Inference in Stochastic Epidemic Models via Multinomial Approximations

Nick Whiteley
School of Mathematics, University of Bristol
and the Alan Turing Institute

Lorenzo Rimella
School of Mathematics, University of Bristol
and the Alan Turing Institute

Abstract

We introduce a new method for inference in stochastic epidemic models which uses recursive multinomial approximations to integrate over unobserved variables and thus circumvent likelihood intractability. The method is applicable to a class of discrete-time, finite-population compartmental models with partial, randomly under-reported or missing count observations. In contrast to state-of-the-art alternatives such as Approximate Bayesian Computation techniques, no forward simulation of the model is required and there are no tuning parameters. Evaluating the approximate marginal likelihood of model parameters is achieved through a computationally simple filtering recursion. The accuracy of the approximation is demonstrated through analysis of real and simulated data using a model of the 1995 Ebola outbreak in the Democratic Republic of Congo. We show how the method can be embedded within a Sequential Monte Carlo approach to estimating the time-varying reproduction number of COVID-19 in Wuhan, China, recently published by Kucharski et al. (2020).

1 Introduction

Compartmental models are used for predicting the scale and duration of epidemics, estimating epidemiological parameters such as reproduction numbers, and guiding outbreak control measures (Brauer, 2008; O'Neill, 2010; Kucharski et al., 2020). They are increasingly important because they allow joint modelling of disease

dynamics and multimodal data, such as medical test results, cell phone and transport flow data (Rubrichi et al., 2018; Wu et al., 2020), census and demographic information (Prem et al., 2020). However, statistical inference in stochastic variants of compartmental models is a major computational challenge (Bretó, 2018). The likelihood function for model parameters is usually intractable because it involves summation over a prohibitively large number of configurations of latent variables representing counts of subpopulations in disease states which cannot be observed directly.

This has lead to the recent development of sophisticated computational methods for approximate inference involving various forms of stochastic simulation (Funk and King, 2020). Examples include Approximate Bayesian Computation (ABC) (Kypraios et al., 2017; McKinley et al., 2018; Brown et al., 2018, 2016), Data Augmentation Markov Chain Monte Carlo (MCMC) (Lekone and Finkenstädt, 2006), Particle Filters (Murray et al., 2018), Iterated Filtering (Stocks, 2019), and Synthetic Likelihood (Fasiolo et al., 2016). These methods continue to have real public health impact, for example the ABSEIR R package of Brown et al. (2018) features in the current UK COVID-19 surveillance protocols (de Lusignan et al., 2020). However the intricacy of these methods, their substantial computational cost arising from use of stochastic simulation, and their dependence on tuning parameters are obstacles to their wider use and scalability. The present work addresses the challenge of finding alternative inference techniques which are computationally simple and easy to use.

Contributions We introduce a new approach to inference in compartmental epidemic models which:

- applies to a class of finite population, partially observed, discrete-time, stochastic models. In contrast to ODE models, these models can account for statistical variability in disease dynamics;

- allows approximate evaluation of the likelihood function for model parameters, and filtering and smooth-

ing for compartment occupation numbers, without any stochastic simulation or algorithm tuning parameters, in contrast to state-of-the-art techniques such as ABC;

- revolves around a computationally simple filtering recursion. The resulting likelihood and smoothing approximations can be combined with e.g., MCMC or Expectation Maximization techniques for parameter estimation;

- is shown to recover ground truth parameter values from synthetic data, and to compare favourably against Data Augmentation MCMC (Lekone and Finkenstädt, 2006), ABC using the ABSEIR R package (Brown et al., 2018) and ODE (Chowell et al., 2004) alternatives analyzing real Ebola outbreak data under a model from Lekone and Finkenstädt (2006);

- is used to extend a method of Kucharski et al. (2020) for estimating the time-varying reproduction number of COVID-19 in Wuhan, China, from an ODE compartmental model to a stochastic model.

2 Preliminaries

2.1 Difficulties of inference in stochastic compartmental models

We use the well-known Susceptible-Exposed-Infective-Recovered (SEIR) model as a simple running example. The new methods we propose are applied to more realistic and complex models in section 5.

SEIR example. The discrete-time stochastic SEIR model is:

$$
\begin{aligned}
S_{t+1} &= S_t - B_t, \\
E_{t+1} &= E_t + B_t - C_t, \\
I_{t+1} &= I_t + C_t - D_t, \\
R_{t+1} &= R_t + D_t,
\end{aligned}
\tag{1}
$$

with conditionally independent, binomially-distributed random variables:

$$
\begin{aligned}
B_t &\sim \mathrm{Bin}(S_t, 1 - e^{-h\beta I_t/n}) \\
C_t &\sim \mathrm{Bin}(E_t, 1 - e^{-h\rho}), \\
D_t &\sim \mathrm{Bin}(I_t, 1 - e^{-h\gamma}),
\end{aligned}
\tag{2}
$$

where $h > 0$ is a time-step size, β, ρ, γ are model parameters, and the process is initialized with nonnegative integers in each of the compartments (S_0, E_0, I_0, R_0) such that $S_0 + E_0 + I_0 + R_0 = n$ and n is the total population size. The interpretation of β is the rate at which an interaction between a susceptible individual and the infective proportion of the population results in

the disease being passed to that individual. The mean exposure and infective periods are respectively $1/\rho$ and $1/\gamma$. The sequence $(S_t, E_t, I_t, R_t)_{t \geq 0}$ is a Markov chain.

