arXiv:1203.5950v2 [stat.AP] 3 Nov 2012

Capturing the time-varying drivers of an epidemic using stochastic dynamical systems

JOSEPH DUREAU*, KONSTANTINOS KALOGEROPOULOS

Department of Statistics , London School of Economics and Political Science, UK

j.dureau@lse.ac.uk

MARC BAGUELIN

Immunisation, Hepatitis and Blood Safety Department, Health Protection Agency, Centre for the Mathematical Modeling of Infectious Diseases, London School of Hygiene and Tropical Medicine

November 6, 2012

Abstract

Epidemics are often modelled using non-linear dynamical systems observed through partial and noisy data. In this paper, we consider stochastic extensions in order to capture unknown influences (changing behaviors, public interventions, seasonal effects etc). These models assign diffusion processes to the time-varying parameters, and our inferential procedure is based on a suitably adjusted adaptive particle MCMC algorithm. The performance of the proposed computational methods is validated on simulated data and the adopted model is applied to the 2009 H1N1 pandemic in England. In addition to estimating the effective contact rate trajectories, the methodology is applied in real time to provide evidence in related public health decisions. Diffusion driven SEIR-type models with age structure are also introduced. population epidemic model; time-varying parameters; Bayesian inference; Particle MCMC

1 Introduction

Epidemic models are often used to simulate disease transmission dynamics, detect emerging outbreaks (Unkel *and others*, 2012), and assess public health interventions (Boily *and others*, 2007). In order to capture the dynamics of epidemics, the main focus is generally made on their intrinsically

[0]To whom correspondence should be addressed.

dynamic elements such as the depletion of susceptibles or the population immunity evolution. Nevertheless, there are time-varying extrinsic factors that are crucial to the epidemic course. These may include social cycles (holidays), public interventions and climatic variations. This has been illustrated for diseases such as cholera, malaria (Cazelles *and others*, 2005; Ionides *and others*, 2006) or influenza (Shaman and Kohn, 2009). These studies were conducted either by relating climatic and incidence time-series (Cazelles *and others*, 2005), which does not disentangle the effect of intrinsic and extrinsic factors, or by experimentally assessing the virus resistance in different climatic conditions (Shaman and Kohn, 2009) requiring an extrapolation to the population scale. Overall, the time-varying nature of epidemics poses a challenging statistical problem stressing the need for suitable computational tools (Ferguson, 2007).

This paper considers a flexible modelling framework that encompasses time-varying aspects of the epidemic via stochastic differential equations. We aim at providing robust inferential procedures, incorporating the uncertainty associated with key parameters and accounting for data and model limitations. In order to provide an accurate and feasible computational toolbox, we provide Markov Chain Monte Carlo (MCMC) algorithms utilising recent developments such as particle MCMC (PMCMC) algorithms (Andrieu *and others*, 2010) and adaptive techniques (Roberts and Rosenthal, 2009). Modelling aspects are presented in Section 2, while the computational framework is presented in Section 3. In Section 4 we evaluate the performance of the proposed adaptive PMCMC schemes on simulated data. In Section 5 we present various applications of the methodology to the 2009 A/H1N1 pandemic, and conclude, in Section 6, with some relevant discussion. Further simulations can be found in the Supplementary Materials.

2 Modelling framework

2.1 Epidemic models with time-varying coefficients

We adopt a SEIR model as a guide in this paper, although the methodology can be applied to other dynamical systems. The model is set in (1); S accounts for susceptible, E for infected but not infective, I for infective, and R for removed individuals. New infections occur at a rate $\beta S_t \frac{I_t}{N}$, implying that the susceptible individuals make effective contacts at rate β (the effective contact rate), and only a fraction $\frac{I_t}{N}$ of these contacts are made with infective individuals. The average period spent in compartments E and I is given by k^{-1} and γ^{-1} respectively.

$$\frac{dS_t}{dt} = -\beta S_t \frac{I_t}{N}, \quad \frac{dE_t}{dt} = \beta S_t \frac{I_t}{N} - kE_t, \quad \frac{dI_t}{dt} = kE_t - \gamma I_t, \quad \frac{dR_t}{dt} = \gamma I_t \quad (1)$$

The basic reproduction number, R_0, represents the number of secondary infections from a primary infected individual in a fully susceptible population. A related quantity is the effective reproduction number, R_t, refers to the number of secondary cases from an infected individual at time t. R_t is a context-dependent quantity of high interest to policy makers as it indicates the possibility for the epidemic to grow ($R_t > 1$) or to decrease ($R_t < 1$) (Anderson and May, 1992).

Epidemic models can be quite detailed (including individual characteristics, geographic information etc.) or basic, such as the SEIR model, that geographically aggregates the cases and assumes deterministic transmission processes, occurring at a given frequency each time infected and susceptible meet. The latter are easier to estimate and interpret, but are based on strong assumptions that could lead to poor inference. In this paper we adopt stochastic extensions of the deterministic SEIR models. The additional dynamic error is likely to contain structural mis-specifications and can subsequently be explored and potentially revised. We focus on large-scale epidemics, for which random effects in transmission processes can be considered to be well-approximated deterministically (Kurtz, 1981). We adopt the paradigm that attributes the model limitations mainly to the time varying nature of the effective contact rate, henceforth denoted as β_t, rather than to the variability in individual characteristics or in transmission processes.

An early approach to estimate R_t can be found in Fine and Clarkson (1982). It can be implemented through discrete generation models or by reconstructing the chain of transmission (Cauchemez *and others*, 2006; Griffin *and others*, 2011). However, as R_t estimates contain both the effects of evolving transmissibility and immunity, quantitative conclusions can hardly be generalised to situations where the immunological situation is different. We therefore concentrate on estimating β_t rather than R_t. A number of approaches use a finite-dimension function space for the trajectory of β_t. Low-dimensional examples can be found in Cauchemez *and others* (2008), in which β_t is modeled as a piece-wise linear function. In some higher-complexity models, as in Cauchemez and Ferguson (2008) and Ionides *and others* (2006), β_t is estimated freely with a few-weeks resolutions. Loosely speaking, as the number of parameters for the trajectory of β_t increases, model-induced biases fade out at the expense of the variance. A compromise is required to improve robustness and is often controlled through a regularising parameter. For example, in He *and others* (2011), β_t is estimated using cubic splines, calibrated via AIC.

2.2 Diffusion driven epidemic models

We consider models where diffusion processes are used for some of the coefficients in (1). Although alternative formulations are possible, as discussed

in Section 2.1, we focus on β_t to get

$$\begin{cases} \frac{dS_t}{dt} &= -\beta_t S_t \frac{I_t}{N}, \quad \frac{dE_t}{dt} = \beta_t S_t \frac{I_t}{N} - kE_t, \quad \frac{dI_t}{dt} = kE_t - \gamma I_t, \quad \frac{dR_t}{dt} = \gamma I_t, \\ dx_t &= \mu_x(x_t, \theta_x)dt + \sigma_x(x_t, \theta_x)dB_t, \quad x_t = h(\beta_t), \end{cases} \tag{2}$$

where $\mu_x(\cdot)$ denotes the drift, $\sigma_x(\cdot)$ the volatility and $h(\cdot)$ is a positive-valued function. The assigned diffusion may capture features such as behaviour changes, preventive measures, seasonal effects, holidays etc. When prior knowledge on β_t is available, it can be reflected in $\mu_x(\cdot)$ and $\sigma_x(\cdot)$; e.g. if the contact rate is expected to converge, an Ornstein Uhlenbeck process can be chosen. Other options may include a sigmoid or a sinusoidal form; see for example (Rasmussen *and others*, 2011). In absence of prior information or when the researcher wants to impose little restrictions, a Brownian motion can be used, with $\mu_x(\cdot) \equiv 0$ and $\sigma_x(\cdot) \equiv \sigma$ (i.e. $\theta_x = \sigma$). This model, with $h(\cdot) \equiv \log(\cdot)$, is henceforth denoted as BM. The obtained output can be either reported or used as an exploratory tool to construct a more structured model; see Section 5.3 for an application. The choice of BM implies a continuous, yet non-differentiable, path satisfying the Markov property. In cases where β_t is believed to evolve as a smooth function in time, higher order Brownian motions could be used. Loosely speaking, these may be regarded as equivalent to non-parametric approaches such as cubic splines (Wahba, 1990), with the model in (2) imposing a prior on β_t and σ being a regularising factor. The rate β_t can be perceived as a product of a smooth and a rough component; the former being a population average of the intrinsic transmission procedure and latter containing extrinsic factors such as the amount of contact among individuals. It is therefore important to build a framework that contains both smooth and rough models.

The above model can be estimated with an Extended Kalman Filter (EKF), as in Cazelles and Chau (1997). EKF allows for fast computations, but is based on Taylor and Gaussian approximations whose error could be non-negligible; see Supplementary Materials for a relevant simulation experiment. Nevertheless, the EKF can still be used as a tool to construct efficient proposal distributions for MCMC schemes. It can also be used to optimize sequential Monte Carlo (SMC) algorithms, but either at a strong computational cost (Särkkä and Sottinen, 2008) or crude time discretisations (Dukic *and others*, 2009). Next, we develop a general framework for efficient MCMC schemes that allow for good approximations.

3 Data augmentation via MCMC for diffusion driven epidemic models

This section presents a general inferential framework for diffusion-driven epidemic models. We adopt the Bayesian paradigm to incorporate parameter

uncertainty and prior information in the estimates of β_t trajectories. The problem can also be cast as estimating partially observed hypoelliptic diffusions, thus presenting various difficulties (Pokern *and others*, 2009). We begin by setting the model and justifying the need for data augmentation. Existing MCMC algorithms are considered but they can lead to extremely inefficient MCMC chains. We address the issue by taking advantage of the specific model structure to construct adaptive PMCMC schemes.

