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, London, UK
Centre for the Mathematical Modeling
of Infectious Diseases,
London School of Hygiene and Tropical Medicine, London, UK

February 1, 2012

Abstract

Epidemics are often modelled using state-space models based on dynamical systems, observed through partial and noisy data. In this paper we develop stochastic extensions to the popular SEIR model with parameters evolving in time, in order to capture unknown influences of changing behaviors, public interventions, seasonal effects etc. Our models assign diffusion processes for the time-varying parameters, and our inferential procedure is based on the particle MCMC algorithm, suitably adjusted to accommodate the features of this challenging non-linear stochastic model. The performance of the proposed computational methods is validated on simulated data and the adopted model is applied to the 2009 A/H1N1 pandemic in England. In addition to estimating the trajectories of the effective contact rate, the methodology is applied in real time to provide evidence in related public health decisions.

Keywords: Infectious disease modelling; Non-parametric methods; Bayesian analysis; Statistical methods in Epidemiology, Particle MCMC

[0]To whom correspondence should be addressed.

1 Introduction

Epidemic models are often used to simulate disease transmission dynamics, detect emerging outbreaks (Unkel *and others*, 2010), and sometimes also to aid in designing and monitoring public interventions (Boily *and others*, 2007). Naturally, in order to capture the dynamics of epidemics, the main focus is generally made on their intrinsically dynamic elements such as the depletion of susceptibles as a disease spreads or the loss of immunity of the population as a virus evolves. Nevertheless, there are time-varying extrinsic factors that also play a key role on the evolution of epidemics, mainly through the effective contact rate, the number of contacts individuals make that could lead to an infection. Indeed, this quantity partly determines the mean number of individuals infected by a person carrying the disease, i.e. the reproduction rate R_t, and can foster or stop an epidemic by bringing R_t over or below 1. Consequently, the majority of public interventions, meant to prevent or mitigate epidemics aim at lowering the effective contact rate, either by protecting the contacts specifically made with infected individuals, or by promoting prevention measures. Examples of the latter can span from washing hands to wearing condoms or sleeping under treated bed-nets. The overall efficiency of a policy will depend on the efficiency of the prevention measures, but also on the exposition (proportion of the population reached by the campaign) and adherence (proportion that effectively change their habits) of the population to the sensibilisation campaign. Timely and accurate estimates of these elements are crucial in order to monitor the efficiency of an intervention and the potential need to intensify it. Nevertheless, the time-varying nature of epidemics poses a challenging statistical problem highlighting the need for suitable computational tools (Ferguson, 2007). In addition to awareness fluctuations, climatic variations also drive the epidemic dynamics through the effective contact rate. This has been clearly illustrated for diseases as cholera, malaria (Cazelles *and others*, 2005; Ionides *and others*, 2006) or influenza (Shaman and Kohn, 2009) for example. These studies were conducted either by relating climatic and incidence time-series (Cazelles *and others*, 2005), or by experimentally assessing the resistance of viruses in different climatic conditions (Shaman and Kohn, 2009). However, approaches of the first type do not disentangle the effect of time varying effective contact rates from the evolving population immunity, thus affecting generalisability to different immunological contexts. Furthermore, experimental results cannot be easily extended to a population level as sociological and behavioral patterns can play a role.

This paper considers a flexible modelling framework that encompasses time-varying aspects of the epidemic via stochastic differential equations and in particular hypoelliptic diffusions. We aim at providing robust inferential procedures, from observations on prevalence or incidence data, 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 suitably tailored Markov Chain Monte Carlo (MCMC) algorithms to the diffusion pathspace. The existing MCMC toolbox for parameters inference on epidemic models is expanded, utilising recent developments such as particle MCMC (PMCMC) algorithms (Andrieu *and others*, 2010). Our proposed methodology is contrasted with existing alternatives as the family of Kalman filters based on classic Gaussian approximations and Taylor expansions (specifically the extended Kalman Filter, EKF, used in Cazelles and Chau (1997)), the Maximum likelihood via Iterated Filtering algorithm (MIF, introduced in Ionides *and others* (2006)), or strong modelling assumptions as crude time discretisations or absence of measurement error (Cori *and others*, 2009). Modelling aspects are presented in Section 2, while the computational framework, and in particular PMCMC methodology, is presented in Section 3. Section 4 contains simulation experiments to illustrate and assess the performance of the proposed algorithms, including comparisons with alternative methods as the EKF. Simulation experiments on more features of PMCMC are presented in the appendices A and B of the Supplementary Materials. In Section 5 we apply our methodology to data from the 2009 H1N1 pandemic provided by the Health Protection Agency. In addition to illustrating the PMCMC algorithm on real data, we discuss how the effective contact rate could have been monitored in real time and whether a second wave after the summer holidays was the most likely scenario. Finally, Section 6 concludes with some relevant discussion.

2 Epidemic models with time-varying coefficients

2.1 Common grounds of epidemic models: key notions and data

Epidemic models are based on the distinction between infected and susceptible individuals, i.e. individuals that could potentially get infected. Depending on the disease and the scope of the study, further details become of interest. For instance, not all infected individuals have to be infective as it can take some time for the infected individual before shedding viruses and becoming infectious; this time interval, referred to as latent period, could be non-negligible in many cases such as influenza. Furthermore, some individuals may be neither infected nor susceptible, e.g. again for influenza a period of immunity is observed after an individual has stopped being ill and infectious. People in this situation are referred to as resistant or removed. Hence, a frequently used model is the "SEIR" accounting for susceptible (S), infected but not infective yet (E), infective (I), and removed (R) individuals. The model is set in 1; 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. Moreover, individuals stay in compartments E and I for an average period of k^{-1} and γ^{-1} respectively.

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

A key quantity here is the basic reproduction number, representing a theoretical quantity measuring the number of secondary infections that would follow the introduction of a primary infected individual in a fully susceptible population. This quantity, denoted with R_0, is often taken as a measure of the intrinsic infectiousness of a virus. A related and somewhat more practical quantity is the effective reproduction number, denoted with R_t, that refers to the number of secondary cases caused by an infected individual at a time t of an epidemic. R_t is a context-dependent quantity of high interest to policy designers as its position relative to 1 indicates the possibility for the epidemic to grow ($R_t > 1$) or to decrease in size ($R_t < 1$).

