Efficient real-time monitoring of an emerging influenza epidemic: how feasible?

Paul J. Birrell¹, Daniela De Angelis¹,², Lorenz Wernsich¹, Brian D. M. Tom¹, Leonhard Held³, Gareth O. Roberts⁴, Richard G. Pebody²

July 3, 2018

¹MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, UK
²Public Health England, London, UK
³Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Switzerland
⁴Centre for Research in Statistical Methodology, University of Warwick, Coventry, UK

e-mail for correspondence: daniela.deangelis@mrc-bsu.cam.ac.uk

Abstract

A prompt public health response to a new epidemic relies on the ability to monitor and predict its evolution in real time as data accumulate. The 2009 A/H1N1 outbreak in the UK revealed pandemic data as noisy, contaminated, potentially biased, and originating from multiple sources, seriously challenging the capacity for real-time monitoring. Here we assess the feasibility of real-time inference based on such data by constructing an analytic tool combining an age-stratified SEIR transmission model with various observation models describing the data generation mechanisms. As batches of data become available, a sequential Monte Carlo algorithm is developed to synthesise multiple imperfect data streams, iterate epidemic inferences and assess model adequacy amidst a rapidly evolving epidemic environment, substantially reducing computation time to ensure timely delivery of real-time epidemic assessments.

KEYWORDS: Sequential Monte-Carlo, Resample-Move, real-time inference, pandemic influenza, SEIR transmission model

1 Introduction

A pandemic influenza outbreak has the potential to place a significant burden upon healthcare systems. Therefore, the capacity to monitor and predict the evolution of an epidemic as data progressively accumulate is a key component of preparedness strategies for prompt public health response.

Statistical inferential approaches have been used in a real-time monitoring context for a number of infectious diseases. Examples include: prediction of swine fever cases in a classical framework Meester et al. (2002); online estimation of a time-evolving effective reproduction number \( R(t) \) for SARS (Wallinga and Teunis, 2004; Cauchemez et al., 2006) for generic emerging disease (Bettencourt and Ribeiro, 2008); and Bayesian inference on the transmission dynamics of avian influenza in the UK poultry industry (Jewell et al., 2009).
These models rely on the availability of direct data on the number of new cases of an infectious disease over time. In practice, as illustrated by the 2009 outbreak of pandemic A/H1N1pdm influenza in the United Kingdom (UK), direct data are seldom available. More likely, multiple sources of data exist, each indirectly informing the epidemic evolution, each subject to possible sources of bias. These sources of data typically come from routine influenza surveillance systems reporting interactions with healthcare services, which means they are often: only a glimpse of the most severe cases; they are subject to the healthcare-seeking behaviours of the population; contaminated with cases of people experiencing influenza-like illness; and can be heavily influenced by governmental advice. This calls for more complex modelling, requiring the synthesis of information from a range of data sources in real time.

In this paper we tackle the problem of online inference and prediction in an influenza pandemic in this more realistic situation. We address this by developing the work of Birrell et al. (2011) who retrospectively reconstructed the A/H1N1 pandemic in a Bayesian framework using multiple data streams collected over the course of the pandemic. In Birrell et al. (2011) posterior distributions of relevant epidemic parameters and related quantities are derived through Markov Chain Monte Carlo (MCMC) methods which, if used in real-time, pose important computational challenges. MCMC is notoriously inefficient for online inference as it requires repeat browsing of the full history of the data as new data accrue. This motivates a more efficient algorithm. Potential alternatives include refinements of MCMC (e.g. Jewell et al., 2009; Banterle et al., 2015) and Bayesian emulation as in Farah et al. (2014), where the model is replaced by an easily-evaluated approximation readily prepared in advance of the data assimilation process. Here, we explore Sequential Monte Carlo (SMC) methods (Doucet and Johansen, 2009). As batches of data arrive at times $t_1, \ldots, t_K$, SMC techniques allow computationally efficient online inference by combining the posterior distribution $\pi_k(\cdot)$ at time $t_k$, $k = 0, \ldots, K$ with the incoming batch of data to obtain an estimate for $\pi_{k+1}(\cdot)$. A further advantage of SMC is that it naturally provides all the necessary posterior predictive distributions to make one-step ahead probabilistic forecasts of the incoming data. In an epidemic context, monitoring the appropriateness of the chosen model is vital to avoid making public health decisions on the basis of mis-specified models. Through formal assessment of the quality of these one-step ahead forecasts, continual, timely checks of model adequacy can be made (Held et al., 2017).

Use of SMC in the real time monitoring of an emerging epidemic is not new. Dureau et al. (2013), Camacho et al. (2015), Dukic et al. (2012), Ong et al. (2010), and Skvortsov and Ristic (2012) are examples of real time estimation and prediction for deterministic and stochastic models describing the dynamics of influenza and Ebola epidemics. These models, again, only include a single source of information that has either been pre-smoothed or is free of any sudden or systematic changes.

In what follows we advance existing literature in three ways: we include a number of data streams, realistically mimicking current data availability in the UK; we consider the situation where a public health intervention introduces a shock to the system, critically disrupting the ability to track the posterior distribution over time; and we demonstrate how the use of SMC can facilitate online assessment of model adequacy.

The paper is organised as follows: in Section 2 the model in Birrell et al. (2011) is reviewed focusing on the data available and the computational limitations of the MCMC algorithm in a real time context; in Section 3 the idea of SMC is introduced and the algorithm of Gilks and Berzuini (2001) is described; Section 4 discusses the types of epidemic predictions required in real-time; in Sections 5 and 6 results are presented from the application of Gilks and Berzuini’s SMC algorithm to data simulated to mimic the 2009 outbreak and illustrate the challenges posed by the presence of the informative observations induced by system shocks; in Sections 7
and 8 adjusted SMC approaches that address such challenges are assessed; we conclude with 
Section 9 in which the ideas explored in the paper are critically reviewed and outstanding issues discussed.

2 A Model For Pandemic Reconstruction

Birrell et al. (2011) describe the transmission of a novel influenza virus among a fixed popu-
lation stratified into A age groups and the subsequent reporting of infections. This is achieved 
through using a deterministic age-structured Susceptible (S), Exposed (E), Infectious (I), Re-
covered (R) transmission model. To remove the memoryless property of the waiting times in 
the E and I states, these states are split into two sub-states, $E_1$ and $E_2$, $I_1$ and $I_2$. The dynamics 
of the system are governed by a system of differential equations:

$$
\begin{align*}
\frac{dS(t,a)}{dt} &= -\lambda(t,a)S(t,a) \\
\frac{dE_1(t,a)}{dt} &= \lambda(t,a)S(t,a) - \frac{2}{d_L}E_1(t,a) \\
\frac{dE_2(t,a)}{dt} &= \frac{2}{d_L}(E_1(t,a) - E_2(t,a)) \\
\frac{dI_1(t,a)}{dt} &= \frac{2}{d_L}E_2(t,a) - \frac{2}{d_I}I_1(t,a) \\
\frac{dI_2(t,a)}{dt} &= \frac{2}{d_I}(I_1(t,a) - I_2(t,a))
\end{align*}
$$

where $d_L$ and $d_I$ are the mean latent and infectious periods respectively. Transmission is driven 
by the time- and age-varying rate $\lambda(t,a)$ at which susceptible individuals become infected. The system in (1) is not chaotic and in practice is evaluated using an Euler approximation at times 
$t_k = k\delta t, k = 0, \ldots, K$. The choice of $\delta t = 0.5$ days is sufficiently small that the probability 
of more than one jump per period is negligible. Under this discretisation, at time $t_k$ the vector 
$(S_{t_k,a}, E_{1t_k,a}, E_{2t_k,a}, I_{1t_k,a}, I_{2t_k,a})$ gives the number of individuals in age group $a \ (a = 1, \ldots, A)$ 
in each disease state. At this time the expected number of new infections is $S_{t_k-1,a}\lambda_{t_k-1,a}$, where 

$$
\lambda_{t_k,a} = 1 - \prod_{b=1}^{A} \left\{ (1 - M_{t_k}^{a,b}R_0(\psi) / d_I)^{I_{1t_k,b} + I_{2t_k,b}} \right\} \delta t. \tag{2}
$$

Here, $R_0(\psi)$ is the basic reproduction number, the expected number of secondary infections 
caused by a single primary infection in a fully susceptible population, parameterised in terms of 
the epidemic growth rate $\psi$. The pattern of transmission between age groups is determined by 
scaled time-varying mixing matrices $M_{t_k}$, with $M_{t_k}^{a,b}$ giving relative rates of effective contacts 
between individuals of each pair of age groups $(a, b)$ at time $t_k$. Quantity $1 - M_{t_k}^{a,b}R_0(\psi) / d_I$ 
is the probability of an individual in strata $a$ not being infected by an infectious individual in 
strata $b$ in the interval $[t_k, t_{k+1})$. When raised to the power of $I_{1t_k,b} + I_{2t_k,b}$, this results in 
the probability of not being infected by any individual in strata $b$. Taking the product over 
all strata gives the probability of not being infected at all. This is an adaptation of the Reed-
Frost epidemic model (e.g. Ball, 1983). The initial conditions of the system are determined 
by: parameter $I_0$, the total number of infectious individuals across all age groups at time $t_0$; 
an assumed equilibrium distribution of infections over the age groups; and an assumption of 
initial exponential growth that determines the relationship between the numbers in the four
disease states. For ease of implementation, a reparameterisation is made from $I_0$ to a parameter denoted $\nu$, the details of which can be found in the Supplementary Information to Birrell et al. (2011). Fixing $d_L$ at two days, denote by $\xi = (\psi, \nu, d_L, m)$ the vector of transmission dynamics parameters, where $m$ parameterise the mixing matrices $M_{tk}$.

The transmission process described by (1) is unobserved. However, there are a number of surveillance sources informing aspects of this process, linked to the transmission system through a number of observational models. See Figure A1 in the Web Appendix for a model schematic.