3.1 Model and data augmentation setup

For ease of exposition we focus on models satisfying (2), but the framework covers models with different ODE systems or more time-varying coefficients, as in Section 5.3. Being in continuous time, t can take any value between t_0 and t_n. We denote the path of the ODE states vector $V_t = \{S_t, E_t, I_t, R_t\}$ between observation times t_i and t_j by $V_{i:j}$. The data, $y_{1:n} = \{y_{t_1}, .., y_{t_n}\}$, usually provide information for I_t at specific times (prevalence data) or for integrals of V_t (incidence data). In either case, we assume that they are obtained with error as the collection procedure is typically associated with additional uncertainty. The noise distribution is denoted with $\mathbb{P}_y$ with density $f(y_{1:n}|V_{0:n}, \theta_y)$. Note that, in the model of (2), V_t can be written as a deterministic function, $g(\cdot)$, of x_t and the parameters $\theta_v = (k, \gamma, V_0)$. This function is the solution of the ODE and can be written as an intractable time integral involving x_t. Hence, the model becomes

$$\begin{cases} dx_t & = \mu_x(x_t, \theta_x)dt + \sigma_x(x_t, \theta_x)dB_t \\ y_{1:n}|V_{0:n}, \theta_y & \sim \mathbb{P}_y(y_{1:n}|V_{0:n}, \theta_y), \quad V_{0:n} = g(x_{0:n}, \theta_v) \end{cases} \tag{3}$$

Denote with $\mathbb{P}_x$ the distribution of the diffusion x_t defined from the SDE above. We require the existence of a unique weak solution which translates into some mild assumptions on $\mu_x(.)$ and $\sigma_x(.)$; e.g. locally Lipschitz with a linear growth bound, see for example Øksendal (2003). The distribution of $\mathbb{P}_x$ may also be viewed as a prior on x_t, or else β_t. The model can now be defined from $\mathbb{P}_y$, $\mathbb{P}_x$, and the assigned priors on $\theta = \{\theta_y, \theta_v, \theta_x\}$, denoted by $\pi(\theta)$

$$\pi(x_{0:n}, \theta|y_{1:n}) \propto f(y_{1:n}|V_{0:n}, \theta_y) \times d\mathbb{P}_x \times \pi(\theta) \tag{4}$$

Given direct observations on x_t, it would have been possible to draw approximation-free inference on $d\mathbb{P}_x$ using the approach of Beskos *and others* (2006). However, this is not possible in our case given the non-linear functionals in $g(\cdot)$ that render (3.4) intractable.We proceed by discretizing the path of x_t, and therefore of β_t and V_t. More specifically, we introduce m points between each pair of successive observation times t_i and t_{i+1} ($i = 0, 1, \ldots, n-1$). When referring to the discrete representation of a path, the superscript *dis* will be used; for example for a step $\delta = \frac{1}{m+1}$, the discrete skeleton of x_t will

be denoted by $x_{0:n}^{dis} = \{x_0, x_\delta, x_{2\delta}, \ldots, x_{t_n}\}$. The presence of $x_{0:n}^{dis}$ allows for approximations of (4) through the Euler-Maruyama scheme to evaluate $d\mathbb{P}_x$

$$\begin{cases} p(x_{\delta:n}^{dis}|x_0, \theta_x) \;=\; \prod_{i:\, t_0 < i\delta \leq t_n} p(x_{i\delta}|x_{(i-1)\delta}, \theta_x), \\ x_{i\delta}|x_{(i-1)\delta} \;\sim\; \mathcal{N}\left\{x_{(i-1)\delta} + \delta\mu_x(x_{(i-1)\delta}, \theta_x),\; \delta\sigma_x(x_{(i-1)\delta}, \theta_x)^2\right\}. \end{cases} \tag{5}$$

Moreover, given $x_{0:n}^{dis}$,the ODE can be solved numerically to obtain $V_{0:n}^{dis}$ and evaluate $f(\cdot)$. The approximation error can be made arbitrarily small by increasing the user-specified parameter m.

3.2 Data augmentation via Gibbs schemes

Model (3) can be put in the context of Chib *and others* (2006), Golightly and Wilkinson (2008) or Kalogeropoulos (2007). In these approaches, a Gibbs scheme can be used to sample from the joint posterior in (4) of $x_{0:n}^{dis}$ and θ. The data augmentation algorithm alternates between drawing $x_{0:n}^{dis}$ given θ, and updating θ conditional on the augmented path $x_{0:n}^{dis}$. The MCMC protocol ensures that the chain provides samples from the marginal posteriors of $x_{0:n}^{dis}$ and θ. Nevertheless, the properties of the algorithm may become unacceptably poor. There are two essential issues associated with such schemes. The first concerns the non-trivial step of sampling on the diffusion pathspace of x_t. The second problem is caused by the high posterior correlations between $x_{0:n}^{dis}$ and θ, leading to reducible chains as m increases (Roberts and Stramer, 2001).

The majority of the literature on data augmentation schemes for diffusions handles the conditional updates of $x_{0:n}^{dis}$ with an independence sampler. As it is difficult to find good proposal distributions for the entire $x_{0:n}^{dis}$, the path is usually split into blocks. Overlapping blocking strategies are essential to ensure that all points are updated and continuity of the path is retained. An alternative way to update $x_{0:n}^{dis}$ is to use the particle filter via the Particle Gibbs algorithm of Andrieu *and others* (2010). But unless the issue of high posterior correlation between $x_{0:n}^{dis}$ and θ is resolved, none of these schemes will improve the overall MCMC performance. The problem is caused by the quadratic variation process of x_t that identifies θ_x. For $\sigma_x(x_s, \theta_x) \equiv \sigma$ we get

$$\lim_{\delta \to 0} \sum_{i:\, t_0 < i\delta \leq t_n} (x_{i\delta} - x_{(i-1)\delta})^2 = \int_{t_0}^{t_n} \sigma^2 ds = \sigma^2(t_n - t_0) \tag{6}$$

Thus, the conditional posterior of σ converges to a point mass as δ tends to 0. In practice this translates into an increasingly slow MCMC algorithm with a convergence rate of $O(m)$ (Roberts and Stramer, 2001). Schemes with a fixed m (Cori *and others*, 2009) could work in some occasions but the approximation error could be substantial. In some cases, the problem can be tackled with suitable reparametrisation. The approach of Roberts and Stramer

(2001) involves transforming to a diffusion $\dot{x}_t$ with unit volatility. An alternative scheme is offered by Chib *and others* (2006) where the driving Brownian motion of x_t is being used. In these algorithms the ODE states vector $V_{0:n}^{dis}$ becomes a function of σ, $\dot{x}_{0:n}$ and θ_v. Hence, in a Metropolis step, every proposed value of σ^* is associated with the corresponding values of $V_{0:n}^{dis*}$. This succeeds into breaking the perfect dependence between $V_{0:n}^{dis}$ and σ, even for $m \to \infty$. But since components of $V_{0:n}^{dis}$ (or functionals thereof) are observed with error, the associated proposed values $V_{0:n}^{dis*}$ should be close to the data for the move to be accepted. As the observation error becomes small and the data increase, this becomes increasingly difficult and leads to very small moves for σ and poor MCMC mixing. More details and simulations supporting this argument are provided in the Supplementary Materials (Appendix E). Consequently, we overcome this issue by updating $x_{0:n}^{dis}$ and θ jointly via the PMCMC algorithm, which is essential as it is not straightforward to implement joint updates with the other approaches mentioned in this section.

3.3 Adaptive Particle Markov Chain Monte Carlo algorithms

Particle filters are SMC algorithms used to recursively explore conditional densities in state space models (Doucet and Johansen, 2011). For given values of θ, N particles $(\tilde{x}_i^j)$ are sequentially propagated from t_0 to t_n. In various time steps t_i, the trajectories that best fit the data $y_{1:i}$ are given more weight through resampling. Algorithm 1 shows how they can be applied in our context. The quantity $L^{i+1}(\theta)$ provides unbiased estimates of $p(y_{1:i}|\theta)$

Algorithm 1 Particle Filter algorithm

Initialise: Set $L^0(\theta) = 1$, $W_0^j = \frac{1}{N}$, sample $(\tilde{x}_0^j)_{j=1,\dots,N}$ from $p(x_0|\theta)$ and calculate $(\tilde{V}_0^j)_{j=1,\dots,N}$ by solving the ODE (for example with the Euler scheme)
for $i = 0$ to $n-1$ **do**
  for $j = 1$ to N **do**
    Sample $(\tilde{x}_{i:i+1}^j)$ from (5) and calculate $(\tilde{V}_{i:i+1}^j)$ by solving the ODE
    Set $\alpha^j = f(y_{i+1}|\tilde{V}_{0:i+1}^j)$
  end for
  Set $W_{i+1}^j = \frac{\alpha^j}{\sum_{k=1}^N \alpha^k}$, and $L^{i+1}(\theta) = L^i(\theta) \times \frac{1}{N}\sum \alpha^j$
  Resample $(\tilde{V}_{0:i+1}^j, \tilde{x}_{0:i+1}^j)_{j=1,\dots,N}$ according to (W_{i+1}^j),
end for

and the resampling step is essential to control the variance of that estimate over time. Algorithm 1 also provides a random sample from $p(x_{1:i}|y_{1:n}, \theta)$. In order to sample from $\pi(x_{1:n}, \theta|y_{1:n})$, the PMCMC algorithm can be used. PMCMC was introduced in Andrieu *and others* (2010) and successfully inte-

grates particle filters in MCMC algorithms. Its implementation is presented in Algorithm 2. The issues of Section 3.2 are now addressed as $x_{0:n}^{dis}$ and θ

Algorithm 2 Particle MCMC algorithm (particle Marginal Metropolis Hastings version)

Initialise: Set current θ value, $\tilde{\theta}$, to an initial value. Use Particle Smoother (PS) according to Algorithm 1 to compute $\hat{p}(y_{1:n}|\tilde{\theta}) = L(\tilde{\theta})$ and sample $\tilde{x}_{1:n}^{\tilde{\theta}}$ from $p(x_{1:n}|y_{1:n}, \tilde{\theta})$

for $It = 1$ to $NIterations$ **do**

  Sample $\tilde{\theta}^*$ from $Q(\tilde{\theta}, .)$

  Use PS to compute $L(\tilde{\theta}^*)$ and sample $\tilde{x}_{1:n}^{\tilde{\theta}^*}$ from $\hat{p}(x_{1:n}|y_{1:n}, \tilde{\theta}^*)$

  Do $\tilde{\theta} = \tilde{\theta}^*$ (and $\tilde{x}_{1:n}^{\tilde{\theta}} = \tilde{x}_{1:n}^{\tilde{\theta}^*}$) with probability $1 \wedge \frac{L(\tilde{\theta}^*)Q(\tilde{\theta}^*,\tilde{\theta})}{L(\tilde{\theta})Q(\tilde{\theta},\tilde{\theta}^*)}$

  Record $\tilde{\theta}$ and $\tilde{x}_{1:n}^{\theta}$

end for

are sampled jointly. In other words $x_{0:n}^{dis}$ is being numerically integrated out, while a sample from its posterior is obtained at each MCMC iteration.