In practice, one typically observes times series of count data associated with some subset of the compartments, perhaps subject to random error, or under-reporting. Given such data, evaluating the likelihood function of the model parameters and initial condition requires the variables associated with unobserved compartments to be marginalized out. In general this is infeasible for models with anything but a small population size n and a small number of compartments.

Stochastic compartmental models also commonly arise in the form of continuous-time pure jump Markov processes, in which transitions of individuals between compartments occur in an asynchronous manner (Bretó, 2018). Likelihood-based inference for such processes is similarly intractable in general. There are rigorous limit theorems which link continuous time Markov process compartmental models to deterministic ODE models in the large population limit, e.g., Kurtz (1970, 1971); Roberts et al. (2015). However the precise nature of the asymptotic is somewhat subtle and not always meaningful in practice: the supplementary materials includes a simple example in which a stochastic model exhibits substantial statistical variation even when the population size is 10^7, and the corresponding ODE limit is pathological. Thus, ODE models are no substitute for stochastic models.

2.2 Notation

In the remainder of the paper, bold upper-case and bold lower-case characters are respectively matrices and column vectors, e.g., $\mathbf{A}$ and $\mathbf{b}$, with generic elements $a^{(i,j)}$ and $b^{(i)}$. The length-m column vector of 1's is denoted $\mathbf{1}_m$. A vector is called a probability vector if its elements are nonnegative and sum to 1. A matrix is called row-stochastic if its elements are nonnegative and its row sums are all 1. The indicator function is denoted $\mathbb{I}[\cdot]$. The element-wise product of matrices is denoted $\mathbf{A} \circ \mathbf{B}$ and the outer product of vectors is denoted $\mathbf{a} \otimes \mathbf{b}$. Element-wise natural logarithm and factorial are denoted $\log \mathbf{A}$ and $\mathbf{A}!$. For positive integers m and n, define $\mathcal{C}_m := \{1, \ldots, m\}$ and $\mathcal{S}_{m,n} := \{\mathbf{x} = [x^{(1)} \cdots x^{(m)}]^{\mathrm{T}} : x^{(i)} \geq 0, i = 1, \ldots, m; \sum_{i=1}^{m} x^{(i)} = n\}$. For $\mathbf{x} \in \mathcal{S}_{m,n}$, define $\boldsymbol{\eta}(\mathbf{x}) := [x^{(1)}/n \cdots x^{(m)}/n]^{\mathrm{T}}$. For a matrix $\mathbf{P}$ (resp. a vector $\boldsymbol{\pi}$) with nonnegative elements summing to 1, $\mathrm{Mult}(n, \mathbf{P})$ (resp. $\mathrm{Mult}(n, \boldsymbol{\pi})$) denotes the distribution of the random matrix (resp. vector) whose elements are the incidence counts obtained from sampling n times with replacement according to $\mathbf{P}$ (resp. $\boldsymbol{\pi}$). This is the usual definition of a multinomial distribution.

3 Model

3.1 A class of compartmental models specified by the transition probabilities of individuals

The general model we consider is defined by: m, the number of compartments; n, the total population size; a length-m probability vector $\boldsymbol{\pi}_0$; and for each $t \geq 1$, a mapping from length-m probability vectors to $m \times m$ row-stochastic matrices, $\boldsymbol{\eta} \mapsto \mathbf{K}_{t,\boldsymbol{\eta}}$. The population at time $t \geq 0$ is a set of n random variables $\{\xi_t^{(1)}, \ldots, \xi_t^{(n)}\}$, each valued in $\mathcal{C}_m$. The counts of individuals in each of the m compartments at time t are collected in a vector $\mathbf{x}_t = [x_t^{(1)} \cdots x_t^{(m)}]^{\mathrm{T}} \in \mathcal{S}_{m,n}$, $x_t^{(i)} := \sum_{j=1}^{n} \mathbb{I}[\xi_t^{(j)} = i]$. For $t \geq 1$ let $\mathbf{Z}_t$ be the $m \times m$ matrix with elements $z_t^{(i,j)} := \sum_{k=1}^{n} \mathbb{I}[\xi_{t-1}^{(k)} = i, \xi_t^{(k)} = j]$, which counts the individuals transitioning from compartment i at $t-1$ to j at t.

