Mathematical Biosciences 255 (2014) 21–32

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Comparison of the performance of particle filter algorithms applied to tracking of a disease epidemic

Daniel M. Sheinson [a,*], Jarad Niemi [b], Wendy Meiring [a]

[a] Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA
[b] Department of Statistics, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 5 July 2013
Received in revised form 28 April 2014
Accepted 27 June 2014
Available online 9 July 2014

Keywords:
Bayesian estimation
Epidemiological modeling
State-space modeling
Sequential Monte Carlo
Particle filtering

ABSTRACT

We present general methodology for sequential inference in nonlinear stochastic state-space models to simultaneously estimate dynamic states and fixed parameters. We show that basic particle filters may fail due to degeneracy in fixed parameter estimation and suggest the use of a kernel density approximation to the filtered distribution of the fixed parameters to allow the fixed parameters to regenerate. In addition, we show that "seemingly" uninformative uniform priors on fixed parameters can affect posterior inferences and suggest the use of priors bounded only by the support of the parameter. We show the negative impact of using multinomial resampling and suggest the use of either stratified or residual resampling within the particle filter. As a motivating example, we use a model for tracking and prediction of a disease outbreak via a syndromic surveillance system. Finally, we use this improved particle filtering methodology to relax prior assumptions on model parameters yet still provide reasonable estimates for model parameters and disease states.

1. Introduction

State-space models are commonly used in the analysis of biological data [1–4]. In particular, these models are frequently used for disease outbreaks to simultaneously model the underlying disease dynamics and the observation process [5–7,3,8]. Together with syndromic surveillance systems [9–13], these models are used to identify emerging disease outbreaks [14], estimate their severity [6], and predict their duration [7]. Typically, the general form of these models may be reasonably assumed, but the models will have unknown fixed parameters that define the disease dynamics and the observation process for a particular outbreak. Based on similar previous outbreaks, the range and likely values for these fixed parameters may be available.

In statistical applications where prior knowledge or beliefs about unknown quantities are available, the Bayesian framework is often convenient for performing statistical analysis. Bayesian inference is conducted through the posterior distribution of any unknown quantities, obtained by updating prior information using observed data. However, the calculation of the posterior distribution in state-space models frequently involves complicated integrals without an explicit analytical form. The most common approach to approximating these posterior distributions is Markov chain Monte Carlo (MCMC) [15]. In a sequential context, e.g. syndromic surveillance, MCMC is inefficient due to the increase in computational cost incurred by the need for the entire MCMC to be rerun as each new observation arrives. Sequential Monte Carlo (SMC) – or particle filtering – methods enable on-line inference by updating the estimate of the posterior as new data become available. Furthermore, SMC methods can be flexible, general, easy to implement, and amenable to parallel computing. For a general introduction, please see [16,17].

Early SMC methods, including the bootstrap filter [18,19] and the auxiliary particle filter [20], assumed all fixed parameters were known. A key defining step in these filters is the use of resampling, which results in particles with low probability being eliminated and particles with high probability being duplicated. When all fixed parameters are known, these filters work remarkably well. In the presence of unknown fixed parameters, these filters suffer dramatically from a *degeneracy* issue due to the fixed parameters (being treated as dynamic states with degenerate evolutions) never being regenerated, and thus only a few distinct values for the fixed parameters remain after a few time points. To combat this degeneracy, a number of alternative approaches have been introduced. In

Abbreviations: BF, bootstrap filter; APF, auxiliary particle filter; KDPF, kernel density particle filter.

* Corresponding author. Tel.: +1 847 609 7824.
E-mail addresses: sheinson@pstat.ucsb.edu (D.M. Sheinson), niemi@iastate.edu (J. Niemi), meiring@pstat.ucsb.edu (W. Meiring).

http://dx.doi.org/10.1016/j.mbs.2014.06.018
0025-5564/

this paper, we focus on the *kernel density particle filter* [21] due to its wide applicability, ease of implementation, and good performance.

The rest of the article proceeds as follows. Section 2 contains a description of state-space models and sequential estimation. Section 3 describes a variety of particle filtering methodologies including the kernel density approach of [21]. In Section 4, we introduce a nonlinear dynamic model for a disease epidemic similar to the one used in Skvortsov and Ristic [3]. In Section 5, we apply the particle filtering methods described in Section 3 to the model described in Section 4. In Section 6, we benefit from the efficiency of more recent particle filtering methods to estimate a more complicated model. In Section 7, we conclude by reviewing more advanced particle filtering methods as well as the general performance and capabilities of SMC algorithms.

2. State-space models

State-space models are a general class of statistical models used for analysis of dynamic data and have been used extensively in modeling disease outbreaks [5,22,6,7,3]. State space models are constructed using an observation equation, $y_t \sim p_{y,t}(y_t|x_t,\theta)$, and a state evolution equation, $x_t \sim p_{x,t}(x_t|x_{t-1},\theta)$, where y_t is the observed response, x_t is a latent, dynamic state, and θ is an unknown fixed parameter, all of which could be vectors. The y_ts are assumed independent given x_t and θ, and the x_ts are assumed independent given x_{t-1} and θ. The distributions $p_{y,t}$ and $p_{x,t}$ are assumed known conditional on the values of θ and x_t in the observation equation and θ and x_{t-1} in the evolution equation. Depending on whether the observations and the states are continuous or discrete, the distributions themselves may be continuous or discrete. The distributions are typically assumed to only vary with x_t and θ, and therefore the t subscript is dropped. For simplicity, we also drop the x and y subscript and instead let the arguments make clear which distribution we are referring to. Thus, the general state-space model is

$$y_t \sim p(y_t|x_t,\theta) \quad x_t \sim p(x_t|x_{t-1},\theta).$$

A fully specified Bayesian model is obtained by also specifying the prior $p(x_0,\theta)$.

The dimension of x_t need not remain constant with respect to t. For instance, we could describe a process where x_t depends on the entire history of states up to t by letting $x_{t-1} = (x_1^*, x_2^*, \ldots, x_{t-1}^*)'$ and defining $x_t = (x_{t-1}, x_t^*)'$, where x_t^* is the new state generated at time t.

Special cases of these state-space models include hidden Markov models [23,24], where the state x_t has discrete support, and dynamic linear models (DLMs) [25], where each distribution is Gaussian whose mean is a linear function of the states and whose variance does not depend on the mean. The disease outbreak models discussed in Section 4 are specific cases of state-space models, but we introduce these models in generality here because the particle filtering methods discussed in Section 3 apply to any model with this form.

2.1. Sequential estimation

When data are collected sequentially, it is often of interest to determine the *filtered distribution*, the distribution of the current state and parameters conditional on the data observed up to that time. This distribution describes all of the available information up to time t about the current state of the system and any fixed parameters. It can be updated recursively using Bayes' rule:

$$p(x_t,\theta|y_{1:t}) \propto p(y_t|x_t,\theta)p(x_t,\theta|y_{1:t-1}) \tag{1}$$

where $y_{1:t} = (y_1,\ldots,y_t)$. Only in special cases can $p(x_t,\theta|y_{1:t})$ be evaluated analytically, e.g. in DLMs when θ is the observation variance [Section 4.3, 24]. When analytical tractability is not present, we turn to numerical methods including deterministic versions, e.g. the extended Kalman filter and the Gaussian sum filter [26], or Monte Carlo versions such as particle filters.

3. Particle filtering

Particle filtering is an SMC inferential technique based on repeated use of importance sampling. It aims to approximate the filtered distribution at time t through a weighted Monte Carlo realization from this distribution in terms of J particles, i.e.

$$p(x_t,\theta|y_{1:t}) \approx \sum_{j=1}^{J} w_t^{(j)} \delta_{\left(x_t^{(j)},\theta^{(j)}\right)} \tag{2}$$

where $\left(x_t^{(j)},\theta^{(j)}\right)$ is the location of the j^{th} particle at time t, $w_t^{(j)}$ is the weight of that particle with $\sum_{j=1}^{J} w_t^{(j)} = 1$, and δ is the Dirac delta function. A variety of SMC techniques have been developed to provide more efficient approximations to Eq. (1) in the sense that with the same computation time a better approximation is achieved. We now review three fundamental particle filtering techniques: the bootstrap filter, auxiliary particle filter, and kernel density particle filter. In Section 5, we compare the efficiency of these techniques in the syndromic surveillance context.

3.1. Bootstrap filter

The first successful version of particle filtering is known as the bootstrap filter (BF) [18,19]. Since this method and the auxiliary particle filter were developed for the situation when θ is known, we will (for the moment) drop θ from the notation. Given an approximation to the filtered distribution at time t as in Eq. (2), to obtain an approximation to the filtered distribution at time $t+1$, perform the following steps for each particle $j = 1,\ldots,J$:

1. Resample: sample an index $k \in \{1,\ldots,j,\ldots,J\}$ with associated probabilities $\left\{w_t^{(1)},\ldots,w_t^{(j)},\ldots,w_t^{(J)}\right\}$,
2. Propagate: sample $x_{t+1}^{(j)} \sim p\left(x_{t+1}\middle|x_t^{(k)}\right)$, and
3. Calculate weights and renormalize:

$$\tilde{w}_{t+1}^{(j)} = p\left(y_{t+1}\middle|x_{t+1}^{(j)}\right) \qquad w_{t+1}^{(j)} = \tilde{w}_{t+1}^{(j)} \Big/ \sum_{l=1}^{J} \tilde{w}_{t+1}^{(l)}.$$

This procedure can be applied recursively beginning with an initial set of weights $w_0^{(j)}$ and locations $x_0^{(j)}$ for all j, usually obtained by sampling from the prior with uniform weights.