While the PMCMC algorithm is theoretically valid even for a single particle, large values of N are usually required for reasonably stable acceptance rates and large moves in the θ space; see the Supplementary Materials for a relevant simulation exercise. It is therefore essential to update the d-dimensional θ at once, marking the proposal $Q(\theta, .)$ crucial to the overall MCMC performance. In this paper we propose to use the adaptive Metropolis algorithm of Roberts and Rosenthal (2009). After transforming the parameters to take values in the real line we use a Normal distribution centered at the current value of θ and with covariance given by $\epsilon\Sigma$. Static random walk metropolis proposals set $\Sigma = I_d$ or $\Sigma = \hat{\Sigma}$ and tune ϵ to obtain acceptance rate of 0.234. Adaptive schemes change the value ϵ for each iteration i through diminishing adaptation; e.g. by $\epsilon_{i+1} = \exp\{\log(\epsilon_i) + \alpha_1^n(\text{AccRate} - 0.234)\}$ where $\alpha_1 = 0.999$ and 'AccRate' denotes the acceptance rate up to iteration i. The covariance matrix Σ_{i+1} can also be updated as

$$\alpha_2 \mathcal{N}\left(\theta, \epsilon\frac{2.38^2}{d}\Sigma_0\right) + (1-\alpha_2)\mathcal{N}\left(\theta, \epsilon\frac{2.38^2}{d}\Sigma_i\right)$$

where α_2 is usually set to 0.05, Σ_i is the posterior covariance matrix estimated by the draws up to i and Σ_0 should be specified in advance. In this paper we enhance the above adaptive algorithms utilising information from the EKF to estimate the covariance $\hat{\Sigma}$ or Σ_0. One choice, EK-Mode, is the observed information matrix at the mode identified by EKF, evaluated through numerical differentiation. Another choice, EK-MCMC, is to run an approximate MCMC scheme based on the EKF approximation of the likelihood and compute the posterior covariance from the draws. Note

that the computational burden of these methods is marginal with regards to the PMCMC. As demonstrated in Section 4, the use of EKF can result in substantial improvement.

4 Simulation Experiments

The proposed algorithms are illustrated and tested on simulated data in this section. We focus on the BM model, where $\log(\beta_t)$ follows a Brownian motion with volatility σ, corresponding to the case of little information on the shape of β_t. The trajectories of β_t were drawn either from the BM model itself (experiment 1) or from a deterministic sigmoid curve (experiment 2). The data y_i, $i = 1, \ldots, 50$ represent noisy observations of weekly new cases of the epidemic $\int_{week\ i} kE_t dt$. We complete the model by assigning a Normal distribution to each $\log(y_i)$ with mean $\log(\int_{week\ i} kE_t dt)$ and variance τ^2. The parameters were tuned to obtain realistic epidemic incidence curves, and observations were generated setting $\tau = 0.1$. The assigned priors were informative for k, γ and $R(t_0)$ and vague for $E(t_0)$, $I(t_0)$, σ and τ, as in Section 5.1. We used 3,000 particles and 100,000 MCMC iterations after a long burn-in period. Fig. 1 shows estimates and 95% pointwise credible intervals of the path, provided by the adaptive PMCMC initialized with EK-MCMC. The posterior output is in good agreement with the simulation trajectories suggesting that the underlying trajectory of β_t can be estimated reasonably well from the partial and noisy observations considered. More can be found in the Supplementary Materials (appendix C), where we also considered a value of $\tau = 0.05$ and obtained similar results.

Next, we use the data of experiment 1 to compare the proposed adaptive PMCMC schemes. Comparison is made in terms of the effective sample size ESS $= (1 + 2\sum_{i\geq 1} \eta(i))^{-1}$, with $\sum_i \eta(i)$ being the sum of the lagged sample auto-correlations, as in Geyer (1992). We record the minimum ESS among the MCMC components and multiply by 100 to monitor the percentage of the total iterations that can be considered as independent. We consider three covariance matrices for each of the two adaptive algorithms defined in Section 3.3: I_d and the ones from EK-Mode and EK-MCMC. For the schemes that adapt ϵ the minimum ESS was 0.008% (I_d), 0.19% (EK-Mode) and 0.54%(EK-MCMC), whereas for the schemes that adapt Σ we got 0.57%, 1.24% and 1.38% respectively. Clearly, adapting Σ is crucial to obtain a reasonable performance, unless the matrices obtained from EK-Mode or EK-MCMC are used. The proposed adaptive algorithms induce substantial improvement that is expected to intensify as the dimension of θ increases.

5 The 2009 A/H1N1 pandemic

5.1 Data, model and estimates

The proposed methodology is illustrated on data from the A/H1N1(2009) pandemic in England between June and December 2009. The data consists of estimates of weekly ILI cases $y_{1:n}$ given by the Health Protection Agency (Baguelin *and others*, 2010). The estimates were obtained from the recorded ILI cases among a selected sample of GPs. They accounted for over-reporting due to similarities in symptoms with other respiratory diseases, based on subsequent virological positivity tests. Corrections for asymptomatic infections and the patients' propensity to consult were also made. Overall the two datasets are different by a multiplicative coefficient $c = 10$, whose value is also supported by a further serological survey (Miller *and others*, 2010). In our analysis c is initially held fixed to 10, but this choice is explored further in Section 5.2. We adopt a model that admits noisy data to reflect the associated uncertainty. The noise model of Section 4 was used, combined with a BM formulation of $\mathbb{P}_x$. Vague priors, $N_{>0}(0, 10^6)$, were put on τ, σ and β_0. The priors for k and γ were obtained from additional data sources (Baguelin *and others*, 2010), the results of which are summarised through Normal distributions that place 95% probability in a symmetric manner between 1.55 and 1.63 days for the latent period k^{-1}, and between 0.93 and 1.23 days for the infectious period γ^{-1}. A Dirichlet distribution was used for the initial proportions in compartments S, E, I, R, constraining the mean of the one in R to be 0.15, its variance 0.15^2, and the means of the other initial proportions to be equal.

The adaptive EK-MCMC algorithm was applied to the data and Fig. 2 depicts the incidence curve together with the posterior mean and pointwise 95% credible intervals. Estimates of β_t are also displayed indicating various changes over time. The changes in β_t are consistent with the argument that schools closure for holidays have been driving the epidemic: different values are observed during school and holidays periods, appearing to be synchronised with schools opening and closing. Posterior summaries for the static parameters, as well as a sensitivity analysis on the priors can be found in the Supplementary Materials. These suggest that inference is quite sensitive to the choice of prior for k and γ, but not for the remaining parameters. It would be interesting to repeat the procedure under an evidence synthesis framework and vague priors.

5.2 Application in real time. Was the first wave waning due to depletion of susceptibles?

In this section the methodology of the paper is applied in real time, i.e. considering partial datasets from June 2009 up to the 20th of July, the 7th of September and the 26th of October. Each time the algorithm is run from

scratch to provide samples from the joint posterior $\pi(x_{1:i}, \theta | y_{1:i})$. From a computational cost point of view this procedure can be improved further by utilising previous MCMC runs, for example under the SMC^2 framework (Chopin *and others*, 2011). We did not pursue this direction further, as the PMCMC algorithm runs quite fast (less than 2 hours on a standard PC). In order to reduce uncertainty, especially at early stages, the value of τ was set to 0.1 rather than being estimated as in Section 5.1. We otherwise use the same model as before. A model with integrated Brownian motion was also fit but BM was chosen in terms of DIC; see Supplementary Materials (Appendix C). The main results are shown in Fig. 3.

On August 1st, the first wave of the epidemic had waned, incidence rates were decreasing and schools had closed. There were two competing scenarios to explain the epidemic decline: (i) holidays had caused the waning of the epidemic by lowering the effective contact rate. Hence, a similar or stronger wave could occur when schools would reopen in September in colder climatic conditions. (ii) The epidemic had stopped independently of holidays because a critical proportion of the population had been infected, conferring a sufficient level of herd immunity to stop the epidemic. In this case, no second wave was to be expected in September. On August 1st there was great uncertainty around the value of c (Baguelin *and others*, 2010), which is crucial in distinguishing between the two scenarios. We therefore conducted the following exercise.

The PMCMC algorithm, run up to August 1st, provides samples from the posterior of the difference in β_t between July 13th (before the decrease in incidence) and August 1st. For $c = 10$, the 97.5% point of this posterior is -0.32, indicating a decrease in β_t. The latter supports scenario (i), as the competing scenario is associated with a zero-decrease in β_t. Nevertheless, as this value depends on c, the algorithm was run for different values of it ranging from 20 to 150. The results appear on Fig. 4. Note that the 97.5% point of interest increases as a function of c and reaches 0 for a correction factor close to 70. As this level seemed unrealistic (Baguelin *and others*, 2010), the experiment provides evidence in favour of scenario (i) highlighting the danger of a second wave in September, that actually occurred. Such evidence can be important for decision-makers, especially when considering implementations of preventive measures as vaccines.

5.3 A multiple age group diffusion driven SEIR model

The analysis of Section 5.1 can be used to construct more structured models. For example, the effect of holidays is evident and may differ from children to adults, thus casting doubts on the assumption of a homogeneous population. It seems more natural to consider a model with two age groups (c:children and a:adults) and target all possible effective contact rates among them. In our notation β^{ca} refers to the effective contact rate from children to adults

and S^c denotes the number of susceptible children. For reasons of parsimony we assign Brownian motions to $\log(\beta_t^{cc})$, $\log(\beta_t^{aa})$ and treat β^{ca}, β^{ac} as constant. We also set we set $\beta^{ca} = \beta^{ac} = b$, in line various multiple age groups epidemic models in different settings (e.g. Whitaker and Farrington (2004)). The dynamic part of the model is now given by

$$\begin{cases} \frac{dS_t^c}{dt} = -S_t^c \left(\beta_t^{cc} \frac{I_t^c}{N^c} + b \frac{I_t^a}{N^a}\right), \quad \frac{dE_t^c}{dt} = S_t^c \left(\beta_t^{cc} \frac{I_t^c}{N^c} + b \frac{I_t^a}{N^a}\right) - kE_t^c, \\ \frac{dS_t^a}{dt} = -S_t^a \left(\beta_t^{aa} \frac{I_t^a}{N^a} + b \frac{I_t^c}{N^c}\right), \quad \frac{dE_t^a}{dt} = S_t^a \left(\beta_t^{aa} \frac{I_t^a}{N^a} + b \frac{I_t^c}{N^c}\right) - kE_t^a, \\ \frac{dI_t^c}{dt} = kE_t^c - \gamma I_t^c, \quad \frac{dR_t^c}{dt} = \gamma I_t^c, \quad \frac{dI_t^a}{dt} = kE_t^a - \gamma I_t^a, \quad \frac{dR_t^a}{dt} = \gamma I_t^a. \end{cases} \tag{7}$$

The data from the A/H1N1(2009) pandemic provide incidence estimates for children and adults separately so they can be used to estimate the model of (7). If only final outcome data were available, not all effective contact rate parameters would be estimable. However, the temporal dataset provides extra information by the relative variation of susceptible and infective population in adults versus children. We applied the EK-MCMC scheme, which was essential in order to obtain reasonable MCMC performance. Fig. 5 depicts the results. Unlike earlier attempts with versions of a multi-group model with a single diffusion driving all contact rates, the fit appears to be good. The trajectory of children seems to be similar with that of Fig. 2 thus stressing their role to the evolution of the epidemic. More details, including posterior summaries for the parameters and information about the priors can be found in the Supplementary Materials (Appendix C).