The number of new age-specific infections in interval $[t_{k-1}, t_k)$, expressed as

$$\Delta_{tk,a} \equiv \Delta_{tk,a}(\xi) = S_{tk-1,a} \lambda_{tk-1,a} \delta t,$$

are indirectly related to surveillance data on health-care burden. Each new infection will develop symptoms of influenza-like illness (ILI) with probability $\phi$. With a further probability, $p_{tk,a}^{conf}$, the symptomatic cases will be virologically confirmed through admission to hospital and/or to an intensive care unit (ICU). Alternatively, with probability $p_{tk,a}^{doc}$, they will choose to contact a primary care practitioner and will be reported as a consultations for ILI alongside individuals attending for non-pandemic ILI. As a result, primary care consultation data are contaminated by a “background” component strongly influenced by the public’s volatile sensitivity to government advice. To identify the consultations attributable to the pandemic strain, complementary data from a sub-sample of swabbed ILI patients provide information on the proportion of consultations with pandemic virus.

Let $e \in \{conf, doc\}$ denote counts of confirmed cases or primary care consultations. The expected number of surveillance counts in the interval $[t_{k-1}, t_k)$ attributable to the pandemic virus,

$$p_{tk,a}^e = \phi p_{tk,a}^{conf} \sum_{l=0}^{k-1} \Delta_{tk-1,a} f(\xi, \sigma_l^2)(l),$$

results from the process of becoming infected and subsequently experiencing a delay (with discretised probability mass function $f(\xi, \sigma_l^2)(\cdot)$) that comprises the time from infection to symptoms (the incubation period), the time from symptoms to the healthcare event, and the time from diagnosis to the report of the healthcare event of interest.

### 2.1 Observational model

The number of susceptibles in age group $a$ at the end of the $k$th time-step, $S_{tk,a} \equiv S_{tk,a}(\xi)$, is informed directly by a series of cross-sectional serological survey data $Z_{tk,a}$ on the presence of immunity-conferring antibodies in the general population. Denote $N_a$ the population size in age group $a$ and $v_{tk,a}^{sero}$ the number of blood sera samples tested in time interval $[t_{k-1}, t_k)$, it is assumed that

$$Z_{tk,a} \sim \text{Bin} \left( v_{tk,a}^{sero}, 1 - \frac{S_{tk,a}}{N_a} \right).$$

Counts of daily numbers of confirmed cases or ILI consultations, $X_{tk,a}^{conf}$, are assumed to have negative binomial distribution expressed here in mean-dispersion parameterisation, such that if $X \sim \text{NegBin}(\mu, \eta)$, then $E(X) = \mu$, var $(X) = \mu(\eta + 1)$. For the confirmed cases $X_{tk,a}^{conf}$,

$$X_{tk,a}^{conf} \sim \text{NegBin} \left( \mu_{tk,a}^{conf}, \eta_{tk,a} \right)$$

whereas the primary care consultations $X_{tk,a}^{doc}$ include contamination by a non-pandemic ILI background component $B_{tk,a}$:

$$X_{tk,a}^{doc} \sim \text{NegBin} \left( \mu_{tk,a}^{doc} + B_{tk,a}, \eta_{tk,a} \right).$$
Here the contamination $B_{t_k,a}$ is appropriately parameterised in terms of parameters $\beta^B$ (see Web Appendix B). The signal $\mu_{t_k,a}^{doc}$ is identified by virological data from sub-samples of size $v_{t_k,a}^{viro}$ of the primary care consultations. The number of swabs testing positive for the presence of the pandemic strain $W_{t_k,a}$ in each sample is assumed to be distributed:

$$W_{t_k,a} \sim \text{Bin} \left( v_{t_k,a}^{viro}, 1 - \frac{B_{t_k,a}}{\mu_{t_k,a}^{doc} + B_{t_k,a}} \right).$$

### 2.2 Inference

Let $\theta = \{ \xi, \phi, \mu_{t_k,a}^{doc}, \eta_{t_k}, \beta^B \}$ denote the vector of all free parameters. Birrell et al. (2011) develop a Bayesian approach and use a Markov Chain-Monte Carlo (MCMC) algorithm to derive the posterior distribution of $\theta$ on the basis of 245 days of primary care consultation and swab positivity data, confirmed case and cross-sectional serological data.

The MCMC algorithm is a naively adaptive random walk Metropolis algorithm, requiring $7 \times 10^5$ iterations, taking over four hours. MCMC is not easily adapted for parallelised computation, although a small speed up can be achieved by parallelising the likelihood component of the posterior distribution of $\theta$ over a small number of CPUs. In total, this required in excess of $6.3 \times 10^6$ evaluations of the transmission model and/or convolutions of the kind in equation (4). Implementation of MCMC in an online fashion, as new data arrive involves the re-analysis of the entire dataset, requiring time for multiple Markov chains to converge.

Although, the runtime might not be prohibitive for real-time inference, the current implementation leaves little margin to consider multiple code runs or alternative model formulations. In a future pandemic there will be a greater wealth of data facilitating a greater degree of stratification of the population (Scientific Pandemic Influenza Advisory Committee: Subgroup on Modelling, 2011). With increasing model complexity comes rapidly increasing MCMC runtimes, which can be efficiently addressed through use of SMC methods.

### 3 An SMC Alternative to MCMC

The model in Section 2 is deterministic and designed for use at a time in a pandemic when stochastic effects are uninfluential. In this case epidemic data are imperfect observations distributed around model outputs and the inferential focus is on $\theta$. Letting $Y_t$ denote the vector of all random quantities in (5)-(8), and let $y_t$ be the observed values of $Y_t$, online inference involves the sequential estimation of posterior distributions $\pi_k(\theta) = p(\theta|y_{1:k}) \propto \pi_0(\theta)p(y_{1:k} | \theta)$, $k = 1, \ldots, K$, where $\pi_0(\theta)$ indicates the prior for $\theta$. Estimation of any epidemic feature, e.g. the assessment of the current state of the epidemic or prediction of its future course, follows from estimating $\theta$.

Suppose at time $t_k$ a set of $n_k$ particles $\{\theta_k^{(1)}, \ldots, \theta_k^{(n_k)}\}$, with associated weights $\{\omega_k^{(1)}, \ldots, \omega_k^{(n_k)}\}$, approximate a sample from the target distribution $\pi_k(\cdot)$. On the arrival of the next batch of data $y_{k+1}$, $\pi_k(\cdot)$ is used as an importance sampling distribution to sample from $\pi_{k+1}(\cdot)$. In practice, this involves a reweighting of the particle set. The particles are reweighted according to the importance ratio, $\tilde{\pi}_{k+1}(\cdot)/\pi_k(\cdot)$, which reduces to the likelihood of the incoming data batch, i.e:

$$\omega_{k+1}^{(j)} \propto \omega_k^{(j)} \frac{\pi_{k+1}(\theta_k^{(j)})}{\pi_k(\theta_k^{(j)})} = \omega_k^{(j)} p(Y_{k+1}|\theta_k^{(j)}).$$
Eventually, many particles will carry relatively low weight, leading to sample degeneracy as progressively fewer particles contribute meaningfully to the estimation of $\pi_k(\cdot)$. A measure of this degeneracy is the effective sample size (ESS) (Liu and Chen, 1995),

$$\text{ESS} \left( \left\{ \omega_k^{(i)} \right\} \right) = \frac{\left( \sum_{j=1}^{n_k} \omega_k^{(j)} \right)^2}{\sum_{j=1}^{n_k} \omega_k^{(j)}},$$

with values of the ESS that are small in comparison to $n_k$ being indicative of the impoverishment of the current particle set.

This degeneracy can be tackled in different ways. Gordon et al. (1993) introduced a resampling step, removing low weight particles and jittering the remainder. This jittering step was formalised by Gilks and Berzuini (2001) using Metropolis-Hastings (MH) steps to rejuvenate the sample. Fearnhead (2002) and Chopin (2002) provide more general treatises of this SMC method, with Chopin (2002) labelling the algorithm ‘iterated batch importance sampling’. This was extended by Del Moral et al. (2006) who unify the static estimation of $\theta$ with the filtering problem (estimation of $x_k$).

Here we adapt the resample-move algorithm of Gilks and Berzuini (2001), investigating its real-time efficiency in comparison to successive use of MCMC. The MH steps rejuvenating the sample constitute the computational bottle-neck in resample-move as they require a browsing of the whole data history to evaluate the full likelihood, not just the most recent batch. For fast inference, the number of such steps should be minimised, without risking Monte Carlo error through sample degeneracy. Their algorithm is laid out in full below. It is presumed that it is straightforward to sample from the prior distribution $\pi_0(\theta)$.

### 3.1 The Algorithm

1. **Set** $k = 0$. Draw a sample $\{\theta_0^{(1)}, \ldots, \theta_0^{(n_0)}\}$ from the prior distribution, $\pi_0(\theta)$, set the weights $\omega_0^{(j)} = 1/n_0$, $\forall j$.

2. **Set** $k = k + 1$. Observe a new batch of data $Y_k = y_k$. Reweight the particles so that the $j^{\text{th}}$ particle has weight $\tilde{\omega}_k^{(j)} \propto \omega_k^{(j)} - 1 p(y_k | \theta_k^{(j)} - 1)$. (10)

3. **Calculate the effective sample size.** Set $\omega_k^{*(j)} = \tilde{\omega}_k^{(j)} / \sum_j \tilde{\omega}_k^{(j)}, \forall j$. If $ESS \left( \left\{ \omega_k^{*(j)} \right\} \right) > \epsilon_L n_{k-1}$ set $\theta_k^{(j)} = \theta_{k-1}^{(j)}$, $\omega_k^{(j)} = \omega_k^{*(j)}$, $n_k = n_{k-1}$ and return to point (2), else go next.

4. **Resample.** Choose $n_k$ and sample $\{\tilde{\theta}_k^{(j)}\}_{j=1}^{n_k}$ from the set of particles $\{\theta_{k-1}^{(j)}\}_{j=1}^{n_{k-1}}$ with corresponding probabilities $\{\omega_k^{*(j)}\}_{j=1}^{n_{k-1}}$. Here, we have used residual resampling (Liu and Chen, 1998). Re-set $\omega_k^{(j)} = 1/n_k$.

5. **Move:** For each $j$, move from $\tilde{\theta}_k^{(j)}$ to $\theta_k^{(j)}$ via a MH kernel $K_k(\tilde{\theta}_k^{(j)}, \theta_k^{(j)}; \gamma)$. If $k < K$, return to point (2).