3.2. Auxiliary particle filter

One problem that arises in implementing the BF is that $w_t^{(j)}$ will be small for particles where $p\left(y_t\middle|x_t^{(j)}\right)$ is small, and these particles will contribute little to the approximation to $p(x_t|y_{1:t})$. The auxiliary particle filter (APF) aims to mitigate this by anticipating which particles will have small weight using a look ahead strategy [20]. Given an approximation to the filtered distribution at time t as in Eq. (2), the APF approximates $p(x_{t+1}|y_{1:t+1})$ by the following:

1. For each particle j, calculate a point estimate of $x_{t+1}^{(j)}$ called $\mu_{t+1}^{(j)}$, e.g.

$$\mu_{t+1}^{(j)} = E\left(x_{t+1}\middle|x_t^{(j)}\right).$$

2. Calculate auxiliary weights and renormalize:

$$\tilde{g}_{t+1}^{(j)} = w_t^{(j)} p\left(y_{t+1}\middle|\mu_{t+1}^{(j)}\right) \qquad g_{t+1}^{(j)} = \tilde{g}_{t+1}^{(j)} \Big/ \sum_{l=1}^{J} \tilde{g}_{t+1}^{(l)}.$$

3. For each particle $j = 1,\ldots,J$,
   (a) Resample: sample an index $k \in \{1,\ldots,j,\ldots,J\}$ with associated probabilities $\left\{g_{t+1}^{(1)},\ldots,g_{t+1}^{(j)},\ldots,g_{t+1}^{(J)}\right\}$,
   (b) Propagate: sample $x_{t+1}^{(j)} \sim p\left(x_{t+1}\middle|x_t^{(k)}\right)$, and
   (c) Calculate weights and renormalize:

$$\tilde{w}_{t+1}^{(j)} = \frac{p\left(y_{t+1}\middle|x_{t+1}^{(j)}\right)}{p\left(y_{t+1}\middle|\mu_{t+1}^{(k)}\right)} \qquad w_{t+1}^{(j)} = \tilde{w}_{t+1}^{(j)} \Big/ \sum_{l=1}^{J} \tilde{w}_{t+1}^{(l)}.$$

The point estimate used in Step 1 can be any point estimate, although the expectation is commonly used. Step 3 is exactly the same as the BF with appropriate modifications to the weight calculation to adjust for the 'look ahead' in steps 1 and 2. APF weights tend to be closer to uniform than BF weights, in which case a better approximation to $p(x_t|y_{1:t})$ is achieved.

The BF and the APF were constructed with the idea that all fixed parameters are known. In order to simultaneously estimate the time-evolving states and fixed parameters using either the BF or APF, it is necessary to incorporate the fixed parameters into the state with degenerate evolutions. That is, one regards the fixed parameters as elements of the state vector x_t and specifies the state evolution equation such that these elements do not change over time. Due to the possible duplication of some particles and elimination of others through resampling, the number of unique values of the fixed parameters in the particle set will decrease over time, resulting in degeneracy in the fixed parameters [21].

3.3. Kernel density particle filter

The particle filter introduced by Liu and West [21] builds on the auxiliary particle filter and provides a general way of fighting degeneracy in fixed parameters. This is done by approximating the set of fixed parameter values by a kernel density estimate and then regenerating values from this approximation. We refer to this filter as the kernel density particle filter (KDPF). This filter approximates $p(x_t,\theta|y_{1:t})$ via Eq. (2). To make the notation transparent, we introduce subscripts for our fixed parameters, e.g. $\theta_t^{(j)}$ represents the value for θ at time t for particle j. This does not imply that the true θ is dynamic, but rather that particle j can have different values for θ throughout time.

Let $\bar{\theta}_t$ and V_t be the weighted sample mean and weighted sample covariance matrix of $\theta_t^{(1)},\ldots,\theta_t^{(J)}$. The KDPF uses a tuning parameter Δ, the discount factor that takes values in $(0,1)$, and two derived quantities $h^2 = 1-((3\Delta-1)/2\Delta)^2$ and $a^2 = 1-h^2$ that determine how smooth the kernel density approximation is. Lower values of Δ result in a smoother approximation. However, the goal here is simply to jitter particles around to refresh values of the fixed parameters and reduce the chance of degeneracy, and so Δ is typically taken to be between 0.95 and 0.99 [21].

Given an approximation to the filtered distribution at time t as in Eq. (2), the KDPF provides an approximation to $p(x_{t+1},\theta|y_{1:t+1})$ by the following steps:

1. For each particle j, set $m_t^{(j)} = a\theta_t^{(j)} + (1-a)\bar{\theta}_t$ and calculate a point estimate of $x_{t+1}^{(j)}$ called $\mu_{t+1}^{(j)}$, e.g. $\mu_{t+1}^{(j)} = E\left(x_{t+1}\middle|x_t^{(j)},\theta_t^{(j)}\right)$.
2. Calculate auxiliary weights and renormalize:

$$\tilde{g}_{t+1}^{(j)} = w_t^{(j)} p\left(y_{t+1}\middle|\mu_{t+1}^{(j)},m_t^{(j)}\right) \qquad g_{t+1}^{(j)} = \tilde{g}_{t+1}^{(j)} \Big/ \sum_{l=1}^{J} \tilde{g}_{t+1}^{(l)}.$$

3. For each particle $j = 1,\ldots,J$,
   (a) Resample: sample an index $k \in \{1,\ldots,j,\ldots,J\}$ with associated probabilities $\left\{g_{t+1}^{(1)},\ldots,g_{t+1}^{(j)},\ldots,g_{t+1}^{(J)}\right\}$,
   (b) Regenerate the fixed parameters: sample $\theta_{t+1}^{(j)} \sim \mathrm{N}\left(m_t^{(k)},h^2V_t\right)$,
   (c) Propagate: sample $x_{t+1}^{(j)} \sim p\left(x_{t+1}\middle|x_t^{(k)},\theta_{t+1}^{(j)}\right)$, and
   (d) Calculate weights and renormalize:

$$\tilde{w}_{t+1}^{(j)} = \frac{p\left(y_{t+1}\middle|x_{t+1}^{(j)},\theta_{t+1}^{(j)}\right)}{p\left(y_{t+1}\middle|\mu_{t+1}^{(k)},m_t^{(k)}\right)} \qquad w_{t+1}^{(j)} = \tilde{w}_{t+1}^{(j)} \Big/ \sum_{l=1}^{J} \tilde{w}_{t+1}^{(l)}.$$

The KDPF adds the kernel density regeneration to the auxiliary particle filter. Here, we use a normal kernel, where $\mathrm{N}(\mu,\Sigma)$ represents the normal distribution with mean μ and covariance matrix Σ.

To use the KDPF with normal kernels, it is necessary to parameterize the fixed parameters so that their support is on the real line. This is not a constraint, but rather a practical implementation detail. We typically use logarithms for parameters that have positive support and the logit function for parameters in the interval $(0,1)$. A parameter ψ bounded on the interval (a,b) can first be rebounded to $(0,1)$ through $(\psi-a)/(b-a)$, and then the logit transformation can be applied. We investigate the sensitivity of the performance of the particle filters to the choice of transformation in Section 5.

3.4. Resampling

In addition to choosing which particle filter algorithm to use, successful implementation also depends on which resampling scheme to use and when to resample. Resampling is sampling (with replacement) random indices between 1 and J, where index j has probability $w^{(j)}$ of being selected. Throughout our discussion, we have explicitly used multinomial resampling, but alternative resampling schemes exist including residual, stratified, and systematic resampling [27]. Residual resampling deterministically samples $\lfloor w^{(j)}J\rfloor$ copies of particle j, for each j, and distributes the remaining $J-\sum_{j=1}^{J}\lfloor w^{(j)}J\rfloor$ particles according to a multinomial distribution with associated probabilities $(w^{(j)}J-\lfloor w^{(j)}J\rfloor)/(J-\sum_{j=1}^{J}\lfloor w^{(j)}J\rfloor)$, where $\lfloor\cdot\rfloor$ is the largest integer less than or equal "$\cdot$". Stratified resampling samples uniformly over the interval $[(j-1)/J,j/J]$, for $j=1,2,\ldots,J$, and calculates the number of copies of particle j according to the empirical cumulative distribution function of the particle indices (i.e. the "inversion method"). Finally, systematic resampling is similar to stratified resampling, except that only one uniform draw is initially sampled from $[0,1/J]$ and the remaining $J-1$ are calculated by adding $(j-1)/J$ to the sampled value prior to applying the inversion method.

Resampling is meant to rebalance the weights of the particles in order to avoid degeneracy, but this introduces additional Monte Carlo variability to the particle sample. Despite systematic resampling only requiring a single uniform draw, Randal et al. [27] show via example that it can introduce more Monte Carlo variability than the other three resampling schemes. In Section 5.5, we discuss some advantages and disadvantages of the different resampling methods when applied to our specific model of a disease outbreak and suggest the use of stratified or residual resampling.

The frequency of resampling should be reduced to balance the loss of information due to degeneracy with the loss of information due to the additional Monte Carlo variability introduced during resampling. Typically, a measure of the nonuniformity of particle weights is used to determine if resampling should be performed at a given iteration of a particle filter. The common measures are

effective sample size, coefficient of variation, and entropy. We use effective sample size [28], a value ranging between 1 and J that can be interpreted as the number of independent particle samples. An effective sample size of J corresponds to all particle weights being equal, and a value of 1 corresponds to one particle weight being 1 with the rest 0. Using this measure of nonuniformity in our runs, we set a threshold of $0.8J$, meaning that if the number of independent samples is less than 80% of the total number of particles at time t, resampling is performed at that time.

The BF, APF, and KDPF algorithms described in the previous sections were constructed under the assumption that resampling is performed at every iteration of the filter. However, in practice, we omit the resampling step in each algorithm at each time point where the effective sample size exceeds $0.8J$. If resampling is not performed, we modify the algorithm at that timepoint by (1) omitting the 'Resample' step, (2) replacing all instances of the sampled index k with the particle index j, and (3) adjusting the calculation of $\tilde{w}_{t+1}^{(j)}$ by multiplying by $w_t^{(j)}$ in the BF and $\tilde{g}_{t+1}^{(j)}$ in the APF and KDPF, i.e. the particle weights get carried over. For the KDPF, regeneration is not performed when resampling is not performed since, in this case, there is no reduction in the number of unique fixed parameter values. When resampling is not performed, the 'Regenerate the fixed parameters' step should read $\theta_{t+1}^{(j)} = \theta_t^{(j)}$.