6 Discussion

In this paper we examined epidemic models where some of the parameters are represented by diffusions or integrals thereof. The main motivation was to account for various time varying drivers (virus evolution, seasonality, schools closure, etc), while maintaining a simple interpretation. We present a unified framework that supports data augmentation MCMC schemes based on fine partitions on the diffusion path. The associated approximation error can be controlled by the user without affecting the MCMC performance and can be viewed as an extension of the approaches by Roberts and Stramer (2001); Chib *and others* (2006) to the more challenging observation regime of this paper. The consideration of the algorithms in a continuous time setting revealed major issues associated with Gibbs data-augmentation schemes. This justifies the use of particle MCMC, which updates paths and parameters jointly, while pointing directions for future research on Gibbs schemes. We also presented a computational machinery based on the PMCMC algorithm (Andrieu *and others*, 2010), that was integrated in an adaptive MCMC context. We consider EKF based adaptive algorithms that can offer substantial improvement, especially in cases with many static parameters. This paper

is one of the first applications of PMCMC in epidemic models and data; standard PMCMC schemes were also used in Rasmussen *and others* (2011).

Initially we relied on a simple SEIR model but such an analysis can be viewed as an exploratory tool towards more structured models; e.g. the age-structured model of Section 5.3 that appears to be an improved representation of reality. This approach can help in developing richer models and testing alternative scenarios for public health interventions, or to bring further insights on extrinsic factors such as climate on the dynamics of epidemics. Moreover, this framework can support multiple sources of data, of potentially different nature: Rasmussen *and others* (2011) has shown how time series and genealogies can be combined in a PMCMC inference framework for more informative estimates. While we worked mainly with influenza time series, the developed methodology can be applied to other cases; current work considers its application as part of the CHARME project (Boily *and others*, 2007). The presented approach may also be thought as an alternative to the white noise modeling of environmental stochasticity introduced in Bretó *and others* (2009), as it offers to the possibility to capture the dynamics of environmental drivers. A potential next step will be to combine environmental with demographic stochasticity, modelling infections as Poisson processes which rates depend on a time-varying β.

The inferential framework presented in this article shares the "plug and play" feature of the Iterated Filtering methodology. While extra care and further study is required for specific models or datasets, its algorithmic aspects can be decoupled from the modeling aspects. This provides the possibility to develop generic inference packages: we are currently working towards its integration in a generic inference platform inspired from the R package POMP.

7 Supplementary Materials

Supplementary material is available online at http://biostatistics.oxfordjournals.org. It contains implementation details for the PMCMC algorithm (Appendix A), a comparison with the EKF (Appendix B), additional information on Sections 4 and 5 (Appendix C). Also the sensitivity analysis (Appendix D) and a detailed exposition of the issues of Section 3.2 (Appendix E).

Acknowledgments

The authors would like to thank John Edmunds, Nikolaos Demiris, Wicher Bergsma and the three anonymous reviewers for their helpful and constructive comments.

References

Anderson, R.M. and May, R.M. (1992). *Infectious diseases of humans: dynamics and control*. Oxford University Press, USA.

Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods (with discussion). *J. Royal Statist. Soc. B* **72**(3), 269–342.

Baguelin, M., Hoek, A.J.V., Jit, M., Flasche, S. *and others*. (2010). Vaccination against pandemic influenza A/H1N1 in England: a real-time economic evaluation. *Vaccine* **28**, 2370–84.

Beskos, A., Papaspiliopoulos, O., Roberts, G.O. and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(3), 333–382.

Boily, MC., Lowndes, CM., Vickerman, P., Kumaranayake, L. *and others*. (2007). Evaluating large-scale HIV prevention interventions: study design for an integrated mathematical modelling approach. *Sex Transm Infect.* **83**(7), 582.

Bretó, C., He, D., Ionides, E. and King, A. (2009). Time series analysis via mechanistic models. *The Annals of Applied Statistics* **3**(1), 319–348.

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. *American journal of epidemiology* **164**(6), 591–597.

Cauchemez, S. and Ferguson, N.M. (2008). Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in london. *Journal of the Royal Society Interface* **5**(25), 885–897.

Cauchemez, S., Valleron., A.J., Boëlle, P.Y., Flahault, A. and Ferguson, N.M. (2008). Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* **452**(7188), 750–754.

Cazelles, B. and Chau, N.P. (1997). Using the Kalman Filter and Dynamic Models to Assess the Changing HIV/AIDS Epidemic. *Mathematical Biosciences* **140**(2), 131–154.

Cazelles, B., Chavez, M., McMichael, A.J. and Hales, S. (2005). Nonstationary Influence of Ei Nino on the Synchronous Dengue Epidemics in Thailand. *PLoS Medicine* **2**(4), 313.

Chib, S., Pitt, M.K. and Shephard, N. (2006). Likelihood based inference for diffusion driven state space models. Working paper.

Chopin, N., Jacob, P.E. and Papaspiliopoulos, O. (2011). SMC^2: A sequential Monte Carlo algorithm with particle Markov chain Monte Carlo updates. arXiv:1101.1528v3.

Cori, A., Boëlle, P.Y., Thomas, G., Leung, G.M. and Valleron, A.J. (2009). Temporal variability and social heterogeneity in disease transmission: the case of SARS in Hong Kong. *PLoS computational biology* **5**(8), e1000471.

Doucet, A. and Johansen, A.M. (2011). A tutorial on particle filtering and smoothing: fifteen years later.

Dukic, V.M., Lopes, H.F. and Polson, N. (2009). Tracking Flu Epidemics Using Google Flu Trends and Particle Learning. Working paper.

Ferguson, N. (2007). Capturing human behaviour. *Nature* **446**(7137), 733–733.

Fine, P.E.M. and Clarkson, J.A. (1982). Measles in England and Wales: an analysis of factors underlying seasonal patterns. *International journal of epidemiology* **11**(1), 5.

Geyer, C.J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science* **7**(4), 473–483.

Golightly, A. and Wilkinson, D. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis* **52**(3), 1674–93.

Griffin, J.T., Garske, T., Ghani, A.C. and Clarke, P.S. (2011). Joint estimation of the basic reproduction number and generation time parameters for infectious disease outbreaks. *Biostatistics* **12**(2), 303.

He, D.H., Dushoff, J., Day, T., Ma., J. and Earn, D.J.D. (2011). Mechanistic modelling of the three waves of the 1918 influenza pandemic. *Theoretical Ecology* **4**, 283–8.

Ionides, E.L., Bretó, C. and King, AA. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **103**(49), 18438.

Kalogeropoulos, K. (2007). Likelihood-based inference for a class of multivariate diffusions with unobserved paths. *Journal of Statistical Planning and Inference* **137**(10), 3092 – 3102.

Kurtz, T.G. (1981). *Approximation of population processes*. Society for Industrial Mathematics.

Miller, E., Hoschler, K., Hardelid, P., Stanford, E. *and others*. (2010). Incidence of 2009 pandemic influenza a H1N1 infection in England: a cross-sectional serological study. *Lancet* **375**(9720), 1100–1108.

Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*. Springer Verlag.

Pokern, Y., Stuart, A.M. and Wiberg, P. (2009). Parameter estimation for partially observed hypoelliptic diffusions. *Journal of the Royal Statistical Society: Series B* **71**(1), 49.

Rasmussen, D.A., Ratmann, O. and Koelle, K. (2011). Inference for Nonlinear epidemiological models using genealogies and time series. *PLoS Computational Biology* **7**(8), e1002136.

Roberts, G.O. and Rosenthal, J.S. (2009). Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics* **18**(2), 349–367.

Roberts, G.O. and Stramer, O. (2001). On Inference for Partially Observed Nonlinear Diffusion Models Using the Metropolis-Hastings Algorithm. *Biometrika* **88**(3), 603–621.

Särkkä, S. and Sottinen, T. (2008). Application of Girsanov Theorem to Particle Filtering of Discretely Observed Continuous-Time Non-Linear Systems. *Bayesian Analysis* **3**(3), 555–584.

Shaman, J. and Kohn, M. (2009). Absolute Humidity Modulates Influenza Survival, Transmission, and Seasonality. *Proceedings of the National Academy of Sciences* **106**(9), 3243.

Unkel, S., Farrington, C., Garthwaite, P.H., Robertson, C. and Andrews, N. (2012). Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A* **175**(1), 49–82.

Wahba, G. (1990). *Spline models for observational data*. Society for Industrial Mathematics.

Whitaker, H. J. and Farrington, C. P. (2004). Infections with Varying Contact Rates: Application to Varicella. *Biometrics* **60**(60), 615–623.