6. **End.**

There are a number of algorithmic choices to be made, including tuning the parameters of the MH kernel ($\gamma$ above) or the rejuvenation threshold, $\epsilon_L$. In a real-time setting, it may not be
possible to tune an algorithm “on the fly”, so the system has to work “out of the box”, either through prior tuning or through being adaptive. In what follows we set $\epsilon_L = 0.5$ (Jasra et al., 2011) and we focus on the key factor affecting the performance of the algorithm in real-time, i.e. the MH kernel.

3.1.1 Kernel Choice

**Correlated Random Walk** A correlated random walk proposes values in the neighbourhood of the current particle:

$$\theta^*|\tilde{\theta}_k^{(j)} \sim N (\tilde{\theta}_k^{(j)}, \gamma \tilde{\Sigma}_k), \tag{11}$$

where $\tilde{\Sigma}_k$ is the sample variance-covariance matrix for the weighted sample $\{\tilde{\omega}_k^{(i)}, \theta_k^{(i)}\}$. The parameter $\gamma$ can be tuned *a priori* to guarantee a reasonable acceptance rate, or, alternatively, asymptotic results for the optimal scaling of covariance matrices (Roberts and Rosenthal, 2001; Sherlock et al., 2010) can be used. Localised moves keep acceptance rates high and will quickly restore the value of the ESS. However, if after re-sampling there are few unique particles then the rejuvenation can result in a highly clustered sample, providing an inaccurate representation of the target distribution.

**Approximate Gibbs’** An independence sampler that proposes (Chopin, 2002):

$$\theta^*|\tilde{\theta}_k^{(j)} \sim N (\bar{\theta}_k, \Sigma_k) \tag{12}$$

where $\bar{\theta}_k$ is the sample mean for the $\{\tilde{\omega}_k^{(i)}, \theta_k^{(i)}\}$. Here, proposals are drawn from a distribution chosen to approximate the target distribution, only weakly-dependent on the current position of the particle. An accept-reject step is still required to correct for this approximation. The quality of the approximation depends on $\pi_{k-1}(\cdot)$ being well represented by the current particle set, there being sufficient richness in the particle weights after the reweighting step and the target density being sufficiently near-Gaussian. Assuming that the multivariate normal approximation to the target is adequate (and it should be increasingly so as more data are acquired) this type of proposal allows for more rapid exploration of the sample space.

Both the correlated random walk and the approximate Gibbs’ methods will be used, both as block updates where a new value for the entire parameter vector is proposed at once, and in component-wise updates where individual or sub-groups of parameter components are proposed in turn, using the appropriate conditional distributions derived from (11) and (12).

4 Epidemic prediction

Until now we have focused on the development of an algorithm for the sequential estimation of the posterior distribution of the model parameters However, it will also be necessary to examine, in real-time, the ability of the model to make epidemic forecasts. These forecasts will be of two different types: long-range predictions and one-step ahead predictions. Long-range forecasts are based on posterior predictive distributions of the form $\pi(y_{(k):(k+l)}|y_{1:(k-1)})$, and are essential to anticipate the epidemic healthcare burden. On the other hand, probabilistic one-step ahead forecasts allow for timely, online assessment of model adequacy through their comparison with the incoming data. The quality of these predictions can be visually assessed through the plotting of probability integral transform (PIT) histograms (Dawid, 1984), as defined for count data by Czado et al. (2009). Denoting the one-step ahead posterior predictive distribution
\( \pi(y_k | y_{1:(k-1)}) \) to have corresponding cumulative distribution function \( F_k \equiv F_k(y_k) \), a histogram of the empirical \( F_k \), \( k = 1, \ldots, K \) should be close to uniform if the predictions are adequate. A more formal method is based on proper scoring rules (Gneiting and Raftery, 2007), which assign a numerical score to each one-step predictive distribution evaluated at the subsequent data point. The suitability of various scoring rules are discussed in Czado et al. (2009) while Seillier-Moiseiwitsch and Dawid (1993) and Held et al. (2017) show how these scoring rules can be aggregated to test a hypothesis of prediction adequacy. In particular Held et al. (2017) calculate the value

\[
 s_{\text{rps}}(F, y) = \sum_{k=0}^{\infty} \left\{ F_k - \mathbb{1}(y \leq k) \right\}^2,
\]

the ranked probability score for each one-step ahead prediction. The mean \( \overline{s}_{\text{rps}} \) of these scores taken over all such predictions is used to form the test statistic

\[
 z_{\text{rps}} = \frac{\overline{s}_{\text{rps}} - \mathbb{E}(\overline{s}_{\text{rps}})}{\text{Var}(\overline{s}_{\text{rps}})},
\]

which has a standard normal distribution under a null hypothesis of prediction adequacy. Overall, if these diagnostics expose a persistent lack of predictive ability, then the epidemic model needs adaptation if it is to be consistent with incoming batches of data.

5 A simulated epidemic

The SMC algorithm’s performance is contrasted against the gold-standard MCMC through its application to simulated epidemic data generated to mimic the timing and dynamics of the 2009 A/H1N1 pandemic in England. Anomally, this epidemic started with an initial burst of infection in Spring, so a 1st May start date is chosen. The epidemic occurs in two waves of infection with a first peak induced by an over-summer school holiday and a second peak towards the end of the year.

We consider two scenarios: in the first scenario direct information on confirmed cases (e.g. hospitalisation, ICU admissions) is available; in the second scenario we observe ILI consultations in primary care that are noisy and contaminated by non-pandemic infections. Alongside confirmed case and consultation data, serological data (see Section 2.1) are available and, in the second scenario, we also assume the existence of companion virological data taken from a sub-sample of the ILI consultations (see Equation (8)). In both scenarios observations are made on 245 consecutive days on a population divided into \( A = 7 \) age groups, and the underlying epidemic curve is characterised by the same parameters, so both confirmed case and primary care consultation data are subject to similar trends and shocks. One such shock arises from an assumed sudden change in the way the confirmed cases and GP consultations are observed. This could occur due to some public health intervention designed to alleviate the burden placed on primary care services as happened in 2009 with the launch of the National Pandemic Flu Service. Table 1 presents the model parameters together with the values used for simulation.

5.1 Confirmed Case Data

For a given set of parameters \( \theta \), the number of confirmed cases, \( \mu_{t_k,a}^{\text{conf}} \), in interval \([t_{k-1}, t_k)\) is given by Equation (4) and count data \( X_{t_k,a}^{\text{conf}} \) are generated from the negative binomial distribution in (6). Note that the overdispersion parameter \( \eta_{t_k} \) is piecewise constant over time, with a breakpoint at the time of the system shock due to the intervention, chosen to be at \( t_k = 83 \).
Table 1: Parameters used in the simulation of (confirmed case) epidemic data

| Parameter | Description | Value |
|-----------|-------------|-------|
| $\eta$   | Dispersion parameters, split either side of a public health intervention at $t_k = 83$, denoted $(\eta_1, \eta_2)$. | $(3.00, 2.15)$ |
| $d_i$    | The mean infectious period. 1 parameter | 3.47 |
| $\phi$   | The proportion of symptomatic infections. 1 parameter | 0.278 |
| $m$      | Multipliers applied to the contact matrices (e.g. to describe the school-holiday effects). 5 parameters | $(0.403, 0.495, 0.0588, 0.301, 0.421)$ |
| $\psi$   | Exponential growth rate. 1 parameter | 0.133 |
| $\nu$    | A reparameterisation of the initial number of infecteds, a function of $I_0$. 1 parameter | -13.9 |
| $p^{conf}$ or $p^{doc}$ | Parameters governing the population propensity of individuals with ILI symptoms to appear in the data. Split either side of the public health intervention at $t_k = 83$, with different rates for adults and children. 4 parameters | 

\[
p = \begin{cases} 
  p_1 & t_k \leq 83, a < 4 \\
  p_2 & t_k \leq 83, a \geq 4 \\
  p_3p_1 & t_k > 83, a < 4 \\
  p_4p_2 & t_k > 83, a \geq 4 
\end{cases} 
\]

\[
p = (0.278, 0.162, 0.137, 0.441)
\]
| $\beta^B$ | Parameters of the piecewise log-linear model for the background ILI consultation rates, with change-points at $t_k = 83, 129, 177$. | See Web Appendix B. |

5.2 Primary Care Consultation Data

The number of consultations due to the pandemic strain $\mu^{doc}_{t_k,a}(\theta)$ is also calculated via the convolution equation (4). The contamination component is added by assuming ‘background’ consultation rates, $B_{t_k,a}$, that develop over time according to a log-linear spline with a discontinuity at $t_k = 83$. With additional age effects to generate separate consultation rates for children (< 16 year-olds) and adults, this spline model is characterised by a 9-dimensional parameter $\beta^B$. The precise specification of this spline is detailed in Appendix B and the log-linear spline used in the simulation is plotted, aggregated over age groups, in Figure 1(D).

In both simulated datasets, the number of consultations will drop markedly due to the intervention, introduced through a sudden change in the parameter $p^{doc}$, the proportion of symptomatic cases that seek consultation. In reality, $p^{doc}_{t_k,a}$ may exhibit greater temporal variation than $p^{conf}_{t_k,a}$ as it depends on behavioural factors and is not a property of the virus. However, in the examples presented here they are parameterised similarly (see Table 1).

The daily consultations are generated from the negative binomial distribution in (7), assuming the same degree of overdispersion used in the generation of the confirmed cases. The companion virological dataset consists of samples of the same size and timing as those available from 2009.

5.3 Serological Data

Serological data arise from the testing of blood sera samples assumed to be taken representatively to give a time-evolving picture of the presence of immunity in the population. Again, the sample timings and sizes are the same as those taken in the 2009 pandemic, with the positive
counts \{Z_{t_k,a}\} \text{ simulated from Equation (5).}

All the data for the second scenario are presented in Figure 1.

6 Results from a resample-move SMC algorithm

In this section we recreate the process of tracking the evolution of an epidemic, comparing the performance of a number of SMC schemes to the 'gold-standard' MCMC algorithm in Section 2.2.