As epidemics are generally only partially observed, most of the available epidemiological data are sparse. For exceptional diseases as SARS, with low prevalence and strong severity and transmissibility, almost all cases can be considered to be carefully monitored. However, for the majority of diseases, monitoring of cases is done through health care attendance which often includes patients consulting for a variety of pathogens while presenting similar symptoms, or alternative syndromic data as telephone or Internet helplines, pharmacy sales, etc (Unkel *and others*, 2010). In the case of influenza, for example, influenza-like-illness cases are reported based on the observation of symptoms such as fever and some additional respiratory symptoms common to several pathogens (influenza A, influenza B, RSV, parainfluenza, adenovirus, rhinovirus, Streptococcus pneumoniae, etc). Additionally, for the most common pathogens, a large part of the infections can lead to no or mild symptoms for which infected people do not consult. Studies have shown that infections by influenza viruses are asymptomatic in a third of the cases and that only around 40% of the infected cases present fevers (Carrat *and others*, 2008). To be able to use the data generated by health care attendance for inference purposes, one has thus first to correct the data for positivity (to assess how many of people presenting the symptoms are actually consulting for a particular pathogen), and then estimate the ratio of people consulting for that pathogen over infections. This would ideally be estimated using serology data (Miller *and others*, 2010), but it is a difficult exercise which can result in inaccurate estimates of the number of cases in population (especially if done in real-time). For example, during the recent A/H1N1(2009) pandemic in the United Kingdom, the estimates of clinical cases were underestimated by a factor 3 (as the ratio of infections over estimated ILI was around 10) making it difficult to infer the likelihood of a second wave with the reopening of school in September '09 (Baguelin *and*

others, 2010). As demonstrated in Section 5, one of the potential practical benefits of the methodology developed in this paper is to provide relevant information on real time decision regarding the epidemic.

2.2 Time varying effective contact rate

Probability models provide a simplified representation of the real world, with different levels of sophistication. Regarding epidemics the most-detailed models account for individual characteristics, geographic distribution of individuals, situation-dependent disease transmission probabilities etc. On the other end of the scale there are models, such as the previously introduced SEIR model, that geographically aggregate the cases and consider that transmission processes are deterministic, occurring at a given frequency each time an infected and a susceptible meet. The level of complexity is often chosen on the basis of the available data. For example, in the case of noisy observations on weekly aggregated cases, SEIR-type models are often adopted. Apart from the argument of limited information in the data at hand, these models are also appealing for being easy to interpret. Nevertheless, there is a number of features of the epidemic that are being overlooked and assumptions that are unlikely to hold, that could potentially lead to poor inference.

In this paper we adopt stochastic extensions of the deterministic SEIR type models. Rather than ignoring deviations from reality, the aim is to introduce them in the model as dynamic system error and adjusting for their effect. This error can then be explored and potentially even re-modelled at an aggregate population level. We focus on large-scale epidemics, for which random effects in transmission processes can be considered to be well-approximated deterministically (Kurtz, 1981). We work under the paradigm stating that the limitations of the classically-used models do not arise from the variability of individual characteristics or the randomness in the transmission process, but mainly from the fact that the effective contact rate (β_t) evolves in time.

An early approach to estimate the time-varying reproduction potential of a disease can be found in P.E.M. Fine and Clarkson (1982). This method provides an estimate of the effective reproduction rate R_t, rather than the effective contact rate β_t, thus mixing the influence of the contact rate and of the population immunity to evaluate the explosive potential of the epidemic. It can be simply estimated by computing the average number of infections caused by each infected individual over a given period of time, which is obtained either through discrete generation models, or by reconstructing the chain of transmission (Cauchemez *and others*, 2006; Griffin *and others*, 2011; Wallinga and Teunis, 2004). 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 dif-

ferent. Not only does the latter limit the use of results on the drivers of R_t in different geographical contexts, but also in the very same region as population immunity can strongly change from one year to another as a mixed result of past epidemics and virus evolution. This is why we concentrate on estimating the trends of β_t rather than R_t.

In a number of existing approaches, the trajectory of β_t has classically been restricted to a finite-dimension function space, comprised of a finite set of parameters estimated from the data. There have been a variety of studies, based on different function spaces: a low-dimensional examples can be found in Cauchemez *and others* (2008), in which β_t is modeled as a piecewise linear functions. 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. Both of these studies inferred richer and more interesting patterns, while posing computational problems to which instructive solutions were brought and will be discussed further. An additional amount of flexibility has been incorporated in Pollicott *and others* (2009), where β_t can literally be any smooth function of time. Generally speaking, as the number of parameters required to estimate the trajectory of β_t increases, model-induced biases fade out at the expense of an increased variance. A compromise is required in order to improve robustness and is typically controlled through a context-specific regularizing parameter. For example, in He *and others* (2011), β_t is estimated using exponentiated cubic splines, the number of which is calibrated based on AIC.

2.3 Diffusion driven epidemic models

We consider models where diffusion processes are used to capture the time varying nature of the coefficients in (1). For reasons discussed in Section 2.2 we focus on the effective contact rate β_t, although alternative formulations are possible. The assigned diffusion is expected to capture time-varying features such as changes in the behaviour, preventive measures, seasonal effects, holidays etc. The model of (1) therefore becomes

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

where $\mu_x(.)$ and $\sigma_x(.)$ refer to the drift and volatility of the diffusion process governing β_t, and $h(.)$ is a positive-valued function to exclude negative contact rates. When prior knowledge on the evolution of β_t is available, it can be reflected in the drift and the volatility function. For example if the contact rate is expected to move towards some kind of equilibrium, 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 such prior information or in cases where the researcher is interested in

exploring the data without imposing restrictions, a Brownian motion can be used setting $\mu_x(.) \equiv 0$ and $\sigma_x(.) \equiv \sigma$ in the model of (2). The obtained output from this model can then be either reported as it is or used as an exploratory tool to formulate a specific drift function, thus leading to a more informative yet less flexible model. We focus on this model, henceforth denoted by BM, with the log function for $h(.)$. Note that this choice for β_t implies a 'rough' path satisfying the Markov and being continuous, yet non-differentiable. In cases where β_t is believed to evolve as a smooth function in time, higher order Brownian motions could be used instead. Loosely speaking, such choices may be regarded as equivalent to non-parametric approaches and in particular cubic splines estimation, as suggested in Wahba (1990), with the model in (2) imposing a prior on β_t and σ playing the role of a regularising factor.

Similar approaches using time-varying parameters include Cazelles and Chau (1997) and Cori *and others* (2009). In the former, an Extended Kalman Filter was used, accounting for both process and observation noise. The latter approach adopts a second-order Brownian motion (integrated Brownian motion) and circumvents the EKF-induced approximations using MCMC on a one-week time-step discretisation of the model. We expand on this approach and develop a general MCMC framework in the next section for constructing accurate and efficient MCMC schemes taking advantage of the specific structure of the previously mentioned models.

3 Bayesian inference via MCMC for diffusion driven epidemic models

This section presents a general inferential framework for diffusion-driven epidemic models. We adopt the Bayesian paradigm, incorporating parameter uncertainty and prior information in the estimates of β_t trajectories. The problem can also be cast as estimating partially observed hypoelliptic diffusions, and therefore presents various difficulties; see for example Pokern *and others* (2009). We begin by setting the model and demonstrating the need for MCMC. Existing MCMC algorithms from other applications are considered but, as we justify and demonstrate, they can lead to extremely inefficient MCMC chains. We address the issue by utilising the recently developed PMCMC methodology, and taking advantage of the specific model structure.