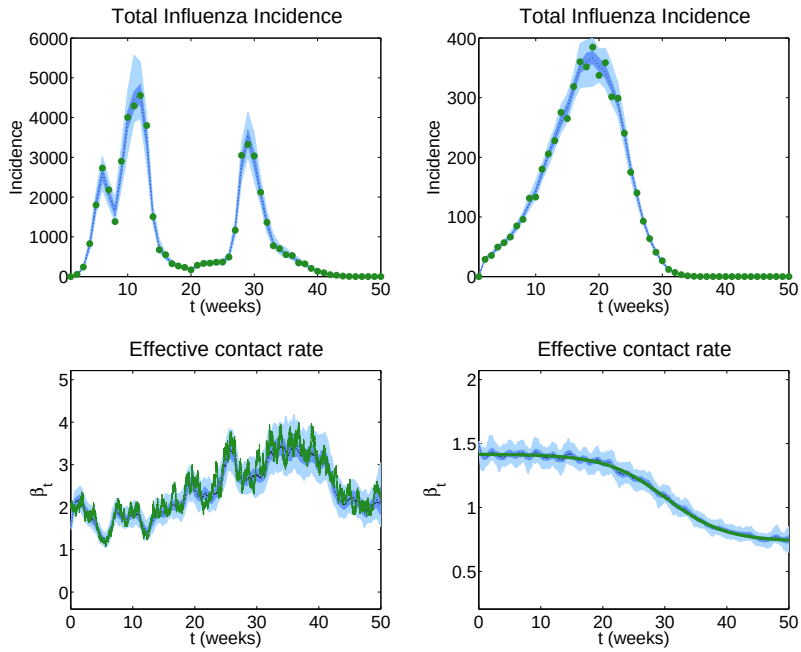

Figure 1: Illustration of how the underlying dynamic of the effective contact rate can be estimated from weekly recorded cases.
Green dots indicate simulated observed incidence (top panels). Green lines indicate simulated effective contact rate trajectories (bottom panels). Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue areas show credible intervals, respectively at 50% and 95% levels.
Top panels: simulated weekly numbers of cases observed with noise, and corresponding model-based offline reconstructions (left: experiment 1, right: experiment 2)
Bottom panels: simulated and estimated trajectory of the effective contact rate (left: experiment 1, right: experiment 2)

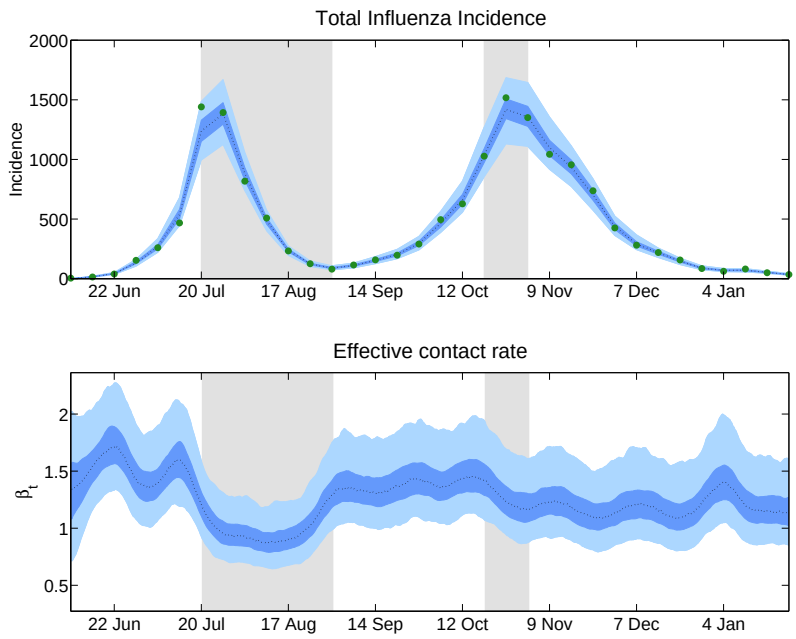

Figure 2: Weekly incidence data from the A/H1N1 2009 influenza pandemic and corresponding offline estimates of the effective contact rate.
Green dots indicate incidence estimates provided by the Health Protection Agency. Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue areas show credible intervals, respectively at 50% and 95% levels. Holidays are indicated by a light grey area.
Top: observations of the weekly total number of A/H1N1 influenza cases in London (per 100 000 inhabs.) and model-based offline reconstruction
Bottom: offline estimates of the effective contact rate.

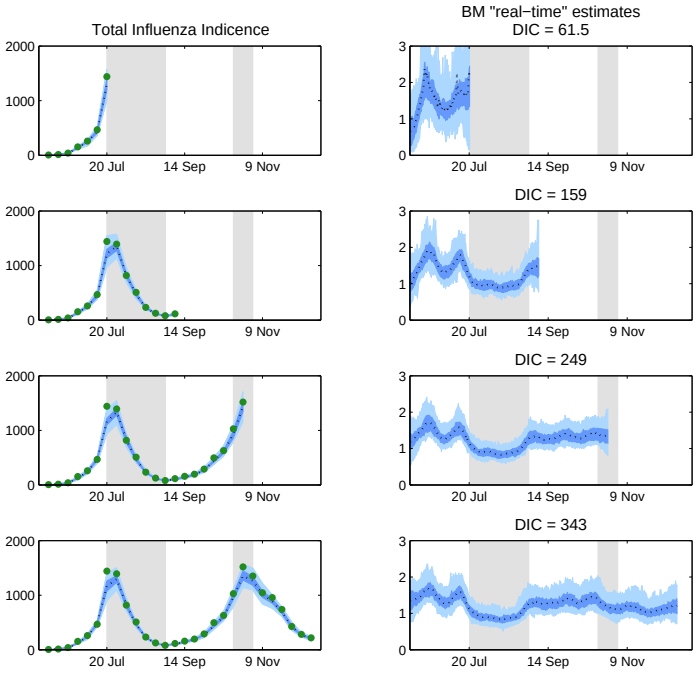

Figure 3: What could have been inferred by carefully following the epidemic in real time?
Green dots indicate observed incidence estimates provided by the Health Protection Agency (left panels). Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue areas respectively indicate 50% and 95% credible intervals of the posterior density. Holidays are indicated by a light grey area.
Left panels: HPA estimates of the weekly total number of A/H1N1 influenza cases in London (per 100 000 inhabs.)
Right panels: "real-time" estimates of the effective contact rate.

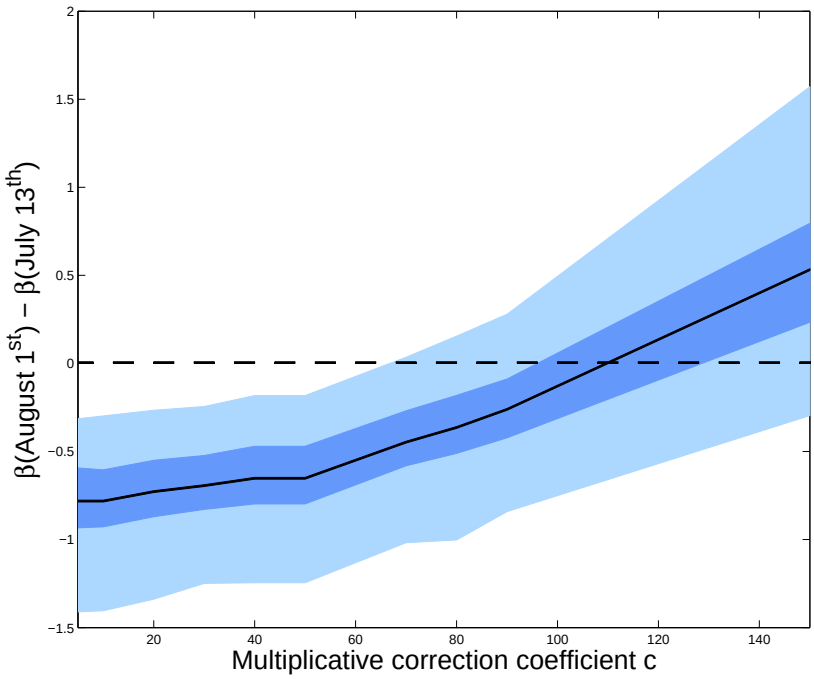

Figure 4: The implication of different scenarios for the real value of under-reporting on the decrease of the effective contact rate between July 13th and August 1st

For each value of c, the mean of the posterior density for $\beta(August\ 1^{st}) - \beta(July\ 13^{th})$ is plotted in black. Dark and light blue areas respectively indicate 50% and 95% credible intervals of the posterior density. The dotted line locates the scenario whith no change in the effective contact rate.

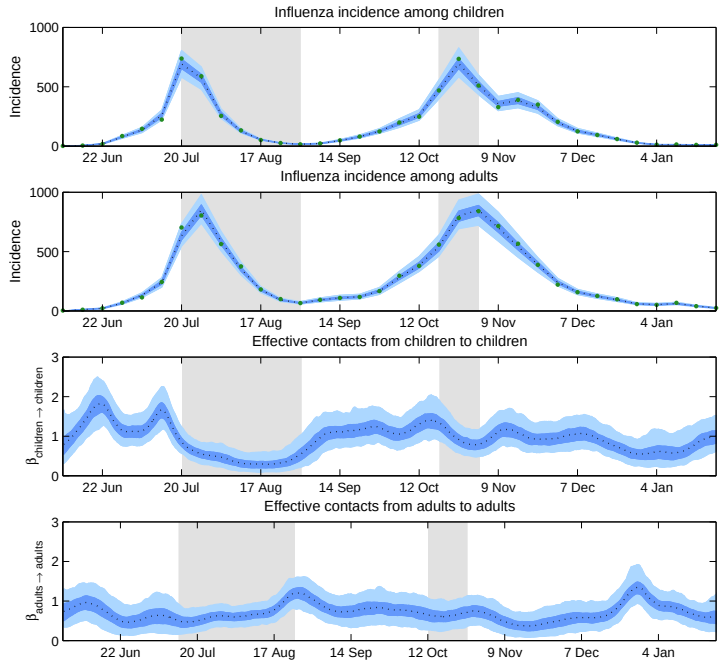

Figure 5: Offline estimates of the effective contact rate among children and adults during the A/H1N1 2009 influenza pandemic using a 2-classes age-structured model and age-specific incidence data.
Green dots indicate observed incidence estimates among each age group provided by the Health Protection Agency (first and second panels). Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue areas respectively indicate 50% and 95% credible intervals of the posterior density. Holidays are indicated by a light grey area.
First panel: HPA estimates of the weekly total number of A/H1N1 influenza cases among children in London (per 100 000 inhabs.)
Second panel: HPA estimates of the weekly total number of A/H1N1 influenza cases among adults in London (per 100 000 inhabs.)
Third panel: offline estimates of the effective contact rate from children to children.
Fourth panel: offline estimates of the effective contact rate from adults to adults.

arXiv:1203.5950v2 [stat.AP] 3 Nov 2012

Capturing the time-varying drivers of an epidemic using stochastic dynamical systems SUPPLEMENTARY MATERIALS

JOSEPH DUREAU*, KONSTANTINOS KALOGEROPOULOS

Department of Statistics , London School of Economics and Political Science, UK

j.dureau@lse.ac.uk

MARC BAGUELIN

Immunisation, Hepatitis and Blood Safety Department, Health Protection Agency, Centre for the Mathematical Modeling of Infectious Diseases, London School of Hygiene and Tropical Medicine

November 6, 2012

Here we provide supplementary information for various parts of the main paper. Appendix A illustrates the effect of various algorithmic parameters and suggestions on how they can be set in practice. Appendix B presents a simulation based comparison of the particle filter with the Extended Kalman Filter (EKF), whereas in Appendix C we provide more details on the analyses of Sections 4 and 5 of the main paper. Appendix D contains a sensitivity analysis for the priors assigned to the parameters of the model in Section 5.1. Finally, in appendix E, more details are given regarding the formulation of the models in continuous time and their potential implications on the associated data augmentation schemes.