The focus of a public health response in the early stages of an emerging epidemic will be on the estimation of some key epidemic parameters from a few initial cases in localised outbreaks. Real time monitoring of the epidemic will begin after this initial stage, taken here to be the first 50 days of the epidemic. An MCMC implementation of the model is carried out at times \( t_k = 50, 70, 83, 120, 164 \) and 245 days and the SMC algorithm will then be used to propagate the MCMC-obtained posteriors over the intervals defined by these timepoints. For example, the MCMC-obtained estimate \( \pi_{50}^{\text{MCMC}}(\theta) \) of \( \pi_{50}(\theta) \) will be used as the starting point for the SMC algorithm over the interval 50-70 days. The SMC algorithm will then give an estimate \( \pi_{70|50}^{\text{SMC}}(\theta) \) for \( \pi_{70}(\theta) \), which will then be compared with \( \pi_{70}^{\text{MCMC}}(\theta) \). The similarity (or divergence) between the two distributions is measured by an approximation to the Kullback-Leibler (KL) divergence of \( \pi_{70|50}^{\text{SMC}}(\theta) \) from \( \pi_{70}^{\text{MCMC}}(\theta) \), obtained by assuming that both distributions are multivariate normal distributions, approximating \( \pi_{70}^{\text{MCMC}}(\theta) \) and \( \pi_{70|50}^{\text{SMC}}(\theta) \) by \( \mathcal{N}(\mu_0, \Sigma_0) \).
Table 2: Scenario 1: Kullback-Leibler statistics and likelihood evaluations per day (‘Run Time’) for each resample-move algorithm.

| Proposal Method | Correlated Component-wise Block approx. Gibbs | Component-wise Block approx. Gibbs |
|----------------|---------------------------------------------|----------------------------------|
| Intervals      | Random-Walk | Gibbs approx. | Gibbs approx. |
| 0-50 KL        | 2.83        | 2.58          | 2.61          |
| Run Time       | 18200       | 16800         | 8000          |
| 51-70 KL       | 2.00        | 0.908         | 1.32          |
| Run Time       | 21000       | 21000         | 8000          |
| 71-83 KL       | 4.44        | 1.06          | 1.60          |
| Run Time       | 26923       | 26923         | 7692          |
| 84-120 KL      | 16.3        | 6.58          | 2.09          |
| Run Time       | 20811       | 17027         | 10000         |
| 121-164 KL     | 0.106       | 0.113         | 0.122         |
| Run Time       | 3182        | 3182          | 4773          |
| 165-245 KL     | 0.339       | 0.471         | 1.15          |
| Run Time       | 8642        | 9506          | 9136          |

and \( N(\mu_1, \Sigma_1) \) respectively:

\[
KL(\pi_{t_k}^{MCMC} \parallel \pi_{t_k}^{SMC}) = \int_\Theta \pi_{t_k}^{MCMC}(\theta) \log \left( \frac{\pi_{t_k}^{MCMC}(\theta)}{\pi_{t_k}^{SMC}(\theta)} \right) d\theta \\
\approx \frac{1}{2} \left\{ \text{tr} \left( \Sigma_1^{-1} \Sigma_0 \right) + (\mu_1 - \mu_0)^T \Sigma_1^{-1} (\mu_1 - \mu_0) - \dim(\theta) + \log \left( \frac{|\Sigma_1|}{|\Sigma_0|} \right) \right\}. \tag{15}
\]

6.1 Scenario 1: Using Confirmed cases and Serology Data

Table 2 reports the KL divergences between the target posterior distributions obtained from SMC (with rejuvenations using only a single MH iteration) and MCMC at \( t_k = 50, 70, 83, 120, 164 \) and 245 days. The correlated random-walk (11) is notably inferior over most of the intervals prior to the interval 84–120 days. Beyond this time, as data accumulate, the divergence between distributions \( \pi_k \) and \( \pi_{k+1} \) is small and the conservative random-walk proposals become progressively more adequate at bridging the gap. The component-wise approximate Gibbs scheme (12) performs better, minimising KL divergence over almost all intervals. As the move step here has many accept-reject steps (one for each component grouping of Table 1), each of low dimension, the overall acceptance rate (the proportion of particles for which at least one component moves) is very high, giving \( ESS \approx n_k \).

Figure 2 illustrates the performance of the approximate Gibbs component-wise proposal kernel comparing the SMC- and MCMC-obtained scatterplots for the parameter components \( \psi \) and \( \nu \) at \( t_k = 70 \) (A), \( t_k = 120 \) (B) and \( t_k = 245 \) (C). There is close correspondence between
the SMC and MCMC obtained distributions at $t_k = 70$ and $t_k = 245$, but substantial departure at $t_k = 120$.

6.2 Scenario 2: Using Primary Care Consultations and Serology Data

With a parameter space that has expanded from 15 to 24 dimensions, the problems identified in Table 2 are magnified (see Table C1 of the Web Appendix). In particular, after $t_k = 83$ the number of new parameters that become active is greater than in Scenario 1 and for the following few days some of these new parameters are not identifiable. Possibly as a consequence, the KL divergences over the interval 83-120, irrespective of the proposal scheme, are arbitrarily high.

6.3 Remarks

It is clear that the basic resample-move SMC algorithm cannot handle the ‘shock’ in the count data occurring at $t_k = 83$. In scenario 1, this shock is accommodated by the model through step changes in the parameters $\eta_{t_k}$ and $p^{\text{e}}_{t_k,a}$, with similar step changes in the levels of background consultation in Scenario 2. For these parameter components, the target marginal posterior distributions move rapidly from day 84 as probability density shifts away from uninformative prior distributions. As an example, Figure 3 shows the change in the marginal posterior for the overdispersion parameter $\eta_{t_k}$ over the interval [84, 90] days. The chosen $\Gamma(0.01, 0.01)$ prior distribution is unbounded and highly non-Gaussian even after transforming to the log-scale. In this case, there is very little prior to posterior overlap and normal proposal distributions represent a poor choice.

For Scenario 1, the 84-120 day interval is the only one over which the block-update approximate Gibbs method gives the best performance (see KL divergence in Table 2). This is attributable to the way the proposal mechanisms modify the ESS as discussed in Section 3.1.1. The inability of the full-block Gibbs updates to restore the ESS, unlike the componentwise algorithms, leads to a rejuvenation step at the arrival of each new data point. This greater number of rejuvenations better enables the tracking of the shifting posterior distributions over time, although it negates any advantages of this algorithm in terms of computation time (see Tables 2 and A1). However, even with the block updates, good correspondence between the SMC- and MCMC-obtained posteriors is not achieved in Scenario 1 until $t_k \approx 100$, and not at all in Scenario 2.

From these initial results it is clear that a modified algorithmic formulation is needed for computationally efficient inference when target posteriors are highly non-Gaussian and/or are moving fast between successive batches of data as a consequence of highly informative observations.

7 Extending the algorithm - handling informative observations

A key feature of any improved SMC algorithm must be that the ESS (9) retains its interpretation as the “required size of an independent sample drawn directly from the target distribution to achieve the same estimating precision attained by the sample contained in the particle set” (Carpenter et al., 1999). A single step of the component-wise algorithms of Section 6 restored the ESS, i.e. $\text{ESS} \approx n_k$. In general, as the proposal scaling tends to zero, i.e. $\gamma \downarrow 0$, acceptance rates will be close to one, resulting in a set of mostly unique particles and a high value for the
Figure 2: Comparison of SMC-obtained posteriors and MCMC-obtained posteriors at \( t_k = 70 \) (A), \( t_k = 120 \) (B) and \( t_k = 245 \) (C) days, via scatter plots for the parameters \( \psi \) and \( \nu \). The grey points in both the left and the right panels represent the MCMC-obtained sample at the beginning of the interval, with the overlaid coloured points representing the SMC or MCMC-obtained samples at the end of the interval. In the SMC-obtained samples, the colour of the plotted points represents the weight attached to the particle, with the red particles being those of heaviest weight.
ESS. However, this would be a highly clustered posterior sample, barely distinguishable from the set of resampled particles. Such a sample is not as informative as an independent sample of size \( n_k \), and so the ESS, as calculated from the particle weights, is no longer a reliable guide to the quality of the sample. Conversely, the ESS (see Section 6.1) ceases to be an adequate guide for identifying rejuvenation times when considering block-update approximate-Gibbs proposals, as it is unable to recover to the threshold \( \epsilon_L n_k \).

We look at three possible improvements to the resample-move algorithm of Section 3, to produce an information-adjusted SMC algorithm that ensures that the ESS remains a good measure of the quality of the sample: we address the timing of rejuvenations; we reconsider the choice of kernels used in the rejuvenations; and we question the number of iterations we need to run the MCMC sampler before the sample is fully rejuvenated.

### 7.1 Timing the rejuvenations: a continuous-time formulation

Here we discuss the idea of rejuvenating at intervening times, including only a fraction of the new batch of data, so that if there is large divergence between consecutive target distributions \( \pi_k \) and \( \pi_{k+1} \), the estimation of intermediate distributions will allow the particle set to move gradually between the two targets (Del Moral et al., 2006). These intermediate distributions are generated via tempering (Neal, 1996), by gradually introducing the new batch of data into the likelihood at a range of ‘temperatures’ \( \delta \in [0, 1] \). These distributions are

\[
\pi_{k, \delta}(\theta) \propto \pi_k(\theta) \left\{ p \left( y_{k+1} | \theta \right) \right\}^\delta.
\]

Assume that batch of data \( y_{k+1} \) arrives uniformly over the \((k + 1)\text{th}\) interval rather than at the end of the interval. The particle weights will develop according to

\[
\tilde{\omega}^{(j)}_{k+\delta} = \omega_{k}^{(j)} \left\{ p \left( y_{k+1} | \theta^{(j)} \right) \right\}^\delta, \quad j = 1, \ldots, n_k.
\]

More generally, denote \( \omega_{k+\delta, \delta_0}^{(j)} \) the weight attached to a particle at an intermediate time \( t_{k+\delta} \) when the previous rejuvenation took place at time \( t_{k+\delta_0} \), with \( \delta_0 = 0 \) corresponding to no prior rejuvenation within the interval \([t_k, t_{k+1}]\). Then, for \( \delta \geq \delta_0 \geq 0 \) and indicator function \( \mathbb{1}_A \),

\[
\tilde{\omega}^{(j)}_{k+\delta, \delta_0} = \left( \omega_{k}^{(j)} + \left( 1 - \omega_{k}^{(j)} \right) \mathbb{1}_{\delta_0 > 0} \right) p \left( y_{k+1} | \theta^{(j)} \right)^{\delta - \delta_0}.
\]
Therefore, if $ESS\left(\{\omega_{k+1,j,0}^{(j)}\}_{j=1}^{n_k}\right) < \epsilon Ln_k$ a further rejuvenation would be proposed at time $\delta^*$, such that $\delta^* = \arg\min_{\delta \in (0,1)} \{ESS(\omega_{k+\delta,0}^{(j)}) - \epsilon Ln_k\}^2$. The number of rejuvenations required within the $(k+1)^{th}$ interval will be in proportion to the degree of particle impoverishment that would have occurred were no rejuvenation to take place.