4. Particle filtering in epidemiological models

The methodology described in Sections 2 and 3 provides a general framework applicable to a wide array of scientific questions [29–31]. As an example, we describe a state-space model of an influenza-like outbreak similar to one used in Skvortsov and Ristic [3], and we explain how it fits into the general framework outlined in Sections 2 and 3. In Section 5, we analyze the performance of the particle filter algorithms in this specific modeling situation.

4.1. Model

We incorporate a compartmental model of an epidemic in which we track the proportion of the population that is susceptible (s_t), infectious (i_t), and recovered (r_t), i.e. no longer able to be infected, at time t. Mathematically, s_t, i_t, and r_t are all nonnegative and $s_t + i_t + r_t = 1$ for all t. When monitoring an epidemic, the true s_t, i_t and r_t are unknown and regarded as hidden states of the model and the observed data are gathered via syndromic surveillance. In our state-space model of an epidemic, the observation equation specifies how the observed data depend on the state of the epidemic and the state equation describes how the epidemic evolves over time.

4.1.1. State equation

First, we describe the state equation. Let $x_t = (s_t, i_t)'$ denote the state of the epidemic at time t (by definition $r_t = 1 - s_t - i_t$ and hence r_t is not needed in the state vector). Initially, we consider a compartmental model of disease transmission that is governed by three parameters:

- β, the contact rate for the spread of illness,
- γ, the recovery time from infection (i.e. the reciprocal of the average infectious period), and
- ν, the mixing intensity of the population.

β, γ, and ν are each restricted to be nonnegative. Define $\theta = (\beta, \gamma, \nu)'$ to be the vector of unknown parameters in our model and let P be the size of the population. Then, we describe the evolution of the epidemic from time t to $t+1$ by

$$x_{t+1}|x_t, \theta \sim \mathrm{N}_\Omega(f(x_t, \theta), Q(\theta)) \tag{3}$$

where

$$f(x_t, \theta) = \begin{pmatrix} s_t - \beta i_t s_t^\nu \\ i_t + \beta i_t s_t^\nu - \gamma i_t \end{pmatrix} \quad Q(\theta) = \frac{\beta}{P^2}\begin{pmatrix} 1 & -1 \\ -1 & 1 + \gamma/\beta \end{pmatrix}$$

and $\Omega = \{(s_t, i_t) : s_t \geqslant 0, i_t \geqslant 0, s_t + i_t \leqslant 1\}$, with $\mathrm{N}_\Omega(\mu, \Sigma)$ representing the truncated normal distribution onto the set Ω with mean and covariance matrix of its corresponding untruncated normal distribution given by μ and Σ, respectively.

In Eq. (3), $Q(\theta)$ is determined by calculating the variances and covariance of s_{t+1} and i_{t+1} in the discrete time approximation of a modified SIR model with stochastic fluctuations [32–34], given by

$$s_{t+1} = s_t - \beta i_t s_t^\nu + \epsilon_\beta$$
$$i_{t+1} = i_t + \beta i_t s_t^\nu - \gamma i_t - \epsilon_\beta + \epsilon_\gamma$$

where ϵ_β and ϵ_γ are random components with $\epsilon_\beta \sim \mathrm{N}(0, \sqrt{\beta}/P)$ and $\epsilon_\gamma \sim \mathrm{N}(0, \sqrt{\gamma}/P)$. The variances of these terms come from a scaling law for stochastic fluctuations in a dynamical system generated by random contacts among the population [35,32,33,3].

The *basic reproductive number*, $R_0 = \beta/\gamma$, is the average number of people infected by one sick person in a population where everyone is susceptible [36]. If $R_0 > 1$, then an epidemic can occur. In many cases, prior information about R_0 for a specific type of infection is more readily available than prior knowledge about β or γ individually.

The mixing parameter ν describes the heterogeneity of social interactions within the population, where $\nu = 1$ corresponds to a population with homogenous mixing, i.e. an infectious person is equally likely to infect any susceptible, and $\nu = 0$ corresponds to a population with no social interaction. Values of $\nu > 1$ represent populations with heterogenous mixing, i.e. an individual is more likely to interact with some people more than others, leading to less severe epidemics than those that would occur in homogenous populations for a fixed R_0 [37,38].

4.1.2. Observation equation

The observed data are positive real numbers related to counts of emergency room visits, prescription sales, or calls to a hotline, for example, and we can observe data from these different streams/sources asynchronously in time. That is, at any time t, we can observe data from any subset of the streams (or possibly none of them). Let $y_{l,t} > 0$ represent data coming from stream l at time t, where $l = 1, 2, \ldots, L$ and $t = 1, 2, \ldots, T$. We model the log of the observations by

$$\log y_{l,t} \sim \mathrm{N}(b_l i_t^{\varsigma_l} + \eta_l, \sigma_l^2) \tag{4}$$

where b_l, ς_l, and σ_l are nonnegative constants [3] and η_l is a real number that determines the baseline level of incoming syndromic data from stream l.

The form of the mean of $\log y_{l,t}$ in Eq. (4) is derived from a simplification of the power-law relationship, described in Skvortsov and Ristic [3] and Ginsberg et al. [13], between syndromic observations and the proportion of the population that is infectious, where b_l is a multiplicative constant that depends on the syndromic data source, ς_l is the power-law exponent, and σ_l is the standard deviation term that determines the magnitude of random fluctuations in the syndromic observations from stream l. Initially, we assume $b_l, \varsigma_l, \sigma_l$, and η_l are known, as in [3], but will relax that assumption in Section 6.

Having formulated the data-generating model, we define $y_t = (y_{1,t}, \ldots, y_{L,t})'$ and specify $p(y_t|x_t, \theta)$, i.e. the likelihood of an observation y_t given x_t and θ, according to $\mathrm{LN}(\mu_t, \Sigma_t)$, where μ_t is an L-length vector with element l equal to $b_l i_t^{\varsigma_l} + \eta_l$, Σ_t is an $L \times L$ diagonal matrix with the l^{th} diagonal equal to σ_l^2, and $\mathrm{LN}(\mu, \Sigma)$ represents the log-normal distribution with mean μ and covariance matrix Σ on the log scale. Elements of y_t may be missing, in which

case the dimensions of y_t, μ_t, and Σ_t shrink by the number of missing elements. If all elements of y_t are empty (i.e. if no syndromic data are observed at time t), then $p(x_t, \theta|y_{1:t}) = p(x_{t-1}, \theta|y_{1:t-1})$.

Since $p(x_{t+1}|x_t, \theta)$ and $p(y_t|x_t, \theta)$ are nonlinear functions of x_t and θ, a closed-form solution to the posterior $p(x_t, \theta|y_{1:t})$ cannot be obtained. Thus, we use the particle filtering techniques described in Section 3 to approximate $p(x_t, \theta|y_{1:t})$ for all t.

5. Comparing particle filters

We now compare the performance of the BF, APF and KDPF using simulated data analogous to that analyzed by Skvortsov and Ristic [3] using the BF. In addition, using the KDPF, we compare the performance of bounded versus unbounded priors on the fixed parameters as well as different resampling schemes. Lastly, we discuss the role of the discount factor Δ when implementing the KDPF and compare results from the KDPF with an MCMC approach.

5.1. Simulation

Forty epidemics were simulated according to Eq. (3) for a population of size $P = 5000$ and $T = 125$ days. True values of parameters θ that we estimate (termed 'unknown parameters') were different for each simulated outbreak, determined by sampling from the log-normal prior distribution, $p(\theta)$, described in Section 5.2. For all simulations, infection was introduced in 10 people in the population at day 0 (i.e. true $i_0 = 10/5000$ and $s_0 = 4990/5000$). Among the 40 simulations, the average time at which the epidemics peak is 57 days and the average proportion of the population that has been infected by $t = 125$ is 74%. The left panel of Fig. 1 shows the evolution of s_t, i_t, and r_t for a single simulation. The evolution of i_t for the remaining 39 simulations are superimposed (light gray).

Data from randomly selected streams at each day were generated from Eq. (4). The right panel of Fig. 1 displays the observed data from the simulated epidemic shown on the left. Values of known constants for $L = 4$ streams were kept the same for each simulation and are given in Table 1 (η_l was set to 0 for all l). Values for b_l, ς_l, and σ_l were chosen to be the same as those used in the numerical study carried out by Skvortsov and Ristic [3]. The values chosen for ς_l were motivated by evidence based on real syndromic data that suggests values close to 1 [39].

Table 1
Values of known constants in model.

l	b_l	ς_l	σ_l
1	0.25	1.07	0.0012
2	0.27	1.05	0.0008
3	0.23	1.01	0.0010
4	0.29	0.98	0.0011

5.2. Particle filter runs

For each simulated data set, the BF, APF, and KDPF were run using J = 100, 1000, 10,000, and 20,000 particles to obtain weighted sample approximations to $p(x_t, \theta|y_{1:t})$ for $t = 1, \ldots, T$. For each J, separate runs using multinomial, residual, stratified, and systematic resampling were implemented [40] and an effective sample size threshold was set at 80% of the total number of particles to determine when to resample particles [28]. For the KDPF, sensitivity to changes in the discount factor Δ was explored by running with $\Delta = 0.9, 0.95, 0.96, 0.97, 0.98,$ and 0.99.

To start each particle filter run, the initial state, x_0, and fixed parameters, θ_0, for J particles were sampled from the prior density $p(x_0, \theta) = p(\theta)p(i_0, s_0)$, where $p(i_0, s_0)$ is the joint pdf of the random variables s_0 and i_0 and $p(\theta)$ is the joint prior density of β, γ, and ν. Since a very small percentage of the population is infected during the initial stage of an epidemic, we let $i_0 \sim \mathrm{N}_{[0,1]}(0.002, 0.0005^2)$ and set $s_0 = 1 - i_0$, i.e. no infected individuals have recovered from illness yet.