Appendix A: details of the PMCMC implementation

In this Appendix, we provide more details for the practical implementation of the PMCMC algorithm presented in this article. We specify how to determine key parameters of the algorithm, i.e. the Euler discretisation time-step δ, the number of particles N_{parts} used in the Particle Smoother (PS), and how to set the Metropolis updates of the parameter vector θ.

[0]To whom correspondence should be addressed.

Determining the Euler discretization time-step

In general, solutions of the nonlinear ODEs encountered in epidemic models are not available in closed form. In order to evaluate $\pi(x_{0:n}, \theta|y_{1:n})$, the trajectory of the system needs to be discretised according to a given time-step δ to provide an approximate solution. The Euler approximation ensures that as δ tends to 0, $\hat{\pi}_\delta(x_{0:n}, \theta|y_{1:n})$ converges to $\pi(x_{0:n}, \theta|y_{1:n})$. In practice a sequence of decreasing δ values is chosen and quantities such as $E[\hat{p}_\delta(\sigma|y_{1:n})]$ or $E[\hat{p}_\delta(\tau|y_{1:n})]$, are monitored. For sufficiently small values of δ, a convergence is generally observed, as shown in Fig. 1 for two different datasets of weekly influenza data. In this case, a $\delta = 0.1$ day would be a reasonable choice. We note at this point that the computational cost is of $O(\delta^{-1})$

Determining an optimal number of particles

The PMCMC algorithm is theoretically valid regardless of the number of particles N_{parts} used in the particle smoother, as shown in Andrieu *and others* (2010). Nevertheless, the smaller the number of particles used in the particle smoother, the more noisy the estimate of the likelihood $\hat{p}(y_{1:n}|\theta)$ becomes. This noise has a negative impact on the acceptance rate of the MCMC algorithm run in the θ space. Consequently, N_{parts} has to be big enough so that it won't affect the acceptance rate, keeping in mind that the cost is of $O(N_{parts})$. Hence, a compromise needs to be achieved. Fig. 2 shows how the acceptance rate increases as the number of particles gets higher. Note that the acceptance rate has a plateau form, indicating that it is perhaps not worthwhile to increase N_{parts} beyond some point. We repeat the experiment for two different values of the measurement error parameter (τ). This figure shows that when the observational noise decreases, more particles are needed, which is explained by the fact that the particles need to be 'closer' to the observations.

Appendix B: Assessing the validity and limitations of the Extended Kalman approximation

Here, we compare the particle filter with the EKF approach to assess the the extent of the gain of avoiding the approximations of the latter. A set of 100, 7-month long, time-series of weekly influenza cases were drawn from the BM model. In order to ensure realistic epidemic datasets, we 'reverse-engineered' randomly selected influenza time-series $(y_{1:n}^{Goog,j})_{j=1}^{100}$ from the freely available Google FluTrend data (Ginsberg *and others*, 2008). For each of the datasets, we obtained estimates of $(\beta_{0:n}^{Sim,j})_{j=1}^{100}$ and the corresponding parameters $(\theta_j)_{j=1}^{100}$. These quantities were then used to generate influenza time-series $(y_{1:n}^{Sim,j})_{j=1}^{100}$. The static parameters of the model (initial condi-

tions, k, γ, σ and τ) were assumed known, to isolate the problem of estimating β_t from accounting for parameter uncertainty and perform more relevant comparisons. We compare the following two estimators of $\beta_i^{Sim,j}$: $\hat{\beta}_i^{Filt,j} = \hat{E}(\beta_i^j | y_{1:i}^j, \theta_j^*)$ obtained from the filtering distribution where $\hat{E}(.)$ denotes Monte Carlo estimates of the relevant expectations, and the the EKF estimator $\hat{\beta}_i^{EKF,j} = \tilde{E}^{EKF}(\beta_i^j | y_{1:i}^j, \theta_j^*)$ where $\tilde{E}^{EKF}(.)$ denotes expectation under EKF. The performance of the estimators is measured through their bias and Mean Squared Error (MSE). The results indicate a better performance for $\hat{\beta}_i^{Filt,j}$. The bias of the estimates provided by the EKF is 0.0285 while use of $\hat{\beta}_i^{Filt,j}$ reduces the bias by about 78% (0.0063). The corresponding relative reduction in MSE is smaller (10%, 0.0270 to 0.0242), indicating a bias-variance tradeoff. Use of the smoothing distribution estimator $\hat{\beta}_i^{Sm,j} = \hat{E}(\beta_i^j | y_{1:n}^{Sim,j}, \theta_j^*)$ is associated with a further 87% (0.0032) reduction in the MSE, while keeping the bias at the same low levels. The estimators $\hat{\beta}_i^{Filt,j}$ and $\hat{\beta}_i^{Sm,j}$ are associated with a tolerable computational cost of 2 hours on a standard PC. In conclusion, the bias introduced by the Extended Kalman approximation is non-negligible with regards to the level of accuracy that can be obtained with exact particle methods on this type of datasets. Nevertheless, this study has shown the approximation to be robust and motivates the use of the approximated model as a proxy for the exact one, for example to initialize the particle MCMC algorithm in the way presented in this article.

Appendix C: details on the simulations and results of sections 4 and 5

Assessing the validity of the particle MCMC through simulations

This section deals with a series of experiments that were conducted in order to assess the validity of the particle MCMC, and to illustrate on different examples how the trajectory $\beta_{1:n}$ could be captured from noisy weekly cases observations. Experiment 1 is based on a trajectory of $\beta_{1:n}$ simulated from a random walk model with volatility $\sigma^2 = 0.07^2$. Two different corresponding epidemic datasets have been generated, for given and equal initial conditions and biological parameters, respectively with observational noise $\tau = 0.1$ (experiment 1.a) and $\tau = 0.05$ (experiment 1.b). Similarly, two epidemic datasets resulting from an effective contact rate following a significantly decreasing sigmoid were generated with respectively $\tau = 0.1$ (experiment 2.a) and $\tau = 0.05$ (experiment 2.b).

For each of these datasets, our proposed methodology was run to estimate σ, τ and $\beta_{1:n}$. cFig. 1 of the main text contains estimates of the incidence time series as well as $\beta_{1:n}$ for the experiments with $\tau = 0.1$ (Exps 1.a and

2.a). Corresponding figures for Experiments 1.b and 2.b are shown in Fig. 3 of this document. Moreover, Table 2 presents the mean, median and 95% credible intervals for the estimates of σ and τ in each of the experiments. In experiment 2 the estimates seem to be in good agreement with the true values, as the latter are contained in the 95% credible intervals. The aim was to assess the robustness of the proposed methodology to model misspecification, fitting a Brownian motion to a smooth sigmoid curve. The algorithm performs reasonably well, succeeding in capturing the trajectories of β_t and all the parameters except for τ which is slightly underestimated. A potential explanation for this is that part of the variability is absorbed from the volatility parameter of the Brownian motion.

A/H1N1 pandemic

We provide here with the corresponding trace plots (Fig. 4) for the parameter estimates of Section 5.2. The trace plots indicate good mixing of the PMCMC algorithm for every parameter. This was achieved by following the procedure presented in Appendix A.

Age-structured model: children and adults

Additional information on the analysis of section 5.3. can be found in Table 3.

Illustration of alternative approaches on the real time example

In Fig. 5, we repeat the real-time analysis, conducted in the Section 5.2 of the main paper, under two alternative approaches. First, we consider a model with an integrated Brownian motion (iBM) on x_t, implying smoother β_t trajectories as opposed to the non-differentiable paths induced by the Brownian motion (BM) formulation. The choice between those models is not trivial and could depend on the context of the epidemic. We decided to adopt the model with Brownian motion (BM) on the basis of the Deviance Information Criterion (DIC) of Spiegelhalter *and others* (2002); as we can see from Fig. 3 of the main document and Fig. 5 of the SM, the BM model is consistently better in that respect. Nevertheless, there is uncertainty on the performance of the DIC criterion in the setting of this. It would be quite interesting to explore this further in the future and compare with alternative model choice criteria.

The second approach adopts and applies the methodology of maximum likelihood via iterated filtering (MIF), introduced in (Ionides *and others*, 2006). Initially, estimates θ are obtained by maximising $p(y_{i:n}|\theta)$ subject to some constraints set by the priors. Second, a particle filter was run with θ fixed to its estimated value. As expected, given that we are not accounting for parameter uncertainty, the resulting pointwise 95% credible intervals are

narrower; roughly 50% on the 6-month dataset and even more at early stages with less information on θ.

Appendix D: Sensitivity analysis

We explore in this section the robustness of the obtained estimates. We concentrate on the example of Section 5.1 in the main paper, where the observational noise was estimated and informative priors were used for the initial proportion of immune individuals in the population ($R(0)$), and the lengths of the latent (k^{-1}) and infectious periods (γ^{-1}). For each of these quantities, the mean of the prior densities have been tilted by respectively -20%, -10%, +10% and +20%. In all cases but for the tilted parameters themselves, the median of the corresponding posterior densities lie both in the 95% and 50% credible intervals of the untitled case. The resulting medians estimates are shown in table 4, along with the originally obtained summary statistics. Furthermore, for all the runs with tilted priors, the resulting $\beta_{1:n}$ trajectories remain completely in the original 50% credible intervals.

Appendix E: details of Gibbs data augmentation scheme

In this section we provide more details on the Gibbs schemes discussed in section 3.3 of the main text. Stochastic epidemic models presented in this paper can be written as

$$\begin{cases} dx_t & = \mu_x(x_t, \theta_x)dt + \sigma_x(x_t, \theta_x)dB_t, \quad 0 < t < t_n \\ y_{1:n}|V_{0:n}, \theta_y & \sim \mathbb{P}_y(y_{1:n}|V_{0:n}, \theta_y), \quad V_{0:n} = g(x_{0:n}, \theta_v) \end{cases} \tag{1}$$

where V_t represents the ODE states vector observed trough partial and noisy data $y_{1:n}$. The rest of the model is defined in section 3.1. Since it contains intractable densities we work with the time discretised versions $x^{dis}_{0:n}$ and $V^{dis}_{0:n}$ and proceed using the Euler approximating scheme. A Gibbs algorithm alternates between updating the trajectories of $x^{dis}_{0:n}$ (and consequently $V^{dis}_{0:n}$) given θ, and vice versa. Nevertheless, as the Euler time step δ goes to 0, the quadratic variation process of x_t uniquely determines the value of θ_x in $\sigma_x(.)$ and the algorithm degenerates (Roberts and Stramer, 2001). In practice this translates into a mixing time of $O(m)$.