3.1 Model setup

For ease of exposition we focus on the family of models satisfying (2), although the developed methodology may cover different and more general models; e.g with different dynamical systems or more time-varying coeffi-

cients. Since these are continuous-time models, t can take any real value between t_0 and t_n. The trajectory of the ODE states vector $V_t = \{S_t, E_t, I_t, R_t\}$ between instants t_i and t_j is denoted by $V_{i:j}$. The data, denoted by $y_{1:n} = \{y_{t_1}, .., y_{t_n}\}$, usually provide information for values of I_t at specific times (prevalence data) or for integrals of V_t (incidence data). In either case, we assume that the measurements on I_t or the integrals of V_t are obtained with error for the reasons mentioned in Section 2.1. The noise distribution is denoted with $\mathbb{P}_y$ with density $f(y_{1:n}|V_{0:n}, \theta_y)$. The structure of the model in (2) allows to write V_t as a deterministic function of x_t and the parameter vector θ_v, which in this cases consists of k, γ and the initial states V_0 and x_0. This function is the solution of the ODE and in most cases will not be available in closed form, but can rather be expressed as a time integral of 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 $\pi(\theta)$

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

As already noted, an exact solution of the ODE will generally not be available in closed form, nor will be the density of $\mathbb{P}_x$. In order to evaluate (4), we apply a time discretisation on the trajectories of x_t and therefore β_t and V_t. More specifically, we introduce m points between each pair of successive observation times t_i and t_{i+1} $(i = 0, 1, \ldots, n-1)$. Regarding notation, when referring to the discretised representation of a trajectory, the superscript dis will be used; for example for a step δ the discrete skeleton of x_t will be denoted with $x^{dis}_{0:n} = \{x_0, x_\delta, x_{2\delta}, \ldots, x_{t_n}\}$. The time discretised path $x^{dis}_{0:n}$ allows for approximate evaluations of (4). The Euler-Maruyama scheme can then be applied, so that $d\mathbb{P}_x$ is approximated by

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

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

3.2 Partial or approximate solutions

Before getting to the details of the MCMC algorithms, we mention two alternative existing approaches. The first one uses the Extended Kalman Filter, based on normal approximations of the observation process and the conditional densities $p(x_i|\theta, y_{1:i})$. Given the normality assumption, all the quantities of interest are tractable or can be approximated using Taylor expansions. This results in a substantial reduction of computational cost, but the performance depends on the error induced by the Taylor and normal approximations. As demonstrated in Section 4 this error can be non-negligible. Nevertheless, the EKF can still be used as a tool to locate good initial values and construct efficient proposal distributions for MCMC algorithms. It is also used to optimize the propagation of particles in SMC algorithms, but either implying a strong computational cost (Särkkä and Sottinen, 2008) or crude time discretisations (Dukic *and others*, 2009).

The second approach is the MIF framework of Ionides *and others* (2006) that, in our context, can be used to maximise $p(y_{1:n}|\theta)$. The optimisation algorithm relies on estimates of $\nabla_\theta p(y_{1:n}|\theta)$ obtained by particle filters (presented in Section 3.4) run in an augmented state space model. Our experience suggests that this method can efficiently obtain the posterior mode of θ in our setting. Nevertheless, our primary aim is to draw inference on β_t while incorporating the uncertainty over the constant parameters θ, rather than fixing them to an estimated value.

3.3 Data augmentation via Gibbs schemes

The model, as defined in (7) can be put in the stochastic volatility models framework of Chib *and others* (2006), Golightly and Wilkinson (2008) or Kalogeropoulos (2007). In these approaches a Gibbs scheme is enforced to sample from the joint posterior of $x_{0:n}^{dis}$ and θ in (4). The data augmentation algorithm alternates between drawing $x_{0:n}^{dis}$ conditional on the current value of θ, and then updating θ from its full conditional posterior as obtained from the augmented likelihood. The MCMC protocol ensures that the chain will provide samples from the marginal posteriors of $x_{0:n}^{dis}$ and θ. While this is in principle the case, the performance the MCMC algorithm may become unacceptably poor or even degenerate. There are two essential issues associated with such schemes. The first concerns the non-trivial step of sampling from the full conditional posterior defined on the diffusion pathspace of x_t. The second problem is caused by the high posterior correlations $x_{0:n}^{dis}$ and θ leading to reducible chains as m increases (Roberts and Stramer, 2001).

The majority of the existing literature on data augmentation algorithms for diffusions handles the conditional updates of $x_{0:n}^{dis}$ with an independence sampler. However, it is difficult to find proposal distributions for the entire $x_{0:n}^{dis}$ to ensure reasonable acceptance rate. A common approach for dealing

with this issue is to split the process into blocks. In order to be consistent with the continuity of the path and ensure that all points are updated, an overlapping blocks scheme is usually adopted. An alternative way to update $x_{0:n}^{dis}$ is to use the particle filter, which in this context would result in the Particle Gibbs algorithm of Doucet and Johansen (2009). Another option is to use the Hamiltonian Monte Carlo algorithms of Beskos *and others* (2012).

However, unless the issue of high posterior correlation between $x_{0:n}^{dis}$ and θ is resolved, none of these updating schemes will improve the overall MCMC performance substantially. 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

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

The implication of (6) is a point mass conditional posterior for σ as $\delta \to 0$. In practice this translates into an increasingly slow MCMC algorithm as the number of imputed points m increases, resulting in a convergence rate of $O(m)$. In some cases, the problem can be tackled with suitable reparameterisation. The approach of Roberts and Stramer (2001) involves transforming to a diffusion $\dot{x}_t$ with unit volatility. An alternative reparemetrisation, more general for multidimensional x_t's, is offered by Chib *and others* (2006); see also Golightly and Wilkinson (2008). The key feature of these algorithms is that the ODE states vector $V_{0:n}^{dis}$ becomes a function of σ and $\dot{x}_t$. Hence, as σ is typically updated through a Metropolis step, every proposed value of σ^* is associated with some proposed values $V_{0:n}^{dis*}$. This succeeds into breaking the perfect dependence between $V_{0:n}^{dis}$ and σ, even for $m \to \infty$. But in our case, it also creates another issue. Since components of $V_{0:n}^{dis}$ (or functionals thereof) are observed with error, the proposed values $V_{0:n}^{dis*}$ should be close to the data for the move to be accepted. Efficient Gibbs schemes can be developed in some specific settings (Choi and Rempala, 2011), but 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 results from simulated data, supporting this argument, are provided in the in Appendix D of Supplementary Materials.

In the next section, we overcome this issue by updating $x_{0:n}^{dis}$ and θ jointly using the PMCMC algorithm. The use of PMCMC is essential as it is not straightforward to construct joint updates with neither of the approaches mentioned in this section.