To investigate the impact of different prior distributions for θ on the performance of the particle filters, the runs described above were performed once using uniform priors on θ and then again using log-normal priors. We first ran the particle filter algorithms using uniform priors on θ that were chosen to be the same as those used in Skvortsov and Ristic [3], i.e. $p(\theta) = p(\beta)p(\gamma)p(\nu)$ with $\beta \sim \mathrm{Unif}(0.14, 0.50)$, $\gamma \sim \mathrm{Unif}(0.09, 0.143)$, and $\nu \sim \mathrm{Unif}(0.95, 1.3)$, where $\mathrm{Unif}(a, b)$ is the continuous uniform distribution on the interval (a, b). These priors allow for values of R_0 in a range of approximately 1–5.5 and an average infectious period in a range of roughly 7–11 days. R_0 values for strains of influenza have been estimated to be around 2–3 [41,36,42], and so while these priors impose restrictive bounds on the parameters, they are not particularly informative for tracking a flu epidemic.

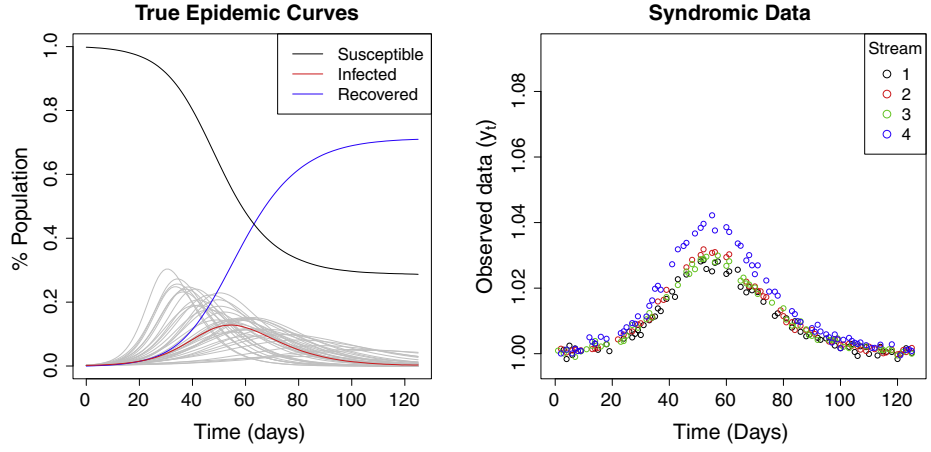

Fig. 1. Simulated epidemic curves (left) and syndromic observations (right) for a single simulated epidemic (colored lines) with $\beta = 0.254, \gamma = 0.111$, and $\nu = 1.246$ along with infectious curves, i_t, for the remaining simulations (light gray). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We then ran the particle filter algorithms using our own log-normal priors on θ that we define by $p(\theta) = p(\beta, \gamma)p(\nu)$ (i.e. β and γ are not independent). When implementing the particle filter algorithms, the prior draws for β were determined by multiplying sampled values of γ by the basic reproductive rate $R_0 = \beta/\gamma$. All parameters were sampled independently with priors $R_0 \sim \text{LN}(0.7520, 0.1768^2)$, $\gamma \sim \text{LN}(-2.1764, 0.1183^2)$, and $\nu \sim \text{LN}(0.1055, 0.0800^2)$. Here, we incorporate prior information on R_0 instead of β directly, since prior knowledge of the basic reproductive number may be easier to obtain than for the contact rate itself. These log-normal priors constrain β, γ and ν to be positive. The mean and variance of the prior distributions on $\log\gamma$ and $\log\nu$ were chosen such that random draws of γ and ν would fall within the bounds of their respective uniform priors with 95% probability, and the mean and variance of $\log R_0$ were chosen such that R_0 would fall between 1.5 and 3 with 95% probability.

It is important to note here that particle filters do not perform well when diffuse priors are placed on unknown states or fixed parameters. This is because priors that are too vague yield a small number of prior draws sampled from areas of high likelihood, resulting in degeneracy of the particle filter after only a few time points. We discuss this challenge in more detail in Section 7.

Logit and log transformations were applied to the components of θ in the manner described at the end of Section 3.3 so that the normal mixture kernel could be used in the KDPF while constraining β, γ, and ν to be within their respective prior domains (i.e. logit was used with uniform priors and log with log-normal priors).

5.3. Comparison of particle filter algorithms under uniform priors

First, we compare the performance of the particle filter algorithms using uniform priors on θ and systematic resampling, since these priors and resampling scheme were used in Skvortsov and Ristic [3]. For ease of comparison, the same prior draws were used in each particle filter for fixed J. Fig. 2 shows 95% credible bounds of $p(\beta|y_{1:t})$, $p(\gamma|y_{1:t})$, and $p(\nu|y_{1:t})$ for $t = 1, 2, \ldots, T$ and J = 100, 1000, 10,000, 20,000 using the simulated data displayed in Fig. 1. Initially, the credible bounds for the BF and APF match those of the KDPF, but quickly degenerate toward a single value due to elimination of unique particles during resampling. Although the time of degeneracy increases as J gets larger, the BF and APF bounds become misleading during the second half of the epidemic, even for J = 20,000. The bounds for the KDPF, on the other hand, have dramatically reduced degeneracy since new values of θ are regenerated from the kernel density approximation.

The KDPF also has an advantage over the BF and APF in terms of computational efficiency. Notice that the bounds for the KDPF become wider as J increases, but they do not change much between J = 10,000 and J = 20,000. This suggests that by 10,000 particles, the weighted sample approximation of $p(x_t, \theta|y_{1:t})$ has converged to the true posterior over the entire epidemic period, unlike with the BF and APF. Even though the bounds for β for the BF and APF seem to roughly match those of the KDPF for J = 20,000 over the first half of the epidemic, the KDPF provides the same measure of uncertainty for J = 10,000 and does not degenerate during the second half of the epidemic.

Estimation of ν is more challenging than for β or γ because of the nonlinear nature of the state equation with respect to this parameter. We notice from the plots for ν with $J \geqslant 10{,}000$ that very little information is gained over the course of the epidemic about this parameter relative to its uniform prior. Furthermore, the 95% credible intervals for ν expand between $t = 70$ and $t = 80$, while we typically expect to see the width of credible intervals for an unknown fixed parameter decrease monotonically over time as data is accumulated. A plausible explanation here is that $p(\nu|y_{1:t})$ has been squeezed against the lower bound of the prior for ν. Rerunning the analysis (results not shown) using our log-normal priors relaxes the prior bounds on ν and shows a shift in the distribution toward higher values around $t = 70$ as opposed to the widening of the interval that we see in Fig. 2.

Table 2 shows that the behavior of the BF, APF, and KDPF illustrated in Fig. 2 is consistent across the 40 simulations. For instance, for J = 20,000, the 95% credible intervals at $t = 125$ for each of β, γ and ν cover the truth for 39 out of 40 (97.5%) simulations using the KDPF. The BF and APF runs using the same number of particles, on the other hand, yield 95% credible intervals at $t = 125$ that cover the truth for no more than 13 out of 40 (32.5%) simulations when considering β, γ, and ν marginally.

5.4. Illustration of the negative impact of priors with truncated support

As mentioned in Section 3.3, implementing the KDPF using a normal kernel density approximation to $p(\theta|y_{1:t})$ to regenerate the fixed parameters requires applying some transformation to the components of θ so that their support is on the real line. The logit function is a convenient choice for mapping fixed parameters with bounded support to the real line; the log function is convenient for fixed parameters with positive support. Thus, we investigate the sensitivity of the results with J = 10,000 to these two types of transformations using the KDPF.

Fig. 3 compares scatterplots of β versus γ sampled jointly from the filtered distribution $p(\beta, \gamma|y_{1:t})$ at $t = 0, 20, 40, 60$. In the top row, the same truncated support prior as in Skvortsov and Ristic [3] is used, i.e. $\beta \sim \text{Unif}(0.14, 0.50)$ and $\gamma \sim \text{Unif}(0.09, 0.143)$ independently. In order to ensure regeneration in the KDPF does not extend past these bounds, a logit transformation was applied in the manner described at the end of Section 3.3 with $a = 0.14$ and $b = 0.50$ for β and $a = 0.09$ and $b = 0.143$ for γ. The kernel density is then created on this transformed space. In the top row of Fig. 3, the samples concentrate on the boundaries of the uniform prior on γ, particularly for $t = 20$ and $t = 40$. This suggests that the truncated support prior bounds on γ are too restrictive to account for the uncertainty in this recovery time.

To test this hypothesis, we reran the KDPF using exactly the same prior draws as those used in the first row of Fig. 3, but we apply a log transformation to θ (to ensure $\beta, \gamma, \nu > 0$) instead of the logit transformation. The results are shown in the second row of Fig. 3. Despite the particles starting within the uniform bounds at $t = 0$, the samples stray outside the uniform bounds for γ, suggesting that the data are informing us that reasonable parameter values can be found outside the bounds that would have been imposed by the truncated support prior.

The bottom row of Fig. 3 displays results from running the KDPF using prior samples taken from the log-normal priors on the components of θ. As in the second row, a log transformation was applied to θ so that the kernel density could be used for jittering particles. The log-normal priors on θ are more informative than the uniform priors in that a greater number of particles are concentrated near the true values of β and γ at $t = 0$. Yet, the distribution of particles at $t = 20$ and $t = 40$ appear more spread out in the bottom row than in the top row because the log-normal priors are less restrictive on the sample space of β and γ than are the truncated support uniform priors.