### 7.2 Choosing kernels - hybrid algorithms.

As discussed in Section 6.3, each of the possible MH kernels has its own distinct strengths which can be exploited by using a combination of kernels. Full block approximate-Gibbs updates are efficient at reducing the clustering that forms around resampled particles. Adding a random walk step would allow the proposal of values outside the space spanned by the principal components of $\Sigma_k$, something of particular necessity if the ESS becomes very small and $\Sigma_k$ is close to singularity.

This motivates a hybridisation of the proposal mechanism, done either by using mixture proposals, e.g. a mixture between the approximate Gibbs’ proposals and full block ordinary random walk Metropolis proposals (Kantas et al., 2014), or, as will be used in the remainder, by augmenting full block approximate Gibbs updates with componentwise random walk proposals.

### 7.3 How many MH iterations? Multiple proposals and intra-class correlation

In the MH-step of the algorithm, there are effectively $n_k$ parallel MCMC chains. Making proposals until all chains have attained convergence would be an inefficiency. In theory, the distribution governing the starting states of these MCMC chains forms a biased sample from the stationary distribution of the MCMC chain. It then seems a reasonable and sufficient requirement that we carry out MH steps until the chains have, to some degree, collectively ‘forgotten’ their starting positions. This can be monitored through an estimate of an intra-class correlation coefficient, $\rho$. To get such an estimate, we divide the particle set into $I$ clusters, each of size $d_i$, $i = 1, \ldots, I$, defined by the parent particle at the resampling stage. For example, if a particular particle is resampled 5 times, it defines a cluster in the new sample with $d_i = 5$. For a univariate summary, $g_{ij} = g(\theta_{ij})$, of the epidemic curve described by the $j^{th}$ particle in the $i^{th}$ cluster, $\theta_{ij}$, $\rho$ is estimated by the analysis of variance intra-class correlation coefficient (Donner and Koval, 1980; Sokal and Rohlf, 1981), given by

$$r_A = \frac{(MS_a - MS_w)/d_0}{(MS_a - MS_w)/d_0 + MS_w},$$

where $MS_a = \frac{1}{I-1} \sum_{i=1}^{I} d_i (\bar{g}_i - \bar{g})^2$, $MS_w = \frac{1}{d(I-1)} \sum_{i=1}^{I} \sum_{j=1}^{d_i} (g_{ij} - \bar{g}_i)^2$ and $d_0 = \bar{d} - \frac{1}{d(I-1)} \sum_{i=1}^{I} (d_i - \bar{d})^2$, represent a between-class mean sum-of-squares, a within-class mean sum-of-squares and an average class size respectively, with $d = \sum_{i=1}^{I} d_i$, $\bar{g}_i = \frac{1}{d_i} \sum_{j=1}^{d_i} g_{ij}$ and $\bar{g} = \frac{1}{d} \sum_{i=1}^{I} \sum_{j=1}^{d_i} g_{ij}$. Clusters with $d_i \leq 1$ are omitted as they have no within cluster variation (Donner and Koval, 1980). Prior to the MH-phase of the algorithm, $r_A$ will be equal to 1, as there is no within-class variation. However, with each iteration of the chosen MH-sampler, $\rho$ will decrease and, in general, so will its estimate $r_A$. We aim to choose a sufficiently small positive threshold for $r_A$ to be the point beyond which there is no longer any value in carrying out further MH proposals to rejuvenate the sample, as particles spawned from different progenitors become indistinguishable from each other. Ideally we will choose this threshold to
Table 3: Performance of the information-adjusted SMC algorithm over the interval 83-120 days by ICC threshold. Traditional filter using daily data (‘discrete’); continuous-time filter (‘continuous’). Cells shaded in green contain KL divergence that is lower than the reference KL divergence distribution.

| ICC threshold  | 0.5 | 0.2 | 0.1 | ICC threshold  | 0.5 | 0.2 | 0.1 |
|----------------|-----|-----|-----|----------------|-----|-----|-----|
| 84 Days (KL target = 0.732) |     |     |     | 90 Days (KL target = 0.159) |     |     |     |
| Continuous    | 1.95| 3.46| 3.48| Continuous    | 0.805| 0.0358| 0.113 |
| Discrete      | 1.22| 1.31| 1.51| Discrete      | 1.22| 1.05 | 0.970 |
| 85 Days (KL target = 0.135) |     |     |     | 100 Days (KL target = 0.135) |     |     |     |
| Continuous    | 0.862| 2.03| 1.68| Continuous    | 0.691| 0.120| 0.0501 |
| Discrete      | 1.50| 1.18| 1.02| Discrete      | 1.15| 0.942| 0.832 |
| 86 Days (KL target = 0.365) |     |     |     | 110 Days (KL target = 0.122) |     |     |     |
| Continuous    | 0.780| 2.01| 2.02| Continuous    | 0.776| 0.167| 0.0799 |
| Discrete      | 1.78| 1.37| 1.24| Discrete      | 1.01| 0.719| 0.630 |
| 87 Days (KL target = 0.276) |     |     |     | 120 Days (KL target = 0.119) |     |     |     |
| Continuous    | 0.282| 0.358| 0.0427| Continuous    | 0.666| 0.278| 0.0842 |
| Discrete      | 1.26| 0.887| 0.696| Discrete      | 0.888| 0.498| 0.552 |

be as large as is practicably possible to minimise the number of rejuvenations required. We shall test our algorithms with thresholds \( r^*_A = 0.1, 0.2, 0.5 \). The ‘attack rate’ of the epidemic, the cumulative number of infections caused by the epidemic, is an obvious choice of summary as it is a key measure of epidemic burden. Recalling (3), the attack rate is formally defined:

\[
g(\theta) = \sum_{t=1}^{\infty} \sum_{a=1}^{A} \Delta_{t,a}(\theta) \frac{1}{N_a}.
\]  

(16)

8 Results from an information-adjusted SMC algorithm

Here we focus mainly on the intervention-spanning day 83 – 120 interval. In what follows, a hybrid algorithm is adopted, using combinations of three thresholds for \( r_A \) with both the continuous and discrete sequential algorithms.

8.1 Scenario 1: Using confirmed case and serology data

8.1.1 Choosing an algorithm: Kullback-Leibler

Further MCMC samples were obtained using data up to and including days 84, 85, 86, 87, 90, 100, 110 and 120, so that KL divergences could be computed at each of these additional time points. In Figure 4(A), KL discrepancies are plotted over time for each combination of algorithms and thresholds. The SMC approximation is closer to the MCMC-obtained posterior distributions for lower values of \( r^*_A \). This is to be expected as a low value of \( r^*_A \) requires a greater number of iterations of the MCMC chain within each rejuvenation (see Figure 4(C)), leading to SMC-derived posteriors more closely resembling the ‘gold-standard’.

It is difficult to interpret the KL divergences in their own right. To construct a reference distribution of such divergences, the MCMC analyses were repeated a further 40 times at each of the times in Table 3, using different starting points and random seeds. The KL divergences of each of these 40 posterior distributions from the original MCMC analysis were then calculated.
Figure 4: (A) Kullback-Leibler divergence over time; (B) ESS for different thresholds using both discrete and continuous time algorithms; (C) Number of proposals required at each rejuvenation time by algorithm. Black and red lines correspond to the use of discrete and continuous algorithms respectively, solid, dashed and dotted lines correspond to the use of the thresholds $r^*_A = 0.1, 0.2$ and 0.5 respectively.
We look for divergences in Table 3 that are typical of this sample of KL divergences. The cells in the table highlighted in green indicate where the KL divergence lies among the lower 95% of sampled KL values at that time, a threshold marked in Table 3 as the ‘KL target’. For \( t_k = 84, 85, 86 \), there is no apparent best performing algorithm as none manage to yield a value for the KL that is typical of an MCMC analysis. From \( t_k = 87 \) onwards, the continuous-time algorithm is much more efficient as for the given \( r^*_A \), fewer iterations are required to attain a much lower KL value (see Figures 4(A) and 4(C)). These results were tested under multiple reruns of the algorithm. For \( r^*_A = 0.5 \), findings from the continuous-time algorithm were highly volatile. Despite this, Figure 4(B) shows that the number and timing of rejuvenations appears independent of the choice of \( r^*_A \) and the decline in the ESS is independent of the quality of the initial sample. In conclusion, the continuous-time algorithm is to be preferred with a threshold of, at most, 0.2. Note that for \( r^*_A = 0.1 \), the continuous-time algorithm has KL typical of an MCMC analysis for all values of \( t_k \geq 87 \).

### 8.1.2 Acceptance rates

Performance of the continuous-time algorithm appears strongly linked to the acceptance rate of the block approximate Gibbs’ proposals. This acceptance rate is particularly low prior to \( t_k = 87 \), before undergoing a sudden step change and increasing from 1-2% to 15-20%, although, for \( r^*_A = 0.5 \), this step change only occurred in about 50% of runs. In contrast, the acceptance rates for the discrete-time algorithm are consistently around 5% throughout, as seen from the number of proposals required over time (Figure 4(C)). The result of this is that, from day 87 onwards, far fewer proposals are required in total for the continuous-time algorithm, even if the number of rejuvenation times increases.