3.4 Particle Markov Chain Monte Carlo algorithms

Particle filters, also called Sequential Monte Carlo algorithms, are used to recursively explore conditional densities in state space models (Doucet and Johansen, 2009). In our context, they can be used for filtering purposes,

in other words to provide samples from $p(x_i|y_{1:i},\theta)$. For given values of θ, they sequentially propagate N particles $(\tilde{x}_i^j)$ from t_0 to t_n along the model, selecting at each time-step i the particles (or trajectories) that best fit the observed data $y_{1:i}$. Algorithm 1 shows how they can be applied to diffusion driven epidemic models. Note that the quantity $L^{i+1}(\theta)$ provides with an

Algorithm 1 Particle Smoother 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

unbiased estimate of the likelihood $p(y_{1:i}|\theta)$. The resampling step is essential to control the variance of that estimate over time. Algorithm 1 also provides a random sample from the point-wise conditional distribution $p(x_i|y_{1:n},\theta)$ which is also known as smoothing.

Our goal however is to sample from the joint posterior $\pi(x_{1:n},\theta|y_{1:n})$. This can be achieved with the PMCMC algorithm, introduced and established theoretically in Andrieu *and others* (2010), that successfully integrates the particle filter in the context of an MCMC algorithm. Its generic implementation is presented in Algorithm 2. Through this process, the issues

Algorithm 2 Particle Monte Carlo Markov Chain algorithm

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

of section 3.3 are addressed as $x_{0:n}^{dis}$ and θ are sampled jointly. In other words $x_{0:n}^{dis}$ is being numerically integrated out, in the context of Andrieu

and Roberts (2009), while a sample from its posterior is obtained at each iteration of the MCMC algorithm.

The PMCMC algorithm is associated with increased computational cost. At each iteration, N particles are propagated forward to obtain a draw from $p(x_{0:n}^{dis}|y_{1:n}, \theta)$. While in theory the algorithm is valid even for a single particle, large values of N are usually required to limit the variability in the estimates of the posterior. This is crucial for controlling the acceptance rate of the MCMC algorithm and therefore its efficiency; see the Appendix B of the Supplementary Materials. Despite the increased computational cost, PMCMC remains feasible in the context of epidemic models. Even for relatively large datasets, the computing time can be limited to less than two hours in a standard specification PC. Implementations in real time are also available to sample from $\pi(x_{1:i}, \theta|y_{1:i})$ as the $i-$th observation becomes available; see Section 5 for an application. In this setting, the SMC^2 algorithm of Chopin *and others* (2011) provides a more efficient way, but the difference would be minor given the short length of epidemic time series.

4 Simulations

4.1 Assessing the validity of the particle MCMC

In this section we illustrate the algorithm and assess its performance on simulated data. We focus on the BM model, where $\log(\beta_t)$ follows a Brownian motion with volatility σ, corresponding to a case where no information regarding the shape of β_t is available. In order to simulate data, we obtain β_t trajectories 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 $\int_{week\ i} kE_t dt$. To preserve positivity, we assumed that each $\log(y_i)$ follows a Normal distribution with mean $\log(\int_{week\ i} kE_t dt)$ and variance τ^2. In both cases the parameters were tuned to obtain realistic epidemic incidence curves, and observation were generated using a 0.1 noise. Fig. 10 shows estimates of the path of the system provided by the PMCMC algorithm, with associated 95% pointwise credible intervals, in a setting where no information was available regarding τ and σ. The posterior output is in good agreement with simulation trajectories, which suggests that the underlying and unobserved trajectory of the effective contact rate can be accurately estimated in spite of uncertainties regarding its variability and the observation noise. More details are provided in the Appendices A and C of the Supplementary Materials.

4.2 Comparison with the EKF

Next, we compare the PMCMC framework with the EKF approach to assess the gains 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 of the previous section. 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 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.

5 The 2009 A/H1N1 pandemic

5.1 Data, model and priors

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 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 respiratorial diseases, based on subsequent virological positivity tests. Corrections for asymptomatic infections and the patients's the propensity to consult were also made. Overall the two datasets are different by a coefficient $c = 10$, which 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 exploited further in Section 5.3.

Given the value of c, the estimates of weekly number of cases in the population $y_{1:n}$ are still uncertain, which was incorporated in our model through measurement error. The noise model of Section 4.1 was used, combined to a BM formulation of $\mathbb{P}_x$. We assigned flat truncated normal priors to τ, σ

and β_0, as in Section 4.1. Consistently with Baguelin *and others* (2010), the prior constrained the latent period k^{-1} between 1.55 and 1.63 days, and the the infectious period γ^{-1} between 0.93 and 1.23 days. Finally, we put a Uniform prior from 0 to 1 to the initial proportions in E and I, whereas a $\mathrm{N}(0.15, 0.15^2)$, truncated at 0 was assigned to the initial proportion in R.

5.2 Parameter estimation

The PMCMC algorithm of section 3.4 was applied to the data and model of the previous section. 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 of the effective contact rate are observed during school and holidays periods, appearing to be synchronised with schools opening and closing. Posterior and prior densities for the static parameters are shown in Fig. 3. Regarding k, γ and $R(0)$, there is not much information in the data as the posterior densities are very close with their prior counterparts. However, there is information on the remaining parameters, that is summarised by their posterior. A higher degree of information could be obtained from additional sources of data such as genealogies, that can be integrated in the PMCMC framework (Rasmussen *and others*, 2011).

5.3 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 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 requires less than two hours which would not have been an issue in an online weekly data setting. In order to reduce uncertainty, especially at early stages, the value of τ^2 was set to 0.1. Although this value is in line with Fig. 3, it is also a plausible value that could have been selected in the absence of the previous analysis.

The main results are shown in Fig. 4. In Fig. 5, we repeat the same analysis 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. 4 and Fig. 5 the BM model is consistently better in that respect. Second, we apply the methodology of MIF in the following way. First, estimates θ were 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.

In the remainder of this section, we describe how the estimate of the posterior density $\pi(x_{1:i}, \theta|y_{1:i})$ could have been used on August 1^{st}. At that time, 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 it a sufficient level of herd immunity to stop the epidemic. In this case, no second wave was to be expected in September. Scenario (ii) was backed further with concerns on whether the correcting multiplicative factor c could have been higher than 10, given the uncertainty at that time (Baguelin *and others*, 2010).

The PMCMC algorithm, run up to August 1st, provides with samples from the posterior of the difference in β_t between July 13^{th} (before the decrease in incidence) and August 1^{st}. The 97.5% point of this posterior is -0.32, indicating a necessary 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. 6. 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 is clearly an unrealistic value, we provide 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.

6 Discussion

We presented a general modelling framework flexible enough to incorporate non-linear dynamical systems, time-varying parameters such as the effective contact rate, stochastic and measurement error, and parameter uncertainty.

The adopted models were based on standard SEIR-type population level formulations. We used stochastic extensions of these models, focusing on the effective contact rate to account for changes in individual behaviours, evolution of the virus, seasonality, schools closure and preventive measures. The main motivation was to account for these factors while keeping the interpretation of the model simple. Along with the modelling framework we also presented a computational machinery based on recently developed PMCMC algorithm developed in Andrieu *and others* (2010), while comparing with alternative methodologies. The problem is quite challenging from a computational point of view, requiring inference for partial and noisily observed hypoelliptic diffusions. The results of this paper are therefore relevant for applications of MCMC in related problems from different fields. Apart from drawing offline inference we illustrated in the 2009 H1N1 pandemic case how such models can be used in real time.