At $t = 60$ in the bottom row of Fig. 3, sampled particle values have moved inside the bounds that would have been imposed by the uniform prior and form an ellipse-shaped distribution similar to what is shown in the second row of Fig. 3 at $t = 60$. This suggests that the tail of points concentrated along the upper uniform bound at $t = 60$ in the top row of Fig. 3 is an artifact of the over-restrictive uniform prior and not influenced by the data. We suggest using log-normal priors on positive elements of θ as opposed to uniform priors which bound the range of possible parameter values. This allows us to use prior knowledge of the epidemic to encourage points to lie in

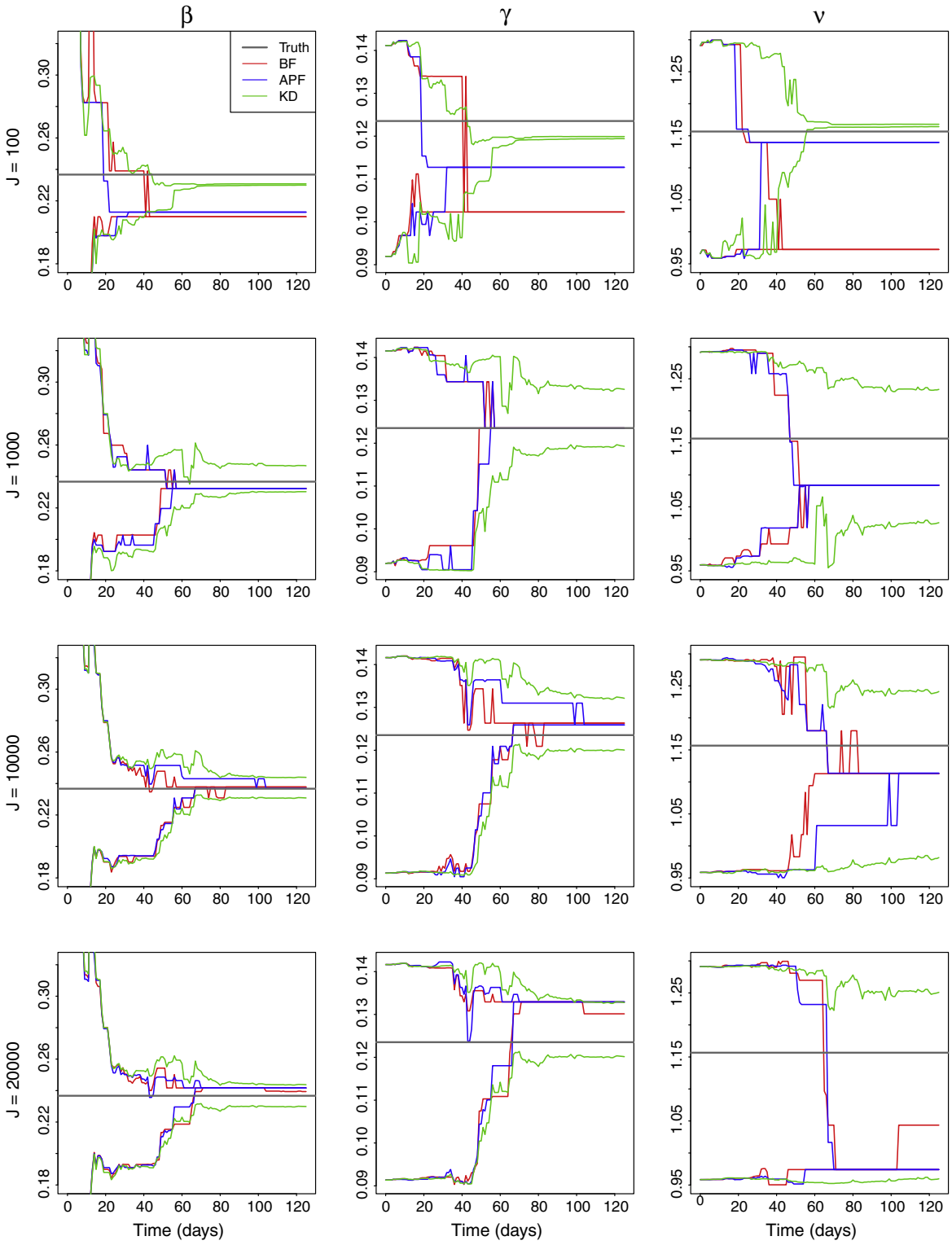

Fig. 2. Sequential 95% credible intervals for β (left column), γ (middle column), and ν (right column) for increasing number of particles (rows) for the BF (red), APF (blue), and KDPF (green), compared with the truth (black lines), when using systematic resampling and uniform priors. Data were generated from the simulated epidemic shown in Fig. 1. For the KDPF, Δ was set to 0.99. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a reasonable range while retaining flexibility in the event of model mis-specification either in the likelihood or the prior.

5.5. Comparison of resampling schemes

As mentioned in Section 3.4, resampling is meant to rebalance the weights of the particles in order to avoid degeneracy, but this comes at the cost of increasing the Monte Carlo variability of the particle sample. Up to this point, we have used only systematic resampling, as in Skvortsov and Ristic [3]. Alternatively, we could have chosen multinomial, residual, or stratified resampling. Randal et al. [27] explains each of these methods in detail and shows that (1) multinomial resampling introduces more Monte Carlo variability than do residual or stratified resampling, (2) residual and strat-

Table 2
Proportion of simulated data sets (out of 40 total) for which 95% credible intervals obtained from the marginal filtered distributions of the fixed parameters (columns) at the end of the epidemic (i.e. $t = 125$) cover the true value used for simulation for increasing number of particles (rows) using the BF, APF, and KDPF.

J	β			γ			v		
	BF	APF	KDPF	BF	APF	KDPF	BF	APF	KDPF
100	0.000	0.000	0.175	0.000	0.000	0.175	0.000	0.000	0.100
1000	0.000	0.000	0.800	0.000	0.000	0.900	0.000	0.000	0.800
10,000	0.150	0.150	0.975	0.150	0.200	0.950	0.175	0.250	0.925
20,000	0.325	0.275	0.975	0.325	0.175	0.975	0.300	0.175	0.975

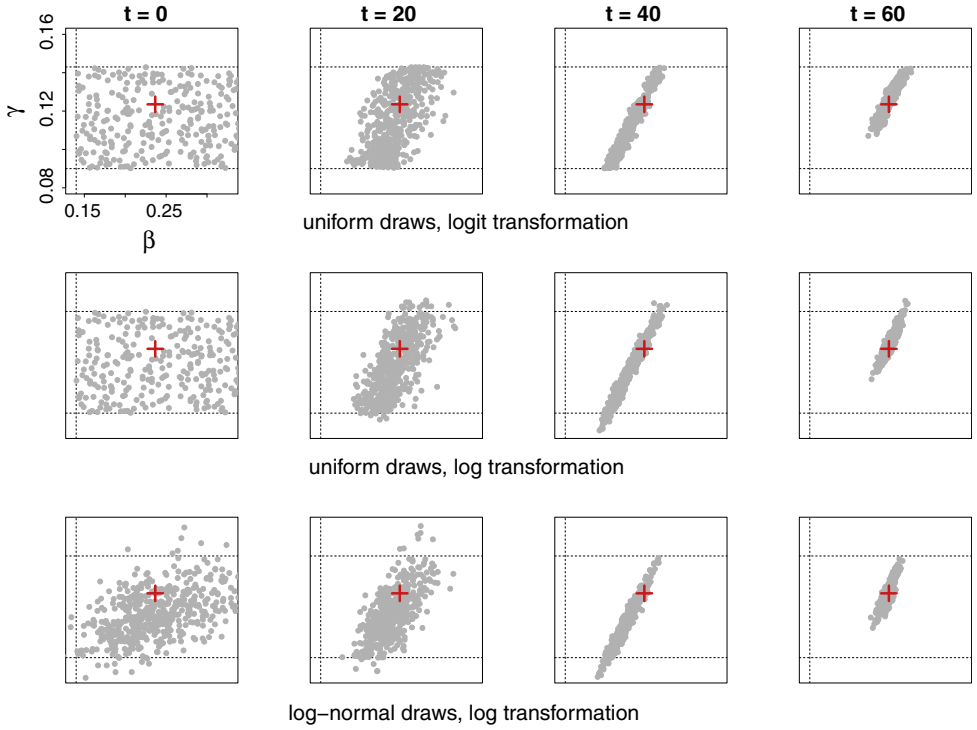

Fig. 3. Scatterplots of β (horizontal) versus γ (vertical) with true values (red crosses) at $t = 0, 20, 40, 60$ days using the KDPF with J = 10,000 particles, systematic resampling, and $\Delta = 0.99$. The logit transformation (top row) on θ shifted and scaled to $(0, 1)$ and log transformation (second row and bottom row) were used before regenerating the fixed parameters. To aid comparison, the same uniform draws of θ were sampled at $t = 0$ in each of the first two rows. Log-normal prior draws were sampled in the bottom row. For demonstration, each panel shows 500 particles sampled from the weighted sample approximation to $p(x_t, \theta|y_{1:t})$. Axes are the same in each panel. Dashed horizontal and vertical lines indicate the bounds of the uniform priors on γ and β, respectively. The upper bound on the uniform prior on β is not shown because it lies outside the range of the horizontal axis.

ified resampling introduce the same amount of Monte Carlo variability, on average, and (3) systematic resampling can introduce more Monte Carlo variability than does multinomial resampling.

With this in mind, we turn to a comparison of different techniques for the resampling step using the KDPF with log-normal priors on θ and $\Delta = 0.99$. To aid in comparison of the different resampling techniques, the same prior draws were used in all particle filter runs for fixed J. We would like to choose the resampling scheme for which the filtered distribution, $p(x_t, \theta|y_{1:t})$, approaches the true posterior the fastest as a function of the number of particles. If the filtered distributions have converged to the true posterior, then 95% credible intervals should cover the true parameter value about 95% of the time.

Using $J = 100$ particles (not shown), sequential 95% credible intervals over the second half of the epidemic for each of β, γ, and v cover the true parameter value for less than half of the 40 simulated data sets, indicating that more particles are needed to approximate the true posterior. Fig. 4 shows that coverage probabilities approach the nominal level for all four resampling techniques as J increases. Multinomial resampling, however, appears to be outperformed by the other three resampling techniques, as coverage for all three model parameters using $J = 1000$ particles dips lower during the second half of the epidemic for multinomial resampling than it does for any of the other three methods. This is also true for β with J = 10,000 particles. By increasing the number of particles to J = 20,000 (not shown), all four resampling methods yield coverage probabilities for each parameter that remain within the 95% confidence bounds around the nominal coverage level throughout the epidemic.

Although residual, stratified, and systematic resampling perform about the same with this specific model, we prefer to use either residual or stratified resampling because of an example shown in [27] where systematic resampling adds more Monte Carlo variability than any of the other three resampling schemes.