### 8.2 Scenario 2: Using primary care consultations and serology data

#### 8.2.1 Choosing an algorithm

As in Section 6.2, results from SMC applied to contaminated count data display many of the phenomena already discussed. Table 4 gives results comparable to those in Table 3 with cells highlighted in green to be interpreted as previously.

Following on from Section 8.1, the discrete-time algorithm has been dropped from consideration and results are presented from an algorithm labelled ‘cts.reduced’ in Table 4. We have seen from Figures 3 that kernel density estimates for the posterior marginal distributions of dispersion parameter \( \eta_2 \) and \( \log(\eta_2) \) (over days 84-90) show that the distribution of \( \eta_2 \) is highly non-Gaussian after \( t_k = 83 \) days, stemming from the unbounded gamma priors placed upon the \( \eta \) parameters. This non-normality leads to very poor acceptance rates for the approximate-Gibbs’ proposals, which, though initially adequate, fall to 0.3% on day 89, illustrated by a peak of over 250 proposals per rejuvenation and over 400 proposals per day in Figures 5(A) and (B) respectively. To improve acceptance rates the ‘cts.reduced’ algorithm was devised. The dispersion parameters are omitted from the block approximate-Gibbs updates and are proposed separately. In terms of the resulting KL divergences, there is no significant drop in performance in moving from the continuous to the ‘cts. reduced’ algorithm as seen by the amount and position of the green cells in Table 4. There is an exception on day 90 where there is an unexpectedly high KL value when the threshold is 0.1. This may arise from the instability that can arise from the estimation of \( \Sigma_0 \) and \( \Sigma_1 \) in Equation (15) on the basis of Monte Carlo samples (Scheuerer and Hamill, 2015). This instability was observed in a small number of the ‘bootstrap’ MCMC samples, but not in significant enough quantity to have any influence over the estimated 95%
Table 4: Performance of the information-adjusted SMC algorithm over the interval 83-120 days, using the continuous filter and the continuous filter alternative with the negative binomial dispersion parameters removed from the block proposals. Here, the parameters describing the background rates of consultation have been removed from the KL calculations.

| ICC threshold | 0.5  | 0.2  | 0.1  | ICC threshold | 0.5  | 0.2  | 0.1  |
|---------------|------|------|------|---------------|------|------|------|
| 84 Days (KL target = 6.06) |     |      |  | 90 Days (KL target = 0.120) |     |      |  |
| Continuous   | 2.92 | 2.87 | 2.83 | Continuous   | 1.80 | 0.353 | 0.0663 |
| Cts. Reduced | 2.97 | 2.85 | 2.86 | Cts. Reduced | 2.10 | 0.0927 | 1.42 |
| 85 Days (KL target = 1.90) |     |      |  | 100 Days (KL target = 0.182) |     |      |  |
| Continuous   | 3.05 | 3.00 | 2.98 | Continuous   | 0.157 | 0.102 | 0.0890 |
| Cts. Reduced | 3.06 | 2.97 | 2.98 | Cts. Reduced | 0.107 | 0.0835 | 0.0701 |
| 86 Days (KL target = 1.94) |     |      |  | 110 Days (KL target = 0.0936) |     |      |  |
| Continuous   | 3.28 | 3.24 | 3.25 | Continuous   | 0.159 | 0.0774 | 0.111 |
| Cts. Reduced | 3.27 | 3.22 | 3.26 | Cts. Reduced | 0.197 | 0.0373 | 0.0348 |
| 87 Days (KL target = 5.44) |     |      |  | 120 Days (KL target = 0.101) |     |      |  |
| Continuous   | 2.54 | 2.45 | 2.42 | Continuous   | 0.136 | 0.0435 | 0.0708 |
| Cts. Reduced | 2.51 | 2.48 | 2.44 | Cts. Reduced | 0.0999 | 0.0423 | 0.0551 |

The ‘cts. reduced’ proposal scheme requires far fewer iterations of the Metropolis-Hastings algorithm over the interval 84-90 days, maintaining acceptance rates of about 10% over this period. Note, that over time, as the target distribution converges to a multivariate normal distribution, the number of moves required for both methods equalise (in fact, the plain continuous-time algorithm is marginally faster) and the benefit of using the ‘cts. reduced’ proposal scheme diminishes (Figure 5).

8.2.2 Parameter estimation

Most of the scatter plots contrasting the posterior distributions obtained under either the ‘continuous’ or ‘cts. reduced’ schemes show a similar level of correspondence to their MCMC-obtained counterparts that observed in Figure S3 for the ‘continuous’ algorithm. However, for the parameters of the background consultation rates (see Figure 1(D)) this is not the case. For the first couple of days post \( t_k = 83 \) days, some of the parameters describing \( B_{t_k,a} \) are only weakly identifiable. Figure 6 highlights this for two weakly identifiable parameter components \( \beta^B_3 \) and \( \beta^B_9 \) showing a clear discrepancy between the MCMC- and the SMC-obtained posterior scatters. The SMC distributions, being based on many short MCMC chains, cover the full posterior distribution adequately. For \( t_k = 85, 86, \) however, the MCMC has difficulty mixing, and this manifests in a particle scatter that is stuck in a sub-region of the full marginal support. KL discrepancies calculated for days where this weak identifiability exists (and it diminishes over time), will therefore be unreliable.

8.2.3 Run-times

Figure 5(C) contrasts the daily run-times for the MCMC and SMC algorithms. In the MCMC implementation, the likelihood calculations can be speeded up by parallelisation across 4-8 threads and was optimally run on a desktop computer with 8 parallel 3.6GHz Intel(R) Core(TM) i7-4790 processors. However, the speed up for SMC is close to linear with increasing parallel-
Figure 5: (A) Number of MH-steps required by the continuous-time SMC algorithms per rejuvenation against the timing of the rejuvenation for both the continuous time algorithms (black and red correspond to with and without $\eta$ in the updates) for values of $r_A^*=0.1$ (solid line), 0.2 (dashed line) and 0.5 (dotted line); (B) Total number of MH-steps required by the continuous-time SMC algorithms per time interval, with $r_A^*=0.1$ and using the continuous-time algorithm (grey bars) and the same algorithm without $\eta$ (magenta bars); (C) The computation time for daily model runs under MCMC (blue line) and SMC (red line).

...and we parallelise the MH steps by sharing the particles (i.e. the parallel MCMC chains) across 255 Intel(R) Xeon(R) CPU E5-262 2.0GHz processors on a high-performance computing cluster. Realistically, in a pandemic, even greater parallelisation would be used. Figure 5 shows that, not only is SMC more efficient on day 84, the day requiring the most MH-updates to rejuvenate the sample, but the run-times then decrease over time, in contrast to the increasing MCMC run-times as more data have to be analysed. On days where the sample does not have to be rejuvenated at all, the run-time is negligible in comparison. Note that the MCMC algorithms are run for $4.5 \times 10^5$ iterations, and as seen in Section 8.2.2, this may well be insufficient with the plain-vanilla MH algorithm used here.

### 8.3 Epidemic forecasting

We choose to predict ILI consultations as a measure of the coming healthcare burden. Using a 20-day forecast horizon, the consultation data seem, in the most part, to be well predicted from the landmark times of 90 days and 164 days (Figure 7(A) and 7(B)) using either the MCMC or SMC. However, there is some discrepancy between the MCMC and the SMC in the calculation of the predictive intervals from day 178 onwards (see Figure 7(B)). This may have arisen due to the MCMC not exploring the full support of the parameter space with the introduction of a changepoint in the piecewise background rate model at day 177. Figure 7(C) shows the PIT histogram for all GP consultation data for all age groups and times from day 84 onwards calculated for the SMC analysis. The PIT histogram is not entirely uniform but shows no consistent under or over-estimation, only a lack of smoothness. If we look at the $z$-statistics testing prediction adequacy (14) there is some evidence for miscalibrated forecasts ($p=0.036$). For the virological and serological data these statistics are -1.27 and 1.19 respectively, showing no clear evidence for miscalibration.
Figure 6: The evolution over time of the marginal joint posterior for two components of the parameter vector $\beta^B$. Comparison between SMC-obtained and MCMC-obtained posterior distributions. Grey points indicate the distribution at the start of the interval.
Figure 7: (A) and (B) Comparison of the observed GP data with posterior predictive distributions obtained using the SMC and MCMC algorithms at day 90 and 164 respectively. Solid lines give posterior medians of the distributions and the dotted lines give 95% credible intervals for the data. Red intervals are obtained via MCMC and blue via SMC. (C) PIT histograms for the one-step ahead predictions of GP ILI consultation data, calculated based on 162 time-points × seven strata.

9 Discussion

This paper addresses the substantive problem of online tracking of an emergent epidemic, assimilating multiple sources of information through the development of an information-adjusted SMC algorithm. When incoming data follow a stable pattern, this process can be automated using standard SMC algorithms, confirming current knowledge (e.g. Dukic et al., 2012; Ong et al., 2010). However, in the likely presence of interventions or any other event that may provide a system shock, it is necessary to adapt the algorithm appropriately. On observing the impact that a new batch of data has on the ESS of a particle set, tailoring of the MH-kernel and selection of suitable thresholds can ensure efficient performance. However, as we have seen, given that prior distributions may not be well chosen and not all models well conceived this might necessitate some careful tinkering.

Having simulated an epidemic where a public health intervention provides a sudden change to the pattern of case reporting, we have constructed a more robust SMC algorithm by tailoring:

(a) the choice of rejuvenation times through tempering;

(b) the choice of the MH-kernel by hybridising local random walk and Gibbs proposals;

(c) a stopping rule for the MH steps based on intra-class correlations to minimise the number of iterations within each rejuvenation.