In the context of diffusion driven epidemic models this problem was dealt with suitable reparametrisations such as the ones in Chib *and others* (2006) or Kalogeropoulos (2007). The latter uses the Lamperti transform, i.e. $x_t \to H(x_t, \theta_x) = \eta(x_t, \theta_x) - \eta(x_0, \theta_x) =: u_t$ where $\eta(\cdot; \theta_x)$ is an antiderivative of $\sigma_x^{-1}(\cdot; \theta_x)$. Assuming that $\sigma_x(\cdot; \theta_x)$ is continuously differentiable, an application of Ito's lemma provides the SDE of the transformed diffusion u_t as:

$$du_t = \nu(u_t; \theta_x)dt + dB_t \ , \quad u_0 = 0 \ , \tag{2}$$

where

$$\nu(u_t;\theta_x) = \frac{\mu_x\big(H^{-1}(u_t,\theta_x),\theta_x\big)}{\sigma_x\big(H^{-1}(u_t,\theta_x),\theta_x\big)} - \frac{1}{2}\sigma_x'\big(H^{-1}(u_t,\theta_x),\theta_x\big)\ .$$

Let $\mathbb{P}^u$ denote the distribution of u_t. Girsanov formula provide its density with respect to that of a standard Brownian motion, denoted by $\mathbb{W}$,

$$\frac{d\mathbb{P}^u}{d\mathbb{W}} = \exp\left\{\int_0^{t_n} \nu(u_s;\theta_x)du_s - \frac{1}{2}\int_0^{t_n} \nu(u_s;\theta_x)^2 ds\right\}, \tag{3}$$

and the state vector can be written as

$$V_{0:n} = h_u(u_{0:n}, x_0, \theta_x, \theta_v). \tag{4}$$

The model can be defined from (3), (6) and (1). It contains intractable quantities but can be accurately approximated given the time discretisation of the diffusion path. An alternative reparametrisation, defined in discrete time, was suggested in Chib *and others* (2006). It uses the transformation below

$$w_t = \frac{x_t - \big(x_{t-\delta} - \delta\mu_x(x_t,\theta_x)\big)}{\sigma_x(x_{t-\delta},\theta_x)}, \forall t. \tag{5}$$

In our setting the driving Brownian motion of $x_{0:n}^{dis}$, denoted by $w_{0:n}^{dis}$ and provided by (5), can be used to provide a discrete skeleton of the state vector

$$V_{0:n}^{dis} = h_w(w_{0:n}^{dis}, x_0, \theta_x, \theta_v). \tag{6}$$

The model is now given by $w_{0:n}^{dis}$, that can be transformed to x_t for which the Euler-Maruyama approximation can be used, and 1 which can be approximated using the discretised state vector $V_{0:n}^{dis}$.

Data augmentations schemes can be used for the models above. Gibbs versions of such schemes will alternate between updating $u_{0:n}^{dis}$ (or $w_{0:n}^{dis}$) and consequently $V_{0:n}^{dis}$ given θ, and θ conditional on either $u_{0:n}^{dis}$ (or $w_{0:n}^{dis}$). The first step can be done either by the overlapping block strategies in Chib *and others* (2006) and Kalogeropoulos (2007) or with a particle filter in the context of a particle Gibbs algorithm. The second step of updating θ_x given $u_{0:n}^{dis}$ or $w_{0:n}^{dis}$ is usually implemented through a random walk Metropolis algorithm:

- Let θ_x^c and $V_{0:n}^{disc}$ be the current values of θ_x and $V_{0:n}^{dis}$ respectively. Propose θ_x^* from $q(\theta_x^*|\theta_x^c)$.
- Compute $V_{0:n}^{dis*} = h_u(u_{0:n}, x_0, \theta_x^*, \theta_v)$
- Accept with probability

$$1 \wedge \frac{\pi(\theta_x^*, V_{0:n}^{dis*}|y_{1:n}, u_{0:n}^{dis*}\theta_v, \theta_y)q(\theta_x^c|\theta_x^*)}{\pi(\theta_x^c, V_{0:n}^{disc}|y_{1:n}, u_{0:n}^{disc}\theta_v, \theta_y)q(\theta_x^*|\theta_x^c)}$$

For the Chib *and others* (2006) formulation, $u_{0:n}^{dis}$ can simply be replaced with $w_{0:n}^{dis}$ in the algorithm above.

Unfortunately both of the above algorithms may perform poorly. The problem is that every proposed value of θ_x implies a proposed trajectory of the ODE states vector V_t. As parts or functionals of this trajectory are observed with error, the proposed value of θ_x will not be accepted unless its associated V_t trajectory is close to these observations. Consequently, only small steps can be made on the θ_x space and the algorithm mixes very slowly. The problem intensifies as the noise variance becomes smaller and as the time horizon of the epidemic increases. Implementations of such algorithms in the simulated and real data of this paper are in line with this argument. Figure 6 displays the posterior draws for σ in the dataset of the simulation experiment 1 of Section 4 in the main paper. The posterior draws of σ were obtained from a particle Gibbs algorithm combined with the algorithm above; note that in this model $u_{0:n}^{dis}$ and $w_{0:n}^{dis}$ are equal. In order to isolate the problem, the algorithm was run on σ and β_t only, and all the other parameters were held fixed at the values they where simulated from (a value of $\tau = 0.1$ was used). The 'true value' of σ was 0.07 and we used a $\delta = 0.1$. As clearly shown in the traceplot the mixing of the chain is quite poor, thus casting doubts on the reliability of its output. The difference in mixing quality with the corresponding traceplot of Fig. 4 (bottom middle plot), corresponding to the PMCMC algorithm, is substantial.

To sum up, both formulations of Gibbs data augmentation schemes (with or without reparametrisation) are very likely to lead to inaccurate and inefficient MCMC algorithms. The use of particle MMH algorithms, termed as PMCMC in this paper, is therefore essential and the main paper focused on its implementation on diffusion driven epidemic models. PMCMC seems to provide a solution to the problem, but future research on Gibbs schemes with alternative reparametrisations addressing this problem, would be very helpful.

References

Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods (with discussion). *J. Royal Statist. Soc. B* **72**(3), 269–342.

Chib, S., Pitt, M.K. and Shephard, N. (2006). Likelihood based inference for diffusion driven state space models. Working paper.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L. *and others*. (2008). Detecting influenza epidemics using search engine query data. *Nature* **457**(7232), 1012–1014.

Ionides, E.L., Bretó, C. and King, AA. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **103**(49), 18438.

Kalogeropoulos, K. (2007). Likelihood-based inference for a class of multivariate diffusions with unobserved paths. *Journal of Statistical Planning and Inference* **137**(10), 3092 – 3102.

Roberts, G.O. and Stramer, O. (2001). On Inference for Partially Observed Nonlinear Diffusion Models Using the Metropolis-Hastings Algorithm. *Biometrika* **88**(3), 603–621.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4), 583–639.

	MSE	**Bias**
EKF	0.0269	0.0285
Particle filter	0.0242	0.0064
Improvement with regards to EKF	-10%	-77%
Particle smoother	0.0032	0.0027
Improvement with regards to P. filter	-87%	-64%

Table 1: Mean Squarred Error and Bias of β_t estimates provided by the EKF, particle filter and particle smoother

		Exp 1.a	**Exp 1.b**	**Exp 1.b**	**Exp 2.b**
τ	Simulation value	**0.1**	**0.05**	**0.1**	**0.05**
	Posterior mean	0.103	0.083	0.078	0.050
	Posterior median	0.103	0.084	0.077	0.050
	Posterior 95% c.i.	$[0.051; 0.152]$	$[0.027; 0.137]$	$[0.063; 0.96]$	$[0.042; 0.060]$
σ	Simulation value	**0.07**	**0.07**	**n.d.**	**n.d.**
	Posterior mean	0.066	0.083	0.016	0.014
	Posterior median	0.064	0.084	0.015	0.014
	Posterior 95% c.i.	$[0.048; 0.090]$	$[0.046; 0.089]$	$[0.010; 0.027]$	$[0.001; 0.021]$

Table 2: Mean, median and 95% confidence intervals for τ and σ estimates in four experiments

	Posterior mean	Posterior median	Posterior 95% c.i.
k^{-1}	1.56	1.55	$[1.53; 1.60]$
γ^{-1}	1.00	1.00	$[0.92; 1.08]$
$\beta_{cc}(0)$	1.44	1.36	$[0.89, 2.30]$
$\beta_{aa}(0)$	1.40	1.39	$[1.25; 1.64]$
β_{ca}	0.30	0.31	$[0.16; 0.50]$
β_{ac}	0.32	0.32	$[0.18; 0.48]$
$E_c(0)$	2.1×10^{-5}	1.9×10^{-5}	$[1.3 \times 10^{-5}; 3.8 \times 10^{-5}]$
$I_c(0)$	1.2×10^{-5}	1.4×10^{-5}	$[0.6 \times 10^{-5}; 2.9 \times 10^{-5}]$
$R_c(0)$	0.13	0.12	$[0.06; 0.26]$
$E_a(0)$	2.1×10^{-5}	2.0×10^{-5}	$[1.1 \times 10^{-5}; 3.5 \times 10^{-5}]$
$I_a(0)$	1.0×10^{-5}	1.0×10^{-5}	$[0.3 \times 10^{-5}; 1.5 \times 10^{-5}]$
$R_a(0)$	0.16×10^{-5}	0.16	$[0.09; 0.28]$
σ_c	0.11	0.10	$[0.08; 0.15]$
σ_a	0.08	0.08	$[0.05; 0.11]$

Table 3: Mean, median and 95% confidence intervals for the parameters of the structured model applied to the A/H1N1 pandemic data