5.6. Additional notes

5.6.1. Discount factor

We recommend, based on the results from Sections 5.3, 5.4, and 5.5, that the KDPF with residual or stratified resampling and prior distributions bounded only by the support of the parameters be used in preference to other choices mentioned. With this implementation

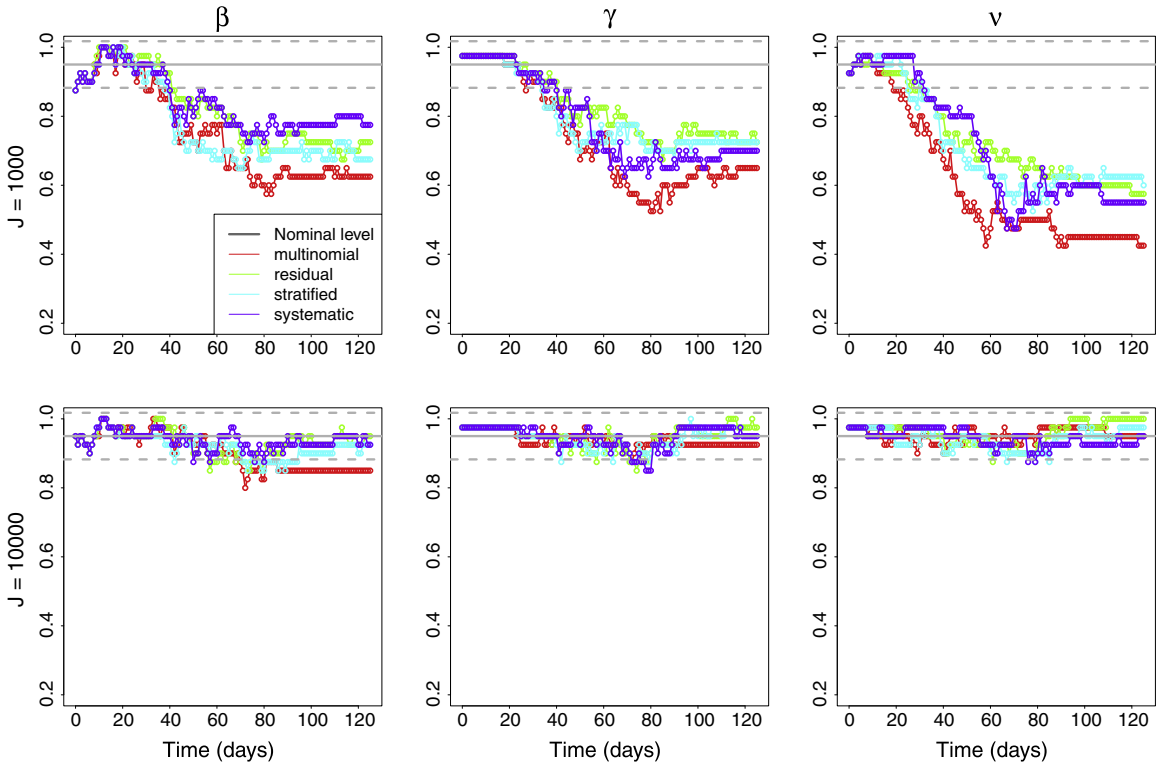

Fig. 4. Proportion of the 40 simulated data sets for which 95% credible intervals for β (left), γ (middle), and ν (right) cover the true value used for simulation for different t (x-axis) and J (rows) using the KDPF with log-normal priors on θ and $\Delta = 0.99$. Solid gray horizontal line denotes nominal coverage (95%) and dashed lines give 95% confidence bounds around the true coverage.

of the particle filter, the practitioner is still left to choose a value for the discount factor, Δ. As mentioned in Section 3.3, Δ is a tuning parameter that determines the smoothness of the kernel density approximation to $p(\theta|y_{1:t})$ when implementing the KDPF. Choosing Δ close to 0 results in a smoother approximation and more substantial jittering of particles while Δ close to 1 leads to a choppier approximation and more subtle jittering of particles.

To test the sensitivity of the KDPF to different values of Δ, we ran the KDPF with log-normal priors and stratified resampling on each of the 40 simulated data sets using different values of Δ (0.9, 0.95, 0.96, 0.97, 0.98, and 0.99) and different numbers of particles (100, 1000, 10,000, and 20,000). We then calculated 95% credible intervals for each of the unknown parameters. The results (not shown) indicate that lower values of Δ lead to a higher proportion of 95% credible intervals covering the truth, but that coverage probabilities for all values of Δ are close to the nominal level for all t if $J \geqslant 10{,}000$. We use $\Delta = 0.99$ because we seek an implementation of the KDPF that works well when enough particles are used to provide a good approximation to the true posterior. Liu and West [21] recommends choosing a value between 0.95 and 0.99.

5.6.2. Comparison with MCMC

For a comparison with our KDPF results, an MCMC analysis was run using the particle MCMC (PMCMC) approach of Andrieu et al. [43]. Using the `pmcmc` function within the R package `pomp` [44], samples from the posterior distribution of the fixed parameters were generated conditional on the first T observations from the simulated data set pictured in Fig. 1 for $T = 5, 10, 15, \ldots, 125$. For each T, three chains consisting of 30,000 PMCMC iterations were generated and 95% credible intervals were calculated for each chain. PMCMC chains were initialized at the true values of the fixed parameters used for simulating the data. Tuning parameters representing the standard deviations of the random-walk proposal distributions of β, γ, and ν were set to 0.005, 0.001, and 0.01, respectively. These values were chosen because they allowed the chains to mix well within reasonable computing time, but it is possible that different values could provide better mixing and hence improved estimates of the fixed parameters within the same computing time. For more information on PMCMC and using functions within the `pomp` package, we refer the reader to Andrieu et al. [43] and King et al. [44].

Fig. 5 compares the performance of the KDPF with PMCMC in terms of marginal 95% credible intervals for the fixed parameters. The intervals obtained from PMCMC samples for each chain at each T are compared with those produced by 20 separate runs of the KDPF on the same data set. The KDPF was run using 20,000 particles, log-normal priors on the components of θ, stratified resampling, and $\Delta = 0.99$. Multiple runs of the KDPF and PMCMC on the same data allows us to assess the uncertainty in the 95% credible intervals for the filtered distributions at each T. For instance, the high variance of the interval estimates for ν in the rightmost panel of the figure demonstrates the challenge in estimating this parameter. The performance of the KDPF compares well with PMCMC in this study, as the bounds of the 95% credible intervals obtained from the PMCMC chains over the course of the epidemic appear to be more variable than the bounds obtained from the separate KDPF runs. Furthermore, a single PMCMC chain run on the full data set (i.e. for $T = 125$) took 8–9 h to complete while the KDPF with $J = 20{,}000$ particles provided results for all time points in about 15 min in our study. Section 7 provides further discussion of different scenarios where either PMCMC or the KDPF might be preferred.

6. Additional unknown parameters

Within SMC approaches, an advantage of using a more computationally efficient algorithm is to allow reduced model assumptions. We therefore turn our focus to extending the analysis

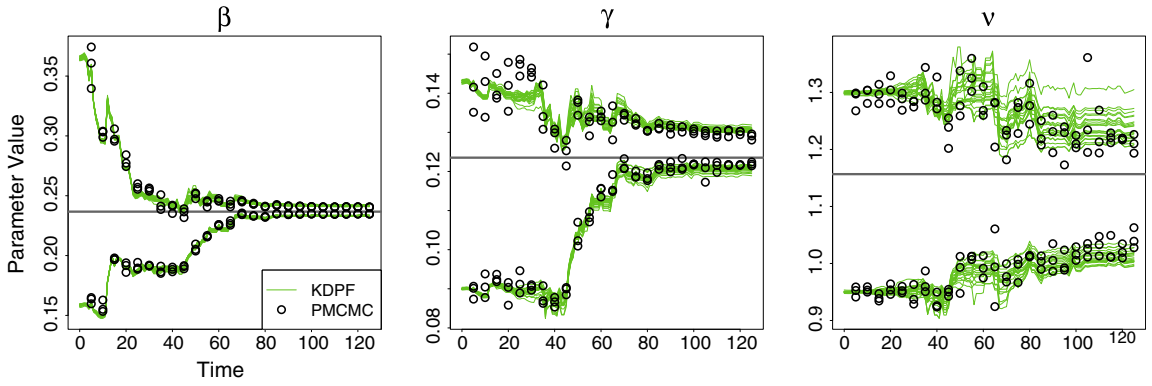

Fig. 5. Sequential 95% credible intervals for β (left), γ (middle), and ν (right) obtained from 20 different runs of the KDPF (green lines) using $J = 20{,}000$ particles, log-normal priors, stratified resampling, and $\Delta = 0.99$ compared with 95% credible intervals obtained from 3 different PMCMC chains (black circles) run for 30,000 iterations on data collected up until day T for $T = 5, 10, \ldots, 125$. All KDPF and PMCMC runs used observations taken from the same simulated data set pictured in Fig. 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