The end result is an algorithm that is: a hybrid of particle filter and population MCMC (Geyer, 1991; Liang and Wong, 2001; Jasra et al., 2007); robust to possible shocks; improves over the plain-vanilla MCMC in terms of run-times needed to derive accurate inference; and can provide all the distributions we need for posterior predictive measures of model adequacy. Throughout we have inevitably made pragmatic choices and alternative strategies could have been adopted. We reflect on these, lessons learned and outstanding questions in what follows.
9.1 Rejuvenation at times of shocks

In the motivating example, a system “shock” occurred at \( t_k = 83 \) leading to a step-change in the values of many parameters, causing the posterior \( \pi_k(\theta) \) to be no longer a good importance distribution for \( \pi_{k+1}(\theta) \), crucially reducing the usefulness of any proposal kernels based on a reweighted sample from \( \pi_k(\theta) \). This is reflected in a severe drop in the ESS. We have seen how it is essential to rejuvenate particle sets at ‘in-between’ times to ensure the ESS does not collapse totally and ensure that sample rejuvenation is always possible in a timely fashion.

Alternative rejuvenation strategies  

A hybrid MH-kernel is introduced in Section 6: first, long-range, low-acceptance proposals are made, followed by short-range high-acceptance componentwise proposals. This is very conservative approach. In many instances, this hybrid is replaced by a mixture distribution, composed of similar short and long-range moves. The adaptive proposal distributions of Fearnhead and Taylor (2013) might take this a step further, tuning the mixture probabilities so that the moves that have the largest expected jumps are proposed more often. This would be an attractive extension to the present case. However, we would still suggest moving at least a proportion of the particles according to random walk proposals, to guard against a reweighted sample from \( \pi_k(\theta) \) providing a degenerate approximation for \( \pi_{k+1}(\theta) \).

The alternative to running long MCMC chains for each particle when there are new parameters in the model, such as those introduced by the shock, is to expand the particle set by cloning each of the particles a number of times. Each cloned particle has a fresh draw from the prior for each of the new parameter components, to maximise the chance of finding good parameter combinations. Upon observing the next batch of data, the expanded particle set could then be reduced down to a more manageable size. However, this would not solve the problem in Scenario 2 where some parameters are not immediately identifiable and sample impoverishment recurs over a number of days.

Minimising the rejuvenation time  

A low value for the ESS is always indicative of depletion, whereas a high value does not guarantee that the sample is adequate. Section 6 illustrated how the ESS can be artificially rejuvenated even when the particle set is not. For the ESS to be useful, it is essential that previous rejuvenation steps result in a sufficiently independent set of values for the margins of interest. This motivated the use of the analysis of variance intra-class correlation coefficient, \( r_A \), to define a stopping rule for the MH-steps. Currently this rule relies on two algorithmic choices: the choice of a univariate function of interest, \( g(\cdot) \) (see Equation (16)), and the choice of the threshold \( r_A^* \), the largest acceptable value for \( r_A \) at the end of the rejuvenation process.

The function \( g(\cdot) \) should depend on model outputs of particular relevance. The predicted attack rate of an epidemic is a quantity of particular interest to public health policymakers throughout an epidemic and depends on all the transmission parameters. However, when the parameter vector is high-dimensional, as in this case, is it reasonable to condense this into a univariate summary to use as a basis for a stopping rule? Convergence of MCMC is typically diagnosed by looking at marginal distributions, so should we be doing something similar here? Does this necessitate the use of multivariate analogs for the intra-class correlation coefficient (for example, see Ahrens, 1976; Konishi et al., 1991)? It is felt here that the univariate \( g \) is adequate as the parameters introduced at the ‘shock’ time are largely nuisance parameters not strongly correlated with the transmission parameters that influence \( g(\cdot) \).

Once \( r_A \) has been suitably defined, a suitable stopping threshold, \( r_A^* \) has to be chosen.
Given the antecedent prescription for defining clusters used here, then $r_A$ truly is a measure of how well the particles have collectively ‘forgotten’ their starting points. A value of $r_A^* = 0.1$ is a sufficiently small threshold except for extreme cases of departure between two successive distributions.

9.2 Benefits of SMC

Model Run Times  From a computational efficiency point of view, the SMC algorithm, because of its highly parallel nature, is faster than the full MCMC analysis. However, this may be an unfair comparison as the MCMC algorithm is based on “plain vanilla” random-walk Metropolis updates and could benefit from significant tuning itself. More sophisticated MCMC algorithms could be used, as exemplified in an epidemic context by Jewell et al. 2009. The use of differential geometric MCMC (Girolami & Calderhead, 2011) or advances in the parallelisation of MCMC (Banterle et al, 2015), for example, could assist with improving MCMC run times. On the other hand, as MCMC steps are the main computational overhead of the SMC algorithm, any development of the MCMC algorithm may prove similarly advantageous to the SMC algorithm. As the target posteriors attain asymptotic normality it should be progressively easier for SMC to move between distributions over time, as can be seen in Figure 5(C) where the daily running time decreases as data accumulate. For any MCMC algorithm, the opposite will be generally true.

Identifiability  Throughout, we have compared candidate MH-kernels via the KL-like statistics measuring the divergence between SMC posteriors from posteriors generated by the “gold-standard” MCMC. We have also constructed a reference distribution for the KL statistic to assess informally the significance of the observed divergences. This, however, assumes the reliability of the MCMC algorithm used. This superiority is called into question by the apparent better performance of the SMC algorithm in the presence of parameter unidentifiability around changepoints. The background ILI rate is modelled using a piecewise log-linear curve, allowing separate value for adults and children, with changepoints at $t_k = 84, 128, 176$ and 245 days. The value of the curve at these knots is given by

$$
\mu + \alpha_{t_k} + \beta_a, \text{ s.t. } \sum_{t_k \in \{84, 128, 176, 245\}} \alpha_{t_k} = \sum_{a \in \{\text{child, adult}\}} \beta_a = 0,
$$

with linear interpolation giving the value of the curve at the intervening points. This results in background consultation rates for days 84, 85, and 86 that include the respective sums (neglecting the age effects) $\mu + \alpha_{84}$, $\mu + 0.98\alpha_{84} + 0.02\alpha_{128}$, $\mu + 0.96\alpha_{128} + 0.04\alpha_{128}$. This gives only weak identifiability of parameters $\mu$ and $\alpha_{84}$. This parameterisation, as shown in Figure 6, can induce convergence problems for MCMC while Figure 7(C) also exhibits the consequences of this problem at the $t_k = 176$ changepoint. Jasra et al (2011) claim that, for their example, SMC may well be superior to MCMC and this is one case where this is certainly true. The population MCMC carried out in the rejuvenation stage achieves good coverage of the sample space, without the individual chains having to do likewise. To improve the MCMC mixing, a reparameterisation or the use of a geometric sampler may improve sampling efficiency.

Predictive Model Assessment  A fundamental goal of real-time modelling is to derive online epidemic forecasts with an appropriate quantification of the associated uncertainty. Being able to assess the predictive adequacy of a model in real-time is therefore crucial. In this case, the ongoing assessment of the quality of one-step ahead forecasts based on posterior predictive
distributions \( p(y_{k+1}|y_k) \) (Dawid, 1984) is an appealing approach. Although test statistics based on the ranked probability scoring rule have been used in Section 4, there are alternate scoring rules that could be considered. The logarithmic score, for a predictive distribution \( p(\cdot) \) and the subsequently realised observation \( y \) is defined to be:

\[
sl_{\text{log}}(P, y) = -\log(p(y)).
\]

Under an SMC scheme, this is approximated by

\[
sl_{\text{log}}(P, y) = -\log(p(y_{k+1}|y_{1:k})) = -\log \left( \int_{\Theta} \pi(\theta|y_{1:k}) p(y_{k+1}|\theta) d\theta \right) 
\approx -\log \left( \sum_{j} \omega_{k}^{(j)} p(y_{k+1}|\theta_{k}^{(j)}) / \sum_{j} \omega_{k}^{(j)} \right) = \log \left( \frac{\sum_{j} \omega_{k}^{(j)} \tilde{\omega}_{k+1}^{(j)}}{\sum_{j} \tilde{\omega}_{k+1}^{(j)}} \right).
\]

Where weights \( \omega_{k}^{(j)} \) and \( \tilde{\omega}_{k+1}^{(j)} \) are routinely calculated as part of the SMC algorithm in Section 3.1 (equation (10)). This provides an additional benefit of SMC as this would require additional computation if the posterior is derived using MCMC. This is an illustration of a more general advantage in that the SMC algorithm generates the full sequence of posterior distributions \( \pi(\theta|y_{1:k}) \) making estimation of the predictive distributions a mere matter of simulation. If the MCMC analyses are not carried out daily, then these are not readily available. Further detail on the calculation and interpretation of these posterior predictive methods is included in the Web Appendix.

9.3 Data availability

To this point, the discussion has centred on algorithmic development and the availability of all data sources in a timely manner has been assumed. Particularly crucial to the feasibility of real-time modelling is the role of serological data. Figure 8 shows epidemic projections sequentially made using only noisy primary care consultation data without serological information: a clear and realistic picture of the epidemic is not available until the epidemic has almost entirely been observed. This poses key questions: are serological samples going to be available in a timely manner, in sufficient quantity and quality, and in the right format? In reality, serological data can be slow to come online. A test has to be developed to identify the antibodies of a (probably) novel virus in blood sera; and there needs to be sufficient time to test samples and report results according to a protocol that ensures unbiased data collection and analysis. In a realistic setting, to accommodate ‘slow’ serological data, particles will have to store the full historical values of \( S_{t_{k+1}} \) in addition to the current state of the epidemic.

In the event that external information on any of the epidemic parameters becomes available that cannot be formally incorporated into the model, in SMC this information can be accounted for through adjustment of the relevant prior distributions. Particles would be reweighted according to the ratio of the new prior to old prior and, if doing so causes a significant drop in the ESS, the particle set can be rejuvenated in the usual way. This provides a clear advantage of SMC over MCMC where the entire dataset would have to be re-analysed.
Figure 8: Sequential epidemic forecasts based on increasing amounts of data, not including serological data. The dark shaded areas represent a current forecast, with the light shaded areas the forecast at the time of the previous plot. Vertical red dashed lines indicate the current time, the black dashed line indicates the time of the previous prediction

9.4 Model Specification

The focus of this work was not to examine in-depth the suitability of the presented epidemic model. However, it has brought to light some features, particularly of the observation model, that can make real-time inference problematic. Changepoints in the values of parameters at key times can limit the performance of both SMC and MCMC algorithms. Often these changepoints represent some landmark time in the epidemic, such as the initiation of a pandemic intervention. However, other changepoints might be necessary simply to maintain an adequate fit to the data and the need for such changepoints has to be identified in real time. Nemeth et al. (2014) provide a prescription for identifying such changepoints and would represent a natural extension to this work. Alternatively, to improve the robustness of the inferences, the piecewise linear quantities describing population reporting behaviour \( (p_{\text{doc}}, B) \) could be described by linked stochastic noise processes. This has the potential to limit the sensitivity of estimates to the presence of changepoints that are not, for whatever reason, foreseeable.

