought under control by the intervention measures. A posterior predictive distribution of final sizes shown in Figure 5(b) does not indicate any evidence of lack of fit (p-value is 0.512). The fit of the model is also evident from Figure 5(a) where the predicted number of cases is consistent with the observed cases.
The Bayes factor computed following the method of Chib and Jeliazkov (2001) was used to compare the proposed model with the model whereby \( f, g^r \) and \( g^\nu \) were assumed to follow a geometric distribution with the same mean as the gamma distributions used here. The geometric distribution is used here as in Lekone and Finkenstädt (2006) as a discrete approximation to the exponential distribution. The theoretical details of the scheme used to calculate the marginal likelihood of each model can be found in Section 2.3 of Chib and Jeliazkov (2001). A Bayes factor of 5.23 in favour of the model with gamma distributions was obtained.

6 Discussion

A simple model that captures the form of distributions of epidemiological determinants has been introduced to estimate the basic reproduction number and to assess the effect of control interventions introduced during the course of the epidemic. MCMC methods were then used to carry out Bayesian inference allowing for integration over latent variables. The performance of the inference method was demonstrated by applying it to simulated data and to data arising from an outbreak of SARS in Hong Kong in 2003. Results obtained for the basic reproduction number from the SARS data are slightly larger than those obtained by Riley et al. (2003) and Lipsitch et al. (2003) but in agreement with those obtained by Wallinga and Tenius (2004). The model also predicts that the control measures introduced during the course of the epidemic were successful in bringing it under control. This was demonstrated by the effective reproduction number crossing the threshold value of one before the end of the epidemic.

Whilst it is appreciated that epidemic models for SARS should incorporate the heterogeneity due to super-spread events, it would be difficult to infer from such models since they would require detailed data which is not readily available in practice. In the case that such data are available, modelling heterogeneity by incorporating it into the form of \( \beta(t) \) or by using a meta-population model as in Riley et al. (2003) will be valuable. The model used here uses three time series of data that can read-
ily be obtained by health practitioners in case of an outbreak and as such is easy to implement. However, better modelling ways of quantifying control interventions are desirable. The back-projection method was in this article used solely to construct an efficient proposal for the Metropolis-Hastings algorithm in order to impute the unobserved process \( B(t) \). This was motivated by the fact that the method reconstructed the SARS epidemic quite well as was shown by Chau and Yip (2003a, b).

The MCMC method can be modified to incorporate variability on the parameters of the gamma distributions used. These are assumed to be fixed in this paper, an assumption that might not be realistic as these are observed with error in real life situations. Computation of the Bayes factor can be quite difficult to compute especially in models with high number of parameters (Chib and Jeliazkov, 2001). However, this difficulty was not encountered in this article as only two parameters \( \beta \) and \( q \) were of interest. In the case that a variation of the model proposed here leads to a high dimensional parameter vector, then blocking methods as suggested by Chib and Jeliazkov (2001) can be used. The modelling approach proposed in this article is easy to implement and requires data that is not difficult to collect and hence can be applied to data sets arising from diseases with similar dynamics to SARS.

Implementation of the methods described in this paper was done with a modern standard computer using the OX language of Doornik (1999). Computation time was less than six hours for each of the Markov chain runs.

Acknowledgements The author would like to thank the University of Warwick (Warwick postgraduate fellowship), the United Kingdom’s national Overseas Research Scheme (ORS) and University of Botswana for jointly funding this research. The manuscript benefited quite a lot from reports by two anonymous reviewers and an associate editor.

Conflict of interests statement
The authors have declared no conflict of interest.

References
Becker, N. G., Watson, L. F., and Carlin, J. B. (1991). A method of non-parametric back-projection and its application to AIDS data. *Statistics in Medicine* **10**, 1527–1542.
Becker, N. G. and Xu, C. (1994). Dependent HIV incidences in back-projection of AIDS incidence data. *Statistics in Medicine* **13**, 1945–1958.
Brookmeyer, R. and Gail, M. H. (1986). Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *Lancet* **2**, 1320–1322.
Chau, P. H. and Yip, P. S. F. (2003a). Monitoring the severe acute respiratory syndrome epidemic and assessing effectiveness of interventions in Hong Kong Special Administrative Region. *Journal of Epidemiology and Community Health* **57**, 766–769.
Chau, P. H. and Yip, P. S. F. (2003b). Modelling the severe acute respiratory syndrome. *Hong Kong Medical Journal* **9**, 310–311.
Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* **96**, 270–281.
Chowell, G., Fenimore, P. W., Castillo-Garsow, M. A., and Castillo-Chavez, C. (2003). SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism. *Journal of Theoretical Biology* **224**, 1–8.
Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., and Hyman, J. M. (2004). The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. *Journal of Theoretical Biology* **229**, 119–126.
Diekmann, O., Heesterbeek, J. A., and Metz, J. A. (1990). On the definition and the computation of the basic reproduction ratio \( R_0 \) in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology* **28**, 365–382.
Donnelly, C. A., Ghani, A. C., Leung, G. M., Hedley, A. J., Fraser, C., Riley, S., Abu-Raddad, L. J., Ho, L., Thach, T., Chau, P., Chan, K., Lam, T., Tse, L., Tsang, T., Liu, S., Kong, J. H. B., Lau, E. M. C., Ferguson, N. M., and Anderson, R. M. (2003). Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *The Lancet* **361**, 1761–1766.
We refer you to our previous publications for further details on the methods and models we have used in our analysis. For instance, the posterior predictive assessment of model fitness via realized discrepancies is discussed in Gelman, A., Meng, X. L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 6, 733–807.

Lekone, P. E. and Finkenstädt, B. F. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics* 62, 1170–1177.

Leung, G. M., Hedley, A. J., Ho, L., Chau, P., Wong, I. O. L., Thach, T. Q., Ghani, A. C., Donnelly, C. A., Fraser, C., Riley, S., Ferguson, N. M., Anderson, R. M., Tsang, T., Leung, P., Wong, V., Chan, J. C. K., Tsui, E., Lo, S., and Lam, T. (2004). The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: an analysis of all 1755 patients. *Annals of Internal Medicine* 141, 662–673.

Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., Gopalakrishna, G., Chew, S. K., Tan, C. C., Samore, M. H., Fisman, D., and Murray, M. (2003). Transmission dynamics and control of severe acute respiratory syndrome. *Science* 300, 1966–1970.

McBryde, E. S., Gibson, G., Pettitt, A. N., Zhang, Y., Zhao, B., and McElwain D. L. S. (2006). Bayesian modelling of an epidemic of severe acute respiratory syndrome. *Bulletin of Mathematical Biology* 68, 889–917.

Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abbu-Raddad, L. J., Hedley, A. J., Leung, G. M., Ho, L., Lam, T., Thach, T. Q., Chau, P., Chan, K., Lo, S., Leung, P., Tsang, T., Ho, W., Lee, K., Lau, E. M. C., and Ferguson, N. M. (2003). Transmission dynamics of the aetiological agent of severe acute respiratory syndrome (SARS) in Hong Kong: the impact of public health interventions. *Science* 300, 1961–1966.

Wallinga, J. and Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology* 160, 509–516.