	τ	k^{-1}	γ^{-1}	β_0	$E(0)$	$I(0)$	$R(0)$	σ
2.5% quantile	0.04	1.55	0.93	0.80	5.2×10^{-6}	1.6×10^{-6}	0.02	0.04
25% quantile	0.09	1.57	1.03	1.16	1.6×10^{-5}	7.0×10^{-6}	0.12	0.05
Median estimate	0.11	1.59	1.08	1.35	2.3×10^{-5}	1.6×10^{-5}	0.17	0.06
75% quantile	0.13	1.60	1.13	1.56	3.1×10^{-5}	2.8×10^{-5}	0.22	0.07
97.5% quantile	0.17	1.63	1.23	2.13	5.2×10^{-5}	6.5×10^{-5}	0.33	0.10
Median when R(0) shifted +10%	0.11	1.59	1.09	1.41	2.0×10^{-5}	1.9×10^{-5}	0.19	0.06
Median when R(0) shifted +20%	0.11	1.59	1.09	1.44	1.8×10^{-5}	2.1×10^{-5}	0.24	0.06
Median when R(0) shifted -10%	0.11	1.59	1.08	1.31	2.2×10^{-5}	2.2×10^{-5}	0.15	0.07
Median when R(0) shifted -20%	0.12	1.59	1.09	1.27	1.9×10^{-5}	2.1×10^{-5}	0.13	0.07
Median when k^{-1} shifted +10%	0.11	1.59	1.09	1.33	1.8×10^{-5}	2.2×10^{-5}	0.15	0.06
Median when k^{-1} shifted +20%	0.12	1.60	1.08	1.34	1.8×10^{-5}	2.3×10^{-5}	0.17	0.06
Median when k^{-1} shifted -10%	0.12	1.58	1.08	1.28	2.0×10^{-5}	2.0×10^{-5}	0.16	0.06
Median when k^{-1} shifted -20%	0.12	1.57	1.09	1.31	1.9×10^{-5}	2.0×10^{-5}	0.16	0.06
Median when γ^{-1} shifted +10%	0.11	1.59	1.12	1.23	2.2×10^{-5}	2.1×10^{-5}	0.15	0.07
Median when γ^{-1} shifted +20%	0.10	1.59	1.14	1.18	2.0×10^{-5}	2.4×10^{-5}	0.15	0.07
Median when γ^{-1} shifted -10%	0.11	1.59	1.06	1.37	1.9×10^{-5}	2.0×10^{-5}	0.16	0.07
Median when γ^{-1} shifted -20%	0.10	1.59	1.02	1.46	2.0×10^{-5}	1.8×10^{-5}	0.17	0.07

Table 4: Original estimates compared to the ones resulting from respectively tilting the priors on $R(0)$, γ^{-1} or k^{-1} by +10, +20, -10 or -20%

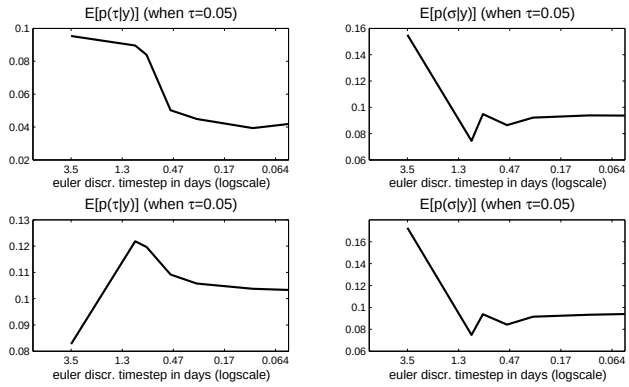

Figure 1: Convergence of the posterior density as the Euler discretization time-step δ decreases (x-axis in the log-scale)

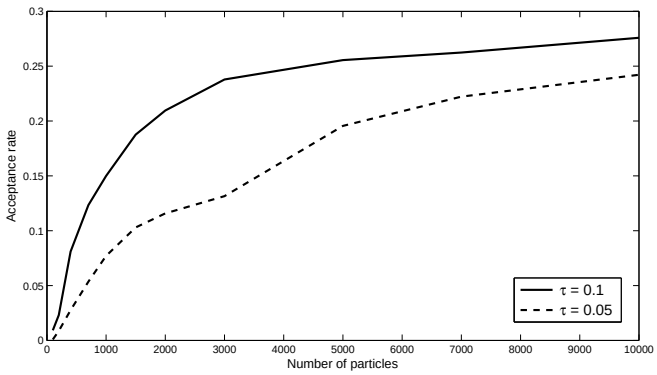

Figure 2: Acceptance rate as a function of N_{parts}, in two situations where the noise amplitude is respectively 10% (full line) and 5% (dotted line).

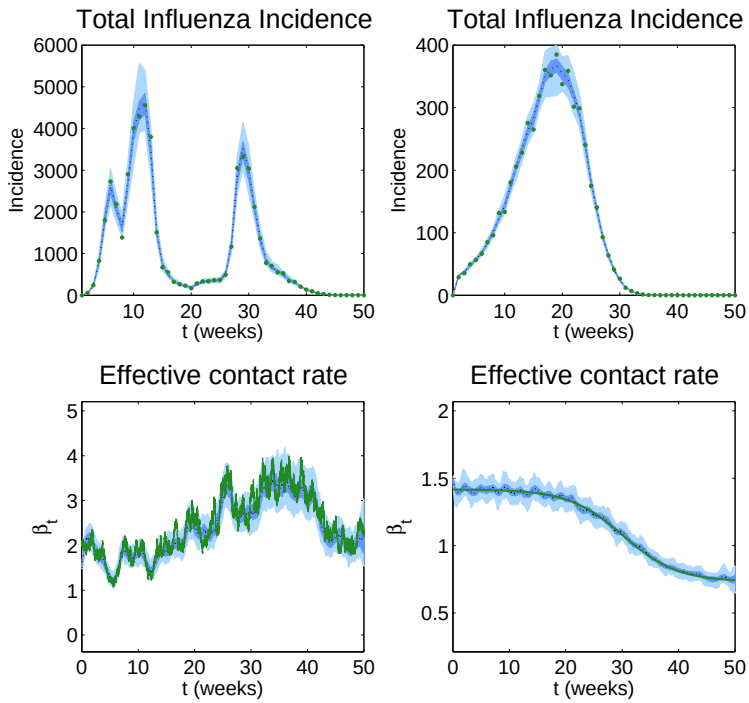

Figure 3: Illustration of how the underlying dynamic of the effective contact rate can be estimated from weekly recorded cases ($\tau = 0.05$)
Green dots indicate simulated observed incidence (top panels). Green lines indicate simulated effective contact rate trajectories (bottom panels). Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue show credible intervals, respectively at 50% and 95% levels.
Top left: experiment 1.b, weekly number of cases observed with noise
Top right: experiment 2.b, weekly number of cases observed with noise
Bottom left: experiment 1.b, simulated and estimated trajectory of the effective contact rate
Bottom right: experiment 2.b, simulated and estimated trajectory of the effective contact rate

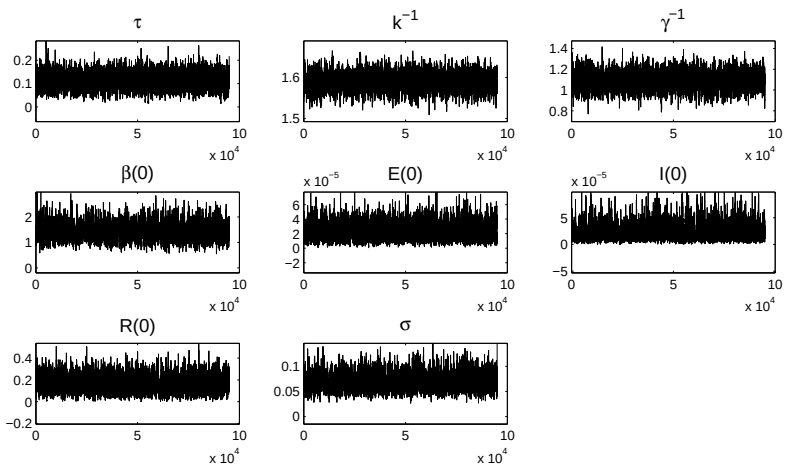

Figure 4: MCMC traceplots for each component of θ

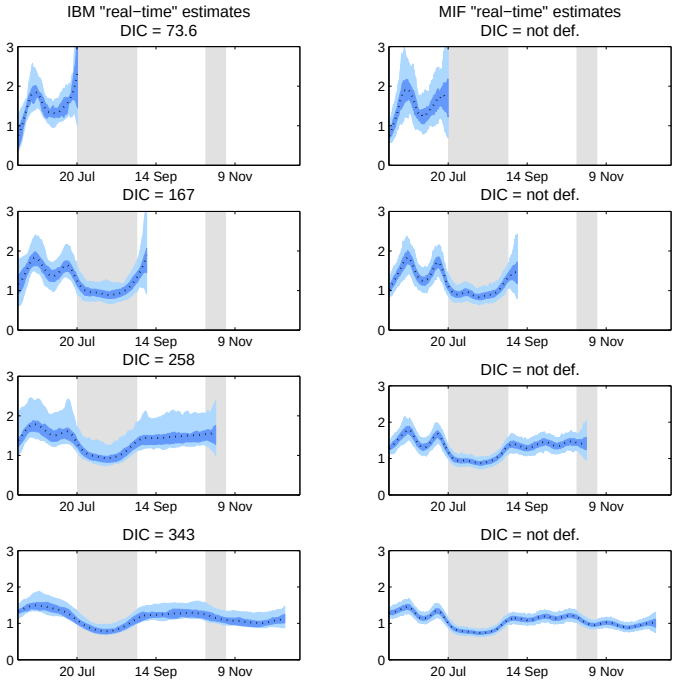

Figure 5: Modeling choices and implications, aiming for robustness
Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue show credible intervals, respectively at 50% and 95% levels.
Left panels: estimates from an alternative modeling approach: exploring the full posterior density of an IBM diffusion model (left)
Right panels: estimates from an alternative methodological approach: exploring the posterior density of a BM diffusion model conditionned on a likelihood maximizing parameter θ^* provided by the MIF algorithm (right)

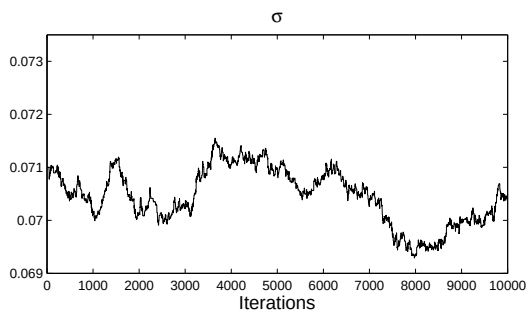

Figure 6: MCMC traceplot for σ when using a Particle Gibbs scheme with reparametrization