9.5 Concluding Remarks

In answer to the question initially posed, we have provided a recipe for online tracking of an emergent epidemic using imperfect data from multiple sources. We have discussed many of the challenges to efficient inference, with particular focus on scenarios where the available information is rapidly evolving and is subject to sudden shocks. We have focused on an epidemic scenario likely to arise in the UK. Nevertheless, our approach addresses modelling concerns common globally (e.g. Shaman and Karspeck, 2012; Wu et al., 2010; Shubin et al., 2016; te Beest et al., 2015) and can form a flexible basis for real-time modelling strategies elsewhere.
Real-time modelling is, however, more than just a computational problem. It does require the timely availability of relevant data, but also needs a sound understanding of any likely biases, and effective interaction with experts. In any country, only interdisciplinary collaboration between statisticians, epidemiologists and database managers can turn cutting edge methodology into a critical support tool for public health policy.

Acknowledgements

This work was supported by the National Institute for Health Research (HTA Project: 11/46/03) (PJB, DDA and GOR) the UK Medical Research Council Unit Programme Numbers U105260566 (DDA) and MC_UP_1302/3 (LW and BDMT), and Public Health England (DDA and RGP).

References

Ahrens, H. (1976) Multivariate variance-covariance components (MVCC) and generalized intraclass correlation coefficient (GICC). *Biometrical J*, 18, 527–533.

Ball, F. (1983) A Threshold Theorem for the Reed-Frost Chain-Binomial Epidemic. *Journal of Applied Probability*, 20, 153–157.

Banterle, M., Grazian, C., Lee, A. and Robert, C. P. (2015) Accelerating Metropolis-Hastings algorithms by Delayed Acceptance. *arXiv*, 1503.00996v2, 27. URL: http://arxiv.org/abs/1503.00996v2.

Bettencourt, L. M. A. and Ribeiro, R. M. (2008) Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS ONE*, 3, e2185.

Birrell, P. J., Ketsetzis, G., Gay, N. G., Cooper, B. S., Presanis, A. M., Harris, R. J., Charlett, A., Zhang, X.-S., White, P., Pebody, R. G. and De Angelis, D. (2011) Bayesian modelling to unmask and predict the influenza A/H1N1pdm dynamics in London. *Proc. Natn. Acad. Sci. USA*, 108, 18238–18243.

Camacho, A., Kucharski, A., Aki-Sawyerr, Y., White, M. A., Flasche, S., Baguelin, M., Pollington, T., Carney, J. R., Glover, R., Smout, E., Tiffany, A., Edmunds, J. J. and Funk, S. (2015) Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: a Real-time Modelling Study. *PLoS Curr*, 7, Web.

Carpenter, J., Clifford, P. and Fearnhead, P. (1999) Improved particle filter for nonlinear problems. *IEE P-Radar Son Nav*, 146, 2+. 

Cauchemez, S., Boëlle, P.-Y., Thomas, G. and Valleron, A.-J. (2006) Estimating in real time the efficacy of measures to control emerging communicable diseases. *Am J Epidemiol*, 164, 591–597.

Chopin, N. (2002) A sequential particle filter method for static models. *Biometrika*, 89, 539–552.

Czado, C., Gneiting, T. and Held, L. (2009) Predictive Model Assessment for Count Data. *Biometrics*, 65, 1254–1261.
Dawid, A. P. (1984) The Prequential Approach. *J. R. Statist. Soc. A*, 147, 278–292.

Del Moral, P., Doucet, A. and Jasra, A. (2006) Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, 68, 411–436.

Donner, A. and Koval, J. J. (1980) The estimation of intraclass correlation in the analysis of family data. *Biometrics*, 36, 19–25.

Doucet, A. and Johansen, A. M. (2009) A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, 12, 656–704.

Dukic, V., Lopes, H. F. and Polson, N. G. (2012) Tracking epidemics with google flu trends data and a state-space seir model. *J. Am. Statist. Ass.*, 107, 1410–1426.

Dureau, J., Kalogeropoulos, K. and Baguelin, M. (2013) Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics*, 14, 541–555.

Farah, M., Birrell, P., Conti, S. and De Angelis, D. (2014) Bayesian emulation and calibration of a dynamic epidemic model for a/h1n1 influenza. *J. Am. Statist. Ass.*, 109, 1398–1411.

Fearnhead, P. (2002) MCMC, sufficient statistics and particle filters. *J Comput Graph Stat*, 11, 848–862.

Geyer, C. J. (1991) Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: The 23rd Symposium on the Interface*, 156–163. Interface Foundation of North America.

Gilks, W. R. and Berzuini, C. (2001) Following a moving target - Monte Carlo inference for dynamic Bayesian models. *J. R. Statist. Soc. B*, 63, 127–146.

Gneiting, T. and Raftery, A. E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Statist. Ass.*, 102, 359–378.

Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc-F*, 140, 107–113.

Held, L., Meyer, S. and Bracher, J. (2017) Probabilistic forecasting in infectious disease epidemiology: The Thirteenth Armitage Lecture. *Statistics in Medicine, In press*.

Jasra, A., Stephens, D. A., Doucet, A. and Tsagaris, T. (2011) Inference for Lévy-driven Stochastic Volatility Models via Adaptive Sequential Monte Carlo. *Scand. J. Statist.*, 38, 1–22.

Jasra, A., Stephens, D. A. and Holmes, C. C. (2007) On population-based simulation for static inference. *Stat Comput.*, 17, 263–279.

Jewell, C. P., Kypraios, T., Christley, R. M. and Roberts, G. O. (2009) A novel approach to real-time risk prediction for emerging infectious diseases: a case study in Avian Influenza H5N1. *Prev vet med*, 91, 19–28.

Kantas, N., Beskos, A. and Jasra, A. (2014) Sequential monte carlo methods for high-dimensional inverse problems: A case study for the navier–stokes equations. *SIAM/ASA Journal on Uncertainty Quantification*, 2, 464–489.
Konishi, S., Khatri, C. and Rao, C. R. (1991) Inferences on multivariate measures of interclass and intraclass correlations in familial data. *J. R. Statist. Soc. B*, 649–659.

Liang, F. and Wong, W. H. (2001) Real-Parameter Evolutionary Monte Carlo With Applications to Bayesian Mixture Models. *J. Am. Statist. Ass.*, 96, 653–666.

Liu, J. S. and Chen, R. (1995) Blind Deconvolution via Sequential Imputations. *J. Am. Statist. Ass.*, 90, 567–576.

— (1998) Sequential Monte Carlo Methods for Dynamic Systems. *J. Am. Statist. Ass.*, 93, 1032–1044.

Meester, R., de Koning, J., de Jong, M. C. and Diekmann, O. (2002) Modeling and real-time prediction of classical swine fever epidemics. *Biometrics*, 58, 178–184.

Neal, R. M. (1996) Sampling from multimodal distributions using tempered transitions. *Stat Comput*, 6, 353–366.

Nemeth, C., Fearnhead, P. and Mihaylova, L. (2014) Sequential Monte Carlo Methods for State and Parameter Estimation in Abruptly Changing Environments. *IEEE T Signal Process.*, 62, 1245–1255.

Ong, J. B. S., Chen, M. I.-C., Cook, A. R., Chyi, H., Lee, V. J., Pin, R. T., Ananth, P. and Gan, L. (2010) Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore. *PloS one*, 5, e10036.

Roberts, G. O. and Rosenthal, J. S. (2001) Optimal Scaling for Various Metropolis-Hastings Algorithms. *Stat Sci*, 16, 351–367.

Scheuerer, M. and Hamill, T. M. (2015) Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities. *Monthly Weather Review*, 143, 1321–1334.

Scientific Pandemic Influenza Advisory Committee: Subgroup on Modelling (2011) Modelling Summary. URL http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@ab/documents/digitalasset/dh_127275.pdf. SPI-M-O Committee document (Accessed 4 February, 2016).

Seillier-Moiseiwitsch, F. and Dawid, A. P. (1993) On testing the validity of sequential probability forecasts. *J. Am. Statist. Ass.*, 88, 355–359.

Shaman, J. and Karspeck, A. (2012) Forecasting seasonal outbreaks of influenza. *Proc. Natn. Acad. Sci. USA*, 109, 20425–20430.

Sherlock, C., Fearnhead, P. and Roberts, G. O. (2010) The random walk metropolis: Linking theory and practice through a case study. *Stat Sci*, 25, 172–190.

Shubin, M., Lebedev, A., Lyytikäinen, O. and Auranen, K. (2016) Revealing the True Incidence of Pandemic A(H1N1)pdm09 Influenza in Finland during the First Two Seasons – An Analysis Based on a Dynamic Transmission Model. *PLoS Comput. Biol.*, 12, 1–3.

Skvortsov, A. and Ristic, B. (2012) Monitoring and prediction of an epidemic outbreak using syndromic observations. *Math Biosci.*, 240, 12–19.
te Beest, D. E., Birrell, P. J., Wallinga, J., De Angelis, D. and van Boven, M. (2015) Joint modelling of serological and hospitalization data reveals that high levels of pre-existing immunity and school holidays shaped the influenza A pandemic of 2009 in the Netherlands. *J. R. Soc. Interface, 12*, 20141244+.

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

Wu, J. T., Cowling, B. J., Lau, E. H. Y., Ip, D. K. M., Ho, L.-M., Tsang, T., Chuang, S.-K., Leung, P.-Y., Lo, S.-V., Lio, S.-H. and Riley, S. (2010) School Closure and Mitigation of Pandemic (H1N1) 2009, Hong Kong. *Emerg Infect Dis, 16*, 538–541.