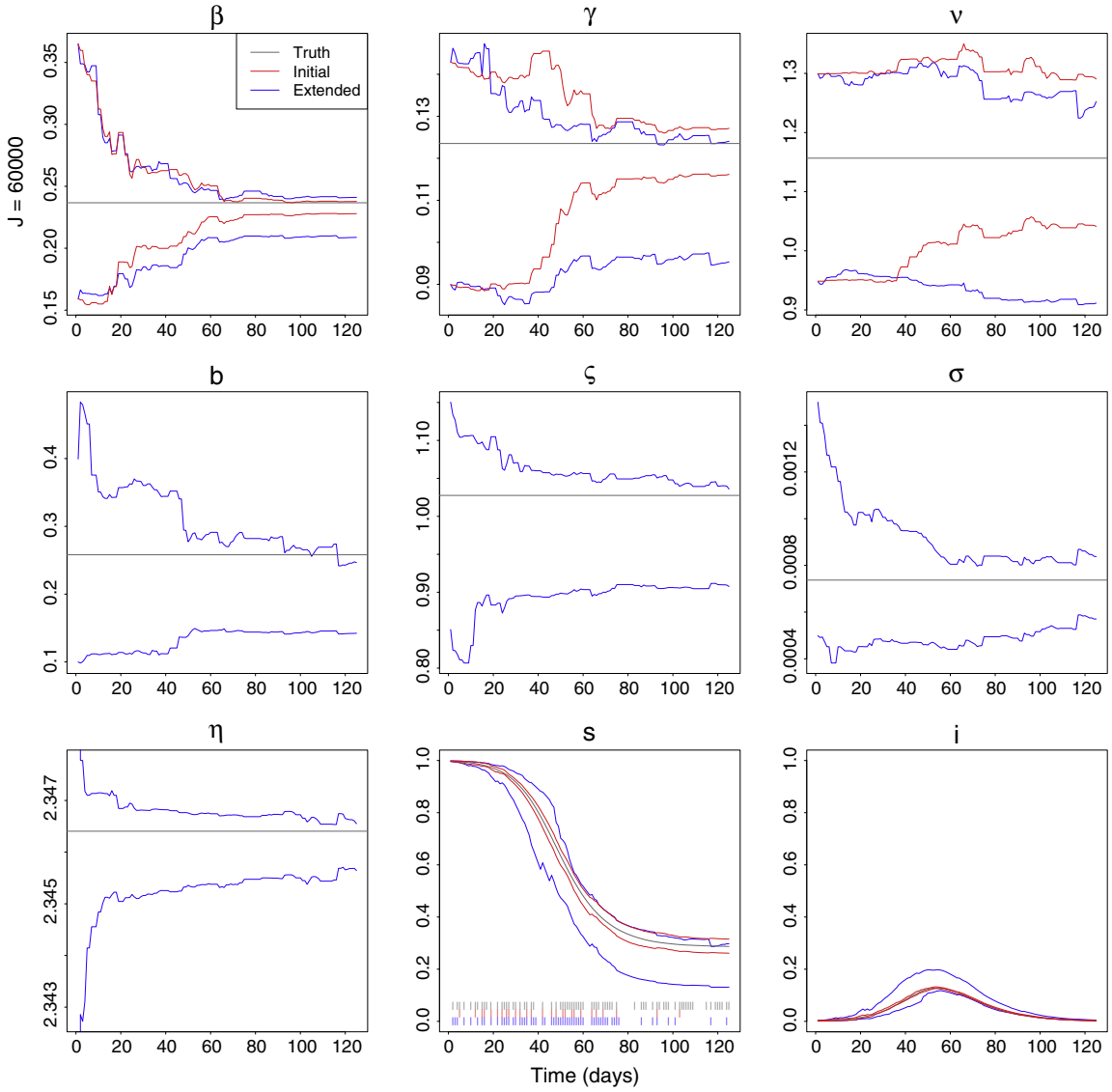

Fig. 6. Sequential 95% credible intervals for the states and fixed parameters from the original (red) and extended (blue) analyses where the KDPF with $J = 60{,}000$ particles was run with stratified resampling and $\Delta = 0.99$. Tick marks are shown along the bottom of the plot for s_t at time points when data were observed (dark gray) and when particles were resampled (blue and red for the extended and original analyses, respectively). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

using the KDPF to include $\{b_l, \varsigma_l, \sigma_l, \eta_l : l \in 1, \ldots, L\}$ as unknown parameters, thereby increasing the number of unknown parameters beyond those considered by Skvortsov and Ristic [3] using the BF. For this section, we consider data coming from only one stream ($L = 1$) and let $\theta = (\beta, \gamma, v, b, \varsigma, \sigma, \eta)'$, dropping the subscript l. Keeping the same simulated evolution of the true epidemic as shown in the left panel of Fig. 1, a single stream of syndromic observations (data y_t) was simulated from Eq. (4) with true values of b, ς, σ, and η set to $0.258, 1.028, 0.000737$ and 2.346, respectively. Days at which data were observed from the single stream were randomly selected.

The KDPF with tuning parameter Δ set to 0.99 was run with $J = 60{,}000$ particles and stratified resampling was used with an effective sample size threshold of $0.8J$. As before, fixed parameter values were regenerated only when resampling was performed. Initial particles for states and parameters were sampled from their prior with $p(x_0)p(\theta) = p(i_0, s_0)p(\beta, \gamma)p(v)p(b)p(\varsigma)p(\sigma)p(\eta)$. The prior for the state and log-normal priors for R_0, γ, and v are the same as those defined in Section 5.2. The priors for b, ς, σ, and η are $b \sim \text{LN}(-1.6090, 0.3536^2)$, $\varsigma \sim \text{LN}(-0.0114, 0.0771^2)$, $\sigma \sim \text{LN}(-7.0516, 0.2803^2)$, and $\eta \sim N(2.5, 1)$ independently. The choice of prior mean and standard deviation on the log scale were made such that random draws of b, ς, and σ on the original scale would be within $(0.1, 0.4)$, $(0.85, 1.15)$, and $(0.0005, 0.0015)$, respectively, with 95% probability. To assess the loss in precision of our estimates due to incorporating additional unknown parameters into our analysis, we compared with results from running the KDPF using 60,000 particles with b, ς, σ, and η assumed known at their true values used for simulating the data (we refer to the run with b, ς, σ, and η assumed known as the initial analysis).

Fig. 6 shows sequential 95% credible intervals for both the extended (blue lines) and the initial (red lines) analyses. Most noticeable from Fig. 6 is that the intervals for β, γ, v, s_t, and i_t are wider for the extended analysis than they are for the initial. This is due to the added uncertainty in b, ς, σ, and η in the extended analysis. Nonetheless, we are still able to obtain credible intervals for the unknown parameters that cover the true values for this simulated data set, as well as intervals for the states that cover the true epidemic curves, using a higher number of particles ($J = 60{,}000$) than was used in the initial KDPF analysis in prior sections.

In Fig. 6, the lines appear choppy or block-like. This results from data coming from only one stream, leading to more time points where no data are available and making the analysis more sensitive to abnormal data. Gaps in the data lead to a lack of resampling of particles and cause more drastic shifts in the filtered distribution once data arrive. For instance, we notice a spike in the s_t curve right after $t = 40$ because of a gap in the data and a shift in the trajectory of data points near the epidemic peak.

Lastly, we comment on a widening of the credible intervals for v. This phenomenon suggests that the log-normal priors used on v are too restrictive, and that our model provides even less insight about this parameter than our prior belief. Scarce knowledge about v is gained over the course of the epidemic in the initial analysis due to the nonlinear nature of the evolution equation with respect to v, and we in fact lose information about v in the extended analysis relative to our specified prior. While the extended analysis could be rerun with a different prior, we present this specific analysis to illustrate the sensitivity of the filtered distribution of v to assumptions about other parameters. The improved efficiency of the KDPF provides insight into this sensitivity (within reasonable computing time) in the absence of assumptions that were made about fixed parameters in the observation equation in both our initial analysis and in the prior BF analysis by Skvortsov and Ristic [3].

7. Discussion

Presented here is a strategy for simultaneous estimation of the current outbreak state and fixed parameters related to disease transmission using syndromic data. We introduce a stochastic epidemiological compartment model of a disease outbreak for data from syndromic surveillance that could possibly be multivariate and have any pattern of missingness. We suggest the use of the kernel density particle filter [21] using priors on fixed parameters that are bounded only by their support. We suggest the use of stratified or residual resampling when effective sample size has dropped markedly and that regeneration of fixed parameter values should only occur when resampling is performed. We showed how this approach is capable of estimating a model with additional unknown fixed parameters.

Advanced techniques exist that are better than the KDPF at fighting particle degeneracy, but require more practitioner input. For example, particle degeneracy could be combated within an SMC algorithm by incorporating a MCMC step to refresh fixed parameter values [45,46]. However, this would require the practitioner to define an MCMC algorithm in addition to the SMC algorithm. In addition to this requirement, the algorithm would no longer be truly sequential as the computational effort increases with time. Alternatively, if the practitioner is willing to modify their model, they can take advantage of a sufficient statistic structure [47], Rao-Blackwellization [48], or both [49]. Possible modifications to the model in Section 4.1 to allow alternative strategies include setting $v = 1$, removing fixed parameters from Q, and eliminating the truncation in Eq. (3).

The KDPF provides a sequential inferential strategy that is easy to implement, applies to a very broad class of models, and reduces particle degeneracy when applied to models with unknown fixed parameters. However, along with its methodological strengths, the algorithm has weaknesses that are reflective of all SMC methods in general. For instance, SMC methods do not perform well in high-dimensional parameter space. In addition, while particle filters perform well when run over fixed-length time intervals, they eventually degenerate if run over long periods of time due to the accumulation of approximation errors. Lastly, as mentioned in Section 5.2, all SMC methods suffer from degeneracy if vague priors are used. A common solution to this problem is to first run an MCMC based on the first few data points to find a reasonable particle cloud from which to draw prior samples [Chapter 5, 24].

In high-dimensional settings, PMCMC methods provide better estimates of unknown states and fixed parameters by using SMC methods to construct efficient proposal distributions for a joint sample of all dynamic states [43]. Since it is likely that more complicated models than what we presented in this paper may be required for monitoring disease outbreaks in real-life situations [50,51], PMCMC may offer a better solution in certain situations such as when x_t and θ are high-dimensional. However, PMCMC is a non-sequential method and only valuable for on-line analysis provided the computation time required is not too burdensome. A sequential analysis could be more valuable for processing data collected at shorter time intervals when an immediate decision regarding an intervention policy is needed [52,7,31]. In addition, efficient comparison of competing models for an epidemic outbreak [51] could be made more feasible by running an SMC algorithm that could assess the fit of the data to multiple models more quickly. SMC methods can also provide an approximation to the marginal likelihood of the data if formal model comparison or model averaging is desired [53,54]. We believe both the KDPF and PMCMC are valuable tools available to the practitioner.