For illustration purposes we relied on a simple SEIR model and unstructured data cases, but the same approach could be followed on more complex models; e.g. accounting for age or risk factors, or decoupling the nature and the frequency of the contacts between individuals in β_t formulations. We have shown that this methodology can be used as an exploration tool, which can help in developing richer models and testing alternative scenarios, either for public health interventions or to bring further insights on the influence of extrinsic factors as climate on the dynamics of epidemics. It should also be noted that this framework is suited to the incorporation of 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 long influenza time series, the developed methodology can be applied to other cases. Current work includes it application as part of the CHARME project (Boily *and others*, 2007) to assess the impact of the large-scale Avahan intervention against HIV in India. Extensions to models with additional dynamic error components, e.g. in the compartments, are also possible. The presented approach may 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 effective contact rate.

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 dataset, its algorithmic aspects can be decoupled from the modeling aspects. This provides the possibility to develop generic inference packages: we are currently working towards integrating it to the R package POMP and in associated future web-based versions.

Acknowledgments

The authors would like to thank John Edmunds, Nikolaos Demiris and Wicher Bergsma for their helpful comments.

7 Supplementary Materials

Supplementary materials contain implementation details for the PMCMC algorithm (Appendix A), details on the MIF approach (Appendix B) and additional information on Sections 4 and 5 (Appendix C). Finally, in Appendix D we present the MCMC algorithms and the issues discussed in Section 3.3.

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 theparameter vector θ.

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. 7 for two different datasets of weekly influenza data. In such a case, $\delta = 0.1$ 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 noisiest the estimate $\hat{p}(y_{1:n}|\theta)$ of the likelihood 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 $O(N_{parts})$; a compromise needs to be achieved. Fig. 8 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. This figure shows that when the observational noise decreases, more particles are needed, which is explained by the fact that the particles need to fit the observations with more accuracy.

A three-steps adaptive algorithm

The task of updating a component of the parameter vector θ requires to run a full particle filter and is therefore associated with high computational cost. The construction of joint updates of θ is therefore essential for an efficient PMCMC algorithm. The state-of the art MCMC schemes require derivatives of the posterior $\pi(\theta|y_{1:n})$ that are not available in closed form. Hence, we use a Random-Walk Metropolis-Hasting algorithm, for which the proposal density $Q(.,\theta)$ is a multivariate normal density centered in θ, with covariance matrix is $\epsilon^2\Sigma$. Here, ϵ is a scale parameter that is easily tuned to achieve a 23% acceptance rate following **?**. The choice of covariance Σ is crucial in order to construct a sampler that moves efficiently in the θ space. The aim here is to match the covariance of the unknown posterior density. We used the following three-steps adaptive algorithm:
step 1: Use an Extended Kalman Filter (EKF) to locate the posterior mode and compute the observed Fisher Information Matrix of the posterior density $p_{EKF}(\theta|y_{1:n})$. Set Σ_1 to be the inverse of this matrix.
step 2: Run the PMCMC for a certain number of iterations, allow for a burn in period, and set Σ_2 to be the estimated posterior covariance.
step 3: Run the final PMCMC.

Appendix B: details on Maximum likelihood via Iterated Filtering

This algorithm was applied in our article on the "real-time" example. It is a popular tool for inference on epidemic models. More specifically, it is used to estimate θ by maximising the likelihood $p(y|\theta)$. Particle filters are are utilised to bypass the generally intractable likelihood by producing unbiased Monte Carlo estimates of it. Local derivatives of the likelihood can significantly improve the maximisation procedure: the iterated filtering approach provides with estimates of such derivatives while offering an efficient way

to find the maximum. The main idea is to replace the epidemic model $\mathcal{M}$ with a related model $\mathcal{M}'$ where the static parameters are allowed to vary in time following simple random walks. The variability of these random walks is progressively brought to zero in order to approach $\mathcal{M}$. For more details, see Ionides *and others* (2006).

Once the MIF has been applied and the parameter θ^* that maximizes $p(y|\theta)$ has been estimated, a particle smoother can be applied to obtain an estimate of the conditional posterior density $p(\beta_{1:n}|\theta^*, y)$, which is then compared to the marginal posterior $p(\beta_{1:n}|y)$ provided by the PMCMC. For a fair comparison, we introduce previous information available regarding k, γ and $R(0)$ for which informative priors are used in the PMCMC. This is done by constraining each of these parameters in a segment $[a; b]$ corresponding to 95% intervals of the corresponding prior densities (i.e. the MIF is run on *logit* transformations of these variables). Results of the comparison are detailed in the main text.

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

Assessing the validity of the particle MCMC

This section deals with a series of experiments aimed at assessing 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 $\beta_{1:n}$. The data fits and estimates of $\beta_{1:n}$ for the experiments with $\tau = 0.1$ (Exps 1.a and 2.a) are shown in Fig. 1 of the main text. Corresponding figures for Exps 1.b and 2.b are shown in Fig. 10 of the SM. Table 1 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 mis-specification fitting Brownian motion to 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.

Comparison with the EKF

Here we give some details regarding the results of the comparison between the particle and the extended Kalman filter. In order to generate diverse but realistic epidemic datasets from known trajectories of the efficient contact rate, 100 time-series were "reverse-engineered" from randomly selected ILI time series freely available from the Google Flu Trend website (Ginsberg *and others*, 2008). These 7-month long time series provide with estimates of the weekly number of ILI cases in different regions or countries, that have been validated against data from national surveillance companies. For each of such time series $y_{1:n}^{Goog}$, the following process was followed:

For each j in 1,..,100

1. Apply our methodology to estimate $p(\beta_{1:n}, \theta | y_{1:n}^{Goog,j})$, with unknown initial conditions and static parameters, with priors set as in Baguelin *and others* (2010) and the other experiments of this paper.

2. Generate a sample $\{\beta_{1:n}^{Sim,j}, \theta_j\}$ from this posterior distribution.

3. Generate a dataset $y_{1:n}^{Sim,j}$ corresponding to $\{\beta_{1:n}^{Sim,j}, \theta_j\}$, with observation errors (of amplitude $\tau = 0.1$)

4. Apply different algorithms (EKF, particle filter, particle smoother) to obtain estimates $\beta_{1:n}^{alg,j}$ of $\beta_{1:n}^{j}$ ($alg \in \{EKF, Filt, Sm\}$) with all initial conditions and static parameters being known

5. Compute the bias and MSE between $E[\beta_{1:n}^{alg,j}]$ and $\beta_{1:n}^{Sim,j}$

The results are presented in Table 2. This tables shows the bias and MSE obtained with each of the three methods. Two comparisons are of interest: EKF versus particle filter, and particle filter versus particle smoother. In the first comparison we evaluate the impact of the Taylor approximations of the EKF as all the other elements, including the observation process and amount of data used ($y_{1:i}$ when estimating β_i), are common to both approaches. The second comparison illustrates the impact of using the smoothing distribution instead of the filtering distribution (using $y_{1:n}$ instead of $y_{1:i}$ when estimating β_i). What this table shows, as mentioned in the text, is a strong reduction in bias when using the EKF instead of the particle filter. However, the reduction in MSE is not as strong as it is compensated by an increase in variance. On the other hand, approximating the exact smoothing distribution brings significant improvement with regards to the filtering one, as the MSE and the bias are reduced by respectively 87 and 64%.

A/H1N1 pandemic

We provide here with the corresponding trace plots (Fig. 9) 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.

Appendix D: details of the Gibbs 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

$$\begin{cases} dx_t & = \mu_x(x_t, \theta_x)dt + \sigma_x(x_t, \theta_x)dB_t \\ y_{1:n}|V_{0:n}, \theta_y & \sim \mathbb{P}_y(y_{1:n}|V_{0:n}, \theta_y), \quad V_{0:n} = g(x_{0:n}, \theta_v) \end{cases} \tag{7}$$

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 $x_{0:n}^{dis}$ and $V_{0:n}^{dis}$ and proceed using the Euler approximating scheme. A Gibbs algorithm alternates between updating the trajectories of $x_{0:n}^{dis}$, and consequently $V_{0:n}^{dis}$, 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 order to overcome this problem, suitable reparametrisation can be applied such as the one 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:

$$du_t = \nu(u_t; \theta_x)dt + dB_t \ , \quad u_0 = 0 \ , \tag{8}$$

where

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

Alternatively the reparametrisation in Chib *and others* (2006) uses the transformation below

$$w_t = \frac{x_t - \big(x_{t-\delta} - \delta\mu_x(x_t, \theta_x)\big)}{\sigma_x(x_{t-\delta}, \theta_x)}, \forall t. \tag{9}$$

In either case the state vector is now written as

$$V_{0:n}^{dis} = h_u(u_{0:n}, x_0, \theta_x, \theta_v) \quad \text{or} \quad V_{0:n}^{dis} = h_w(w_{0:n}, x_0, \theta_x, \theta_v), \tag{10}$$

and the model is then defined by (8) or (9), (10) and 7. Although non trivial, the problem of updating $u_{0:n}^{dis}$ or $w_{0:n}^{dis}$, and consequently $V_{0:n}^{dis}$, given θ can be carried out. This 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. Nevertheless, the step of updating θ_x given $u_{0:n}^{dis}$ or $w_{0:n}^{dis}$ leads to quite poor performance in this setting. Such a step 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

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

The problem with the above algorithm is that for every proposed value of θ_x corresponds to 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. Hence 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 the time horizon of the epidemic increases. The performance of this algorithm in the simulated and real data of this paper was unacceptably poor.

References

Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B* **72**(3), 269–342.

Andrieu, C. and Roberts, G.O. (2009). The pseudo-marginal approach for efficient monte carlo computations. *The Annals of Statistics* **37**(2), 697–725.

Baguelin, M., Hoek, A.J.V., Jit, M., Flasche, S. *and others*. (2010). Vaccination against pandemic influenza A/H1N1v in england: a real-time economic evaluation. *Vaccine* **28**(12).

Beskos, A., Kalogeropoulos, K. and E., Pazos. (2012). Advanced hybrid Monte Carlo methods on diffusion pathspace. Working paper.

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

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

Carrat, F., Vergu, E., Ferguson, N.M., Lemaitre, M. *and others*. (2008). Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American journal of epidemiology* **167**(7), 775.

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, NP. (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*.

Choi, B. and Rempala, G.A. (2011). Inference for discretely observed stochastic kinetic networks with applications to epidemic modeling. *Biostatistics*.

Chopin, N., Jacob, P.E. and Papaspiliopoulos, O. (2011). SMC^2: A sequential monte carlo algorithm with particle markov chain monte carlo updates. *Arxiv preprint arXiv:1101.1528*.

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. (2009). A tutorial on particle filtering and smoothing: fifteen years later. *Handbook of Nonlinear Filtering*.

Dukic, V.M., Lopes, H.F. and Polson, N. (2009). Tracking flu epidemics using google flu trends and particle learning. *SSRN eLibrary, http://ssrn. com/paper* **1513705**.

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

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.

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

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* **1**, 1–6.

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., Andrews, N. and Zambon, M. (2010). Incidence of 2009 pandemic influenza a H1N1 infection in England: a cross-sectional serological study. *The Lancet* **375**(9720), 1100–1108.

Øksendal, B. (2003). *Stochastic differential equations: an introduction with applications*. Springer Verlag.

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

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.

Pollicott, M., Wang, H. and Weiss, H. (2009). Recovering the time-dependent transmission rate from infection data. *Arxiv preprint arXiv:0907.3529*.

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 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.

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

Unkel, S., Farrington, C., Garthwaite, P.H., Robertson, C. and Andrews, N. (2010). Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A*.

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

Wallinga, J. and Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology* **160**(6).

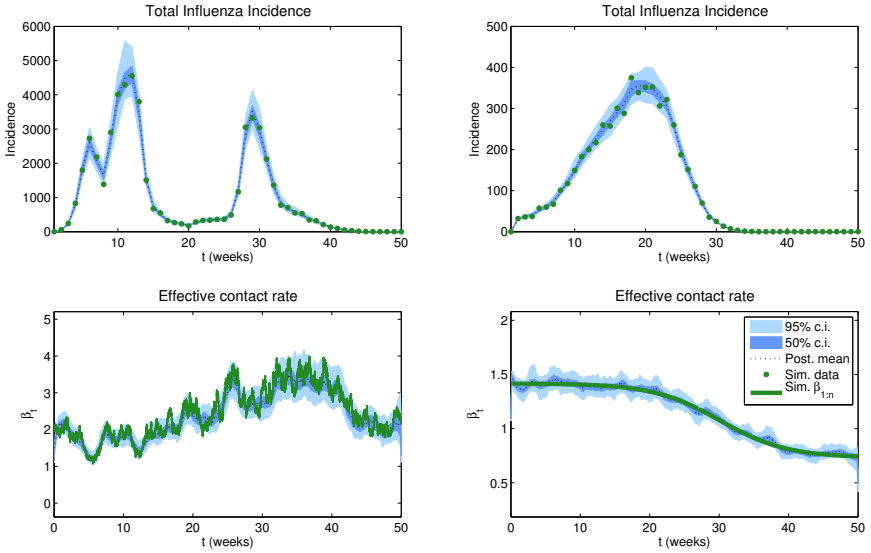

Figure 1: Illustration of how the underlying dynamic of the effective contact rate can be estimated from weekly recorded cases
a. Experiment 1: weekly number of cases observed with noise (top left)
b. Experiment 2: weekly number of cases observed with noise (top right)
c. Experiment 1: simulated and estimated trajectory of the effective contact rate (bottom left)
d. Experiment 2: simulated and estimated trajectory of the effective contact rate (bottom right)