In this paper, we outline a strategy for real time tracking of a disease epidemic using data from syndromic surveillance, but this

strategy can be applied to many other fields requiring on-line data analysis. We present improved particle filtering methods in general within the framework of sequential estimation of states and unknown fixed parameters in state-space models to inspire future work in epidemiological modeling and other scientific areas as well.

References

[1] W.-Y. Tan, Z. Ye, Estimation of HIV infection and incubation via state space models, Math. Biosci. 167 (2000) 31–50.
[2] L.-J. Kao, An application of a two-level non-Gaussian state-space model in the analysis of longitudinal papilloma count data, Math. Biosci. 199 (2006) 121–140.
[3] A. Skvortsov, B. Ristic, Monitoring and prediction of an epidemic outbreak using syndromic observations, Math. Biosci. 240 (2012) 12–19.
[4] M.M. Saito, S. Imoto, R. Yamaguchi, H. Sato, H. Nakada, M. Kami, S. Miyano, T. Higuchi, Extension and verification of the SEIR model on the 2009 influenza A (H1N1) pandemic in Japan, Math. Biosci. 246 (2013) 47–54.
[5] Martínez-Beneito, D. Conesa, A. López-Quílez, A. López-Maside, Bayesian Markov switching models for the early detection of influenza epidemics, Stat. Med. 27 (2008) 4455–4468.
[6] D. Merl, L.R. Johnson, R.B. Gramacy, M. Mangel, A statistical framework for the adaptive management of epidemiological interventions, PloS One 4 (2009) e5807.
[7] M. Ludkovski, J. Niemi, Optimal dynamic policies for influenza management, Stat. Commun. Infect. Dis. 2 (2010).
[8] S. Unkel, C. Farrington, P.H. Garthwaite, C. Robertson, N. Andrews, Statistical methods for the prospective detection of infectious disease outbreaks: a review, J. R. Stat. Soc.: Ser. A (Stat. Soc.) 175 (2012) 49–82.
[9] K.J. Henning, Overview of syndromic surveillance. What is syndromic surveillance, MMWR Morb Mortal Wkly Rep 53 (Suppl) (2004) 5–11.
[10] M. Wagner, A. Moore, R. Aryel, Handbook of Biosurveillance, Elsevier, 2006.
[11] A.G. Wilson, G.D. Wilson, D.H. Olwell, Statistical Methods in Counterterrorism: Game Theory Modeling Syndromic Surveillance and Biometric Authentication, Springer, 2006.
[12] A.M. Hakenewerth, A.E. Waller, A.I. Ising, J.E. Tintinalli, North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) and the National Hospital Ambulatory Medical Care Survey (NHAMCS): comparison of emergency department data, Acad. Emerg. Med. 16 (2009) 261–269.
[13] J. Ginsberg, M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, L. Brilliant, Detecting influenza epidemics using search engine query data, Nature 457 (2009) 1012–1014.
[14] D. Neill, A. Moore, G. Cooper, A Bayesian spatial scan statistic, in: Y. Weiss, B. Schölkopf, J. Platt (Eds.), Advances in Neural Information Processing Systems, vol. 18, MIT Press, Cambridge, MA, 2006, pp. 1003–1010.
[15] A.E. Gelfand, A.F.M. Smith, Sampling-based approaches to calculating marginal densities, J. Am. Stat. Assoc. 85 (1990) 398–409.
[16] A. Doucet, N. De Freitas, N. Gordon, Sequential Monte Carlo Methods in Practice, Springer-Verlag, New York, 2001.
[17] O. Cappé, S.J. Godsill, E. Moulines, An overview of existing methods and recent advances in sequential Monte Carlo, Proc. IEEE 95 (2007) 899–924.
[18] N.J. Gordon, D.J. Salmond, A.F.M. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proc. Part F: Commun. Radar Signal Process. 140 (1993) 107–113.
[19] G. Kitagawa, Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, J. Comput. Graph. Stat. 5 (1996) 1–25.
[20] M.K. Pitt, N. Shephard, Filtering via simulation: auxiliary particle filters, J. Am. Stat. Assoc. 94 (1999) 590–599.
[21] J. Liu, M. West, Combined parameter and state estimation in simulation-based filtering, in: A. Doucet, J.F.G. De Freitas, N.J. Gordon (Eds.), Sequential Monte Carlo Methods in Practice, Springer-Verlag, New York, 2001, pp. 197–217.
[22] R.E. Watkins, S. Eagleson, B. Veenendaal, G. Wright, A.J. Plant, Disease surveillance using a hidden Markov model, BMC Med. Infor. Decis. Making 9 (2009) 39, http://dx.doi.org/10.1186/1472-6947-9-39.
[23] O. Cappé, E. Moulines, T. Rydén, Inference in Hidden Markov Models, Springer Science + Business Media, 2005.
[24] G. Petris, S. Petrone, P. Campagnoli, Dynamic Linear Models, Springer, 2009.
[25] M. West, J. Harrison, Bayesian Forecasting and Dynamic Models, 2nd ed., Springer-Verlag Inc, New York, 1997.
[26] D.L. Alspach, H.W. Sorenson, Nonlinear Bayesian estimation using Gaussian sum approximations, IEEE Trans. Autom. Control AC-17 (1972) 439–448.
[27] D. Randal, O. Cappé, E. Moulines, Comparison of resampling schemes for particle filtering, in: Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005, pp. 64–69.
[28] J.S. Liu, R. Chen, W.H. Wong, Rejection control and sequential importance sampling, J. Am. Stat. Assoc. 93 (1998) 1022–1031.
[29] J. Carpenter, P. Clifford, P. Fearnhead, An improved particle filter for non-linear problems, IEE Proc., Radar Sonar Navigation 146 (1999) 2–7.
[30] C.M. Carvalho, H.F. Lopes, Simulation-based sequential analysis of Markov switching stochastic volatility models 51 (2007) 4526–4542.
[31] V. Dukic, H.F. Lopes, N.G. Polson, Tracking epidemics with google flu trends data and a state-space seir model, J. Am. Stat. Assoc. 107 (2012) 1410–1426.
[32] O.A. van Herwaarden, J. Grasman, Stochastic epidemics: major outbreaks and the duration of the endemic period, J. Math. Biol. 33 (1995) 581–601.
[33] C.E. Dangerfield, J.V. Ross, M.J. Keeling, Integrating stochasticity and network structure into an epidemic model, J. R. Soc. Interface 6 (2009) 761–774.
[34] R.M. Anderson, C. Fraser, A.C. Ghani, Epidemiology transmission dynamics and control of SARS: the 2002–2003 epidemic, Philos. Trans. R. Soc. B Biol. Sci. 359 (2004) 1091–1104.
[35] O. Ovaskainen, B. Meerson, Stochastic models of population extinction, Trends Ecol. Evol. 25 (2010) 643–652.
[36] J.M. Heffernan, R.J. Smith, L.M. Wahl, Perspectives on the basic reproductive ratio, J. R. Soc. Interface 2 (2005) 281–293.
[37] P.D. Stroud, S.J. Sydroiak, J.M. Riese, J.P. Smith, S.M. Mniszewski, P.R. Romero, Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing, Math. Biosci. 203 (2006) 301–318.
[38] A.S. Novozhilov, On the spread of epidemics in a closed heterogeneous population, Math. Biosci. 215 (2008) 177–185.
[39] C. Chew, G. Eysenbach, Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak, PLoS One 5 (2010) e14118.
[40] A.J. Niemi, smcUtils: Utility Functions for Sequential Monte Carlo, 2012. <https://github.com/jarad/smcUtils> R package version 0.2.2.
[41] C.E. Mills, J.M. Robins, M. Lipsitch, Transmissibility of 1918 pandemic influenza, Nature 432 (2004) 904–906.
[42] S. Zhang, Estimating transmissibility of seasonal influenza virus by surveillance data, J. Data Sci. 9 (2011) 44–64.
[43] C. Andrieu, A. Doucet, R. Holenstein, Particle Markov chain Monte Carlo methods, J. R. Stat. Soc., Ser. B: Meth. 72 (2010) 269–342.
[44] A.A. King, E.L. Ionides, C. Breto, S.P. Ellner, M.J. Ferrari, B.E. Kendall, M. Lavine, D. Nguyen, D.C. Reuman, H. Wearing, S.N. Wood, Statistical Inference for Partially Observed Markov Processes, 2014. <http://pomp.r-forge.r-project.org> r package version 0.49-2.
[45] W.R. Gilks, C. Berzuini, Following a moving target: Monte Carlo inference for dynamic Bayesian models, J. R. Stat. Soc., B 63 (2001) 127–146.
[46] G. Storvik, Particle filters in state space models with the presence of unknown static parameters, IEEE Trans. Signal Process. 50 (2002) 281–289.
[47] P. Fearnhead, Markov chain Monte Carlo, sufficient statistics, and particle filters, J. Comput. Graph. Stat. 11 (2002) 848–862.
[48] A. Doucet, S. Godsill, C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, Stat. Comput. 10 (2000) 197–208.
[49] C. Carvalho, M. Johannes, H. Lopes, N. Polson, Particle learning and smoothing, Stat. Sci. 25 (2010) 88–106.
[50] J. Shaman, A. Karspeck, Forecasting seasonal outbreaks of influenza, Proc. Nat. Acad. Sci. 109 (2012) 20425–20430.
[51] A. Bhadra, E.L. Ionides, K. Laneri, M. Pascual, M. Bouma, R.C. Dhiman, Malaria in northwest india: data analysis via partially observed stochastic differential equation models driven by Levy noise, J. Am. Stat. Assoc. 106 (2011) 440–451.
[52] D. Merl, L. Johnson, R. Gramacy, M. Mangel, A statistical framework for the adaptive management of epidemiological interventions, PLoS One 4 (2009) e5807.
[53] A. Doucet, A.M. Johansen, A tutorial on particle filtering and smoothing: fifteen years later, Handbook Nonlinear Filtering 12 (2009) 656–704.
[54] Y. Zhou, A.M. Johansen, J.A.D. Aston, Towards automatic model Comparison: an adaptive sequential Monte Carlo approach. CRiSM Technical Report, University of Warwick, 13-04. Also available as ArXiv manuscript 1303.3123, 2013.