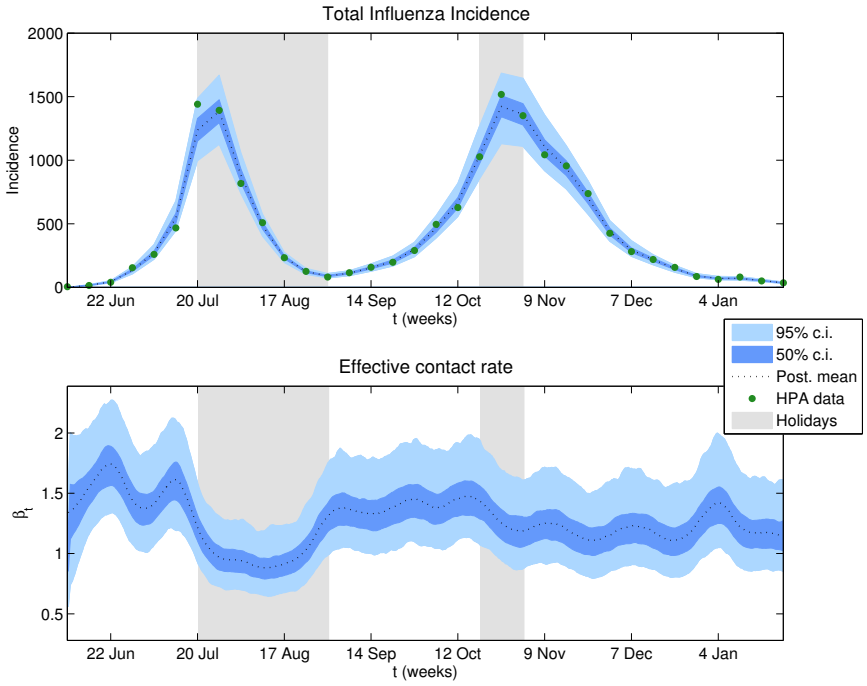

Figure 2: Offline estimates of the effective contact rate during the 2009 A/H1N1 pandemic
a. HPA estimates of the weekly total number of influenza cases in London (per 100 000 inhabs.) (top)
b. offline estimates of the effective contact rate. Dark and light blue show credible intervals, respectively at 95% and 50% levels. The mean estimate is plotted in black, and holidays are indicated by a light grey area. (bottom)

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

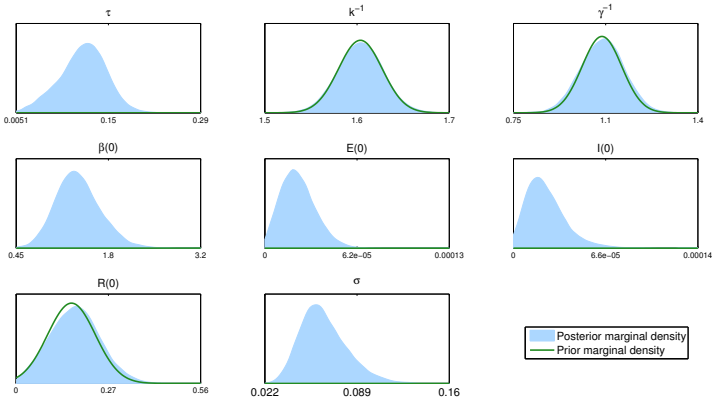

Figure 3: Posterior distribution of the constant parameters obtained from the 2009 H1N1 incidence data
Informative priors are used for k^-1, γ^-1 and R(0), for which the posterior shows no additional information. Informative posterior densities are obtained for the other quantities.

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

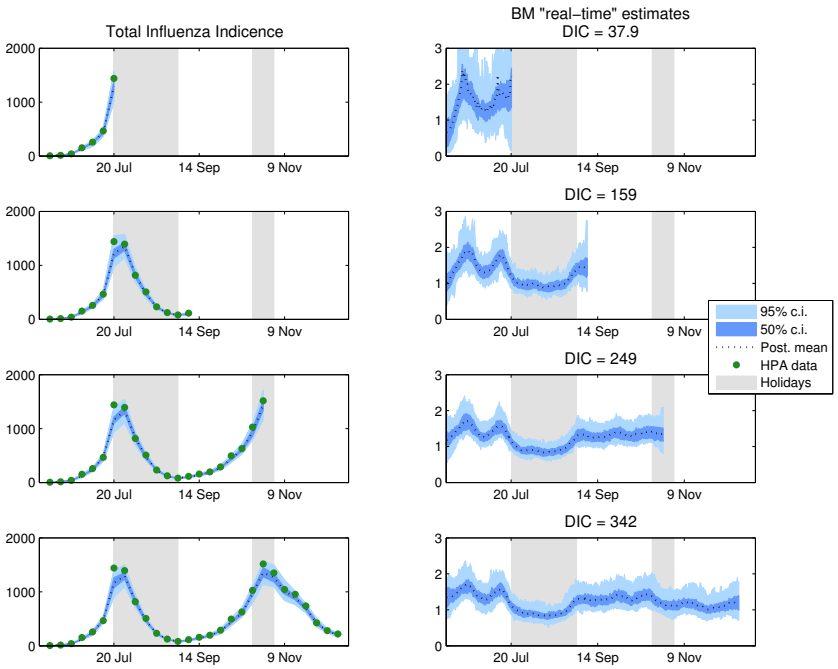

Figure 4: What could have been inferred by carefully following the epidemic in real time?
a. HPA estimates of the weekly total number of influenza cases in London (per 100 000 inhabs.) (left)
b. "real-time" estimates of the effective contact rate. Dark and light blue show credible intervals, respectively at 95% and 50% levels. The mean estimate is plotted in black, and holidays are indicated by a light grey area. (right)

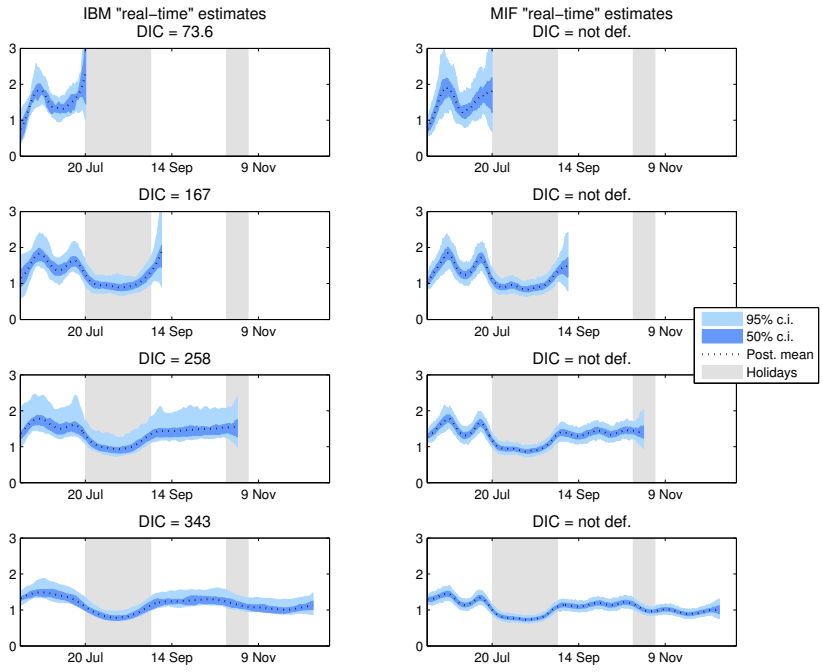

Figure 5: Modeling choices and implications, aiming for robustness
a. Estimates from an alternative modeling approach: exploring the full posterior density of an IBM diffusion model (left)
b. Estimates from an alternative methodological approach: exploring the posterior density of a BM diffusion model conditionned on a likelihood maximizing parameter θ^* provided by the MIF algorithm (right)

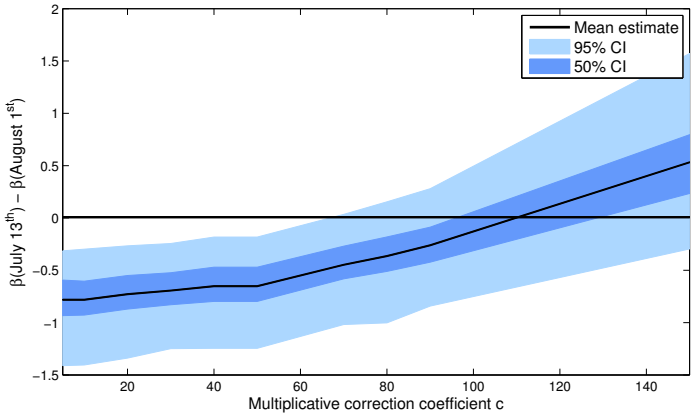

Figure 6: The implication of different scenarios for the real value of underreporting on the decrease in condom use between July 13^{th} and August 1^{st}

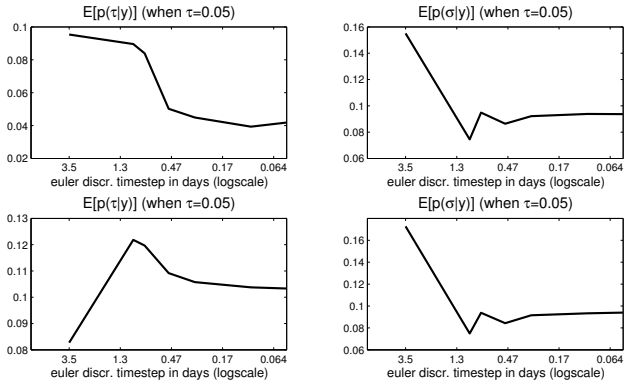

Figure 7: Convergence of the posterior density as the Euler discretization time-step δ decreases (x-axis in the log-scale)

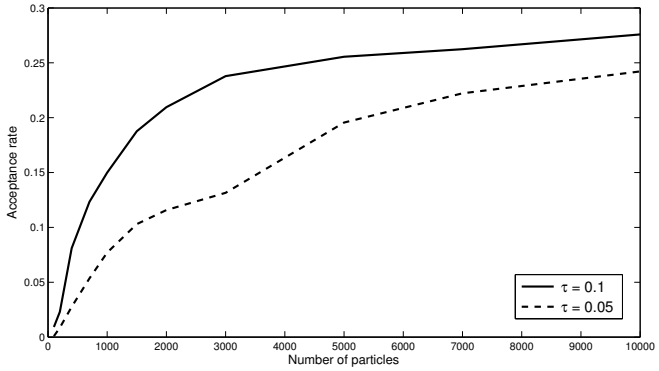

Figure 8: Acceptance rate as a function of N_{parts}, in two situations where the noise amplitude is respectively 10% (full line) and 5% (dotted line).

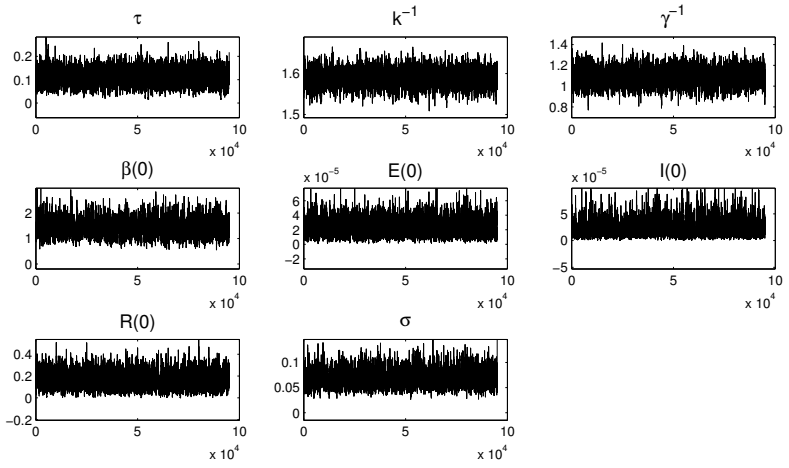

Figure 9: MCMC traceplots for each component of θ

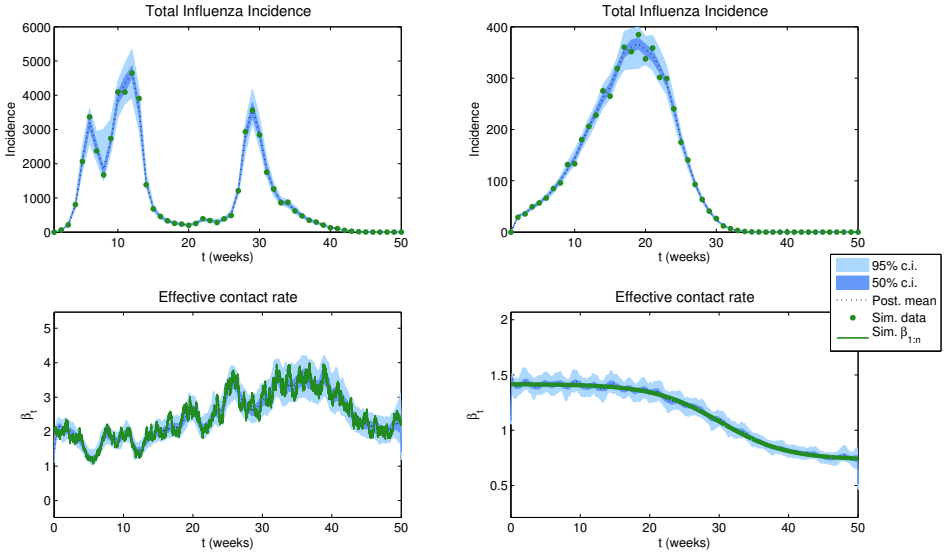

Figure 10: Illustration of how the underlying dynamic of the effective contact rate can be estimated from weekly recorded cases ($\tau = 0.05$)
a. Experiment 1.b: weekly number of cases observed with noise
b. Experiment 2.b: weekly number of cases observed with noise
c. Experiment 1.b: simulated and estimated trajectory of the effective contact rate
d. Experiment 2.b: simulated and estimated trajectory of the effective contact rate