The sequence $\{\xi_t^{(1)}, \ldots, \xi_t^{(n)}\}_{t \geq 0}$ is constructed to be a Markov chain: the members of the initial population $\{\xi_0^{(1)}, \ldots, \xi_0^{(n)}\}$ are i.i.d. with $p(\xi_0^{(i)} = j) = \boldsymbol{\pi}_0^{(j)}$, and given $\{\xi_{t-1}^{(1)}, \ldots, \xi_{t-1}^{(n)}\}$, $\{\xi_t^{(1)}, \ldots, \xi_t^{(n)}\}$ are conditionally independent, with $\xi_t^{(i)}$ drawn from the $\xi_{t-1}^{(i)}$'th row of $\mathbf{K}_{t,\boldsymbol{\eta}(\mathbf{x}_{t-1})}$. It follows from this prescription that the sequence of matrices $(\mathbf{Z}_t)_{t \geq 0}$ is also a Markov chain. Indeed, conditional on $\mathbf{Z}_{t-1}$, and hence automatically on $\mathbf{x}_{t-1}$ since $\mathbf{Z}_t \mathbf{1}_m = \mathbf{x}_{t-1}$, the rows of $\mathbf{Z}_t$ are independent, and the distribution of the ith row of $\mathbf{Z}_t$ is $\mathrm{Mult}(x_{t-1}^{(i)}, \mathbf{K}_{t,\boldsymbol{\eta}(\mathbf{x}_{t-1})}^{(i,\cdot)})$, where $\mathbf{K}_{t,\boldsymbol{\eta}(\mathbf{x}_{t-1})}^{(i,\cdot)}$ is the ith row of $\mathbf{K}_{t,\boldsymbol{\eta}(\mathbf{x}_{t-1})}$. Moreover, noting $\mathbf{1}_m^{\mathrm{T}} \mathbf{Z}_t = \mathbf{x}_t^{\mathrm{T}}$, we observe $(\mathbf{x}_t)_{t \geq 0}$ is also a Markov chain, but we shall not need an explicit formula for its transition probabilities.

SEIR example The SEIR model in (1)-(2) is equivalent to taking $m = 4$,

$$(\mathbf{K}_{t,\boldsymbol{\eta}})^{(i,j)} = \begin{cases} e^{-h\beta\eta^{(3)}} & i = j = 1 \\ 1 - e^{-h\beta\eta^{(3)}} & i = 1 \text{ and } j = 2 \\ e^{-h\rho} & i = j = 2 \\ 1 - e^{-h\rho} & i = 2 \text{ and } j = 3 \\ e^{-h\gamma} & i = j = 3 \\ 1 - e^{-h\gamma} & i = 3 \text{ and } j = 4 \\ 1 & i = j = 4 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

for all $t \geq 1$, and identifying $[x_t^{(1)} \; x_t^{(2)} \; x_t^{(3)} \; x_t^{(4)}]^{\mathrm{T}}$ with respectively the counts of susceptible, exposed, infective and recovered individuals at time t.

We consider two observation models.

3.2 Observations derived from $(\mathbf{x}_t)_{t \geq 1}$

In this scenario, the observation at time $t \geq 1$ is a length-m vector $\mathbf{y}_t$ with elements $y_t^{(i)}$ which are conditionally independent given $\mathbf{x}_t$, and:

$$y_t^{(i)} \sim \mathrm{Bin}(x_t^{(i)}, q_t^{(i)}). \quad (4)$$

We shall collect the parameters $q_t^{(i)} \in [0, 1]$ in a length-m vector $\mathbf{q}_t$. When conducting likelihood-based inference for $\mathbf{x}_t$ using this model, if $y_t^{(i)}$ is a missing observation, then in the likelihood function associated with (4) one should take $y_t^{(i)}$ to be 0, set $q_t^{(i)} = 0$.

3.3 Observations derived from $(\mathbf{Z}_t)_{t \geq 1}$

In this scenario, the observation at time $t \geq 1$ is a $m \times m$ matrix $\mathbf{Y}_t$ with elements $y_t^{(i,j)}$ which are conditionally independent given $\mathbf{Z}_t$, and:

$$y_t^{(i,j)} \sim \mathrm{Bin}(z_t^{(i,j)}, q_t^{(i,j)}). \quad (5)$$

The parameters $q_t^{(i,j)} \in [0, 1]$ from (5) are collected into a $m \times m$ matrix $\mathbf{Q}_t$. Missing data are handled by putting a 0 in place of the missing $y_t^{(i,j)}$ and setting $q_t^{(i,j)} = 0$.

SEIR example In practice, one typically observes, at each time step, counts of *new* infectives rather than the total number of infectives, subject to some random under-reporting or missing data. How can such data be represented in the model? Due to the definition of $\mathbf{Z}_t$, the number of new infectives at time t is exactly $z_t^{(2,3)}$, since the only way an individual can transition to being infective (compartment 3) is by first being exposed (compartment 2). Therefore in this case $y_t^{(2,3)}$ following (5) is a count of newly infectives at time t, subject to binomial random under-reporting with parameter $q_t^{(2,3)}$, as required.

4 Inference

We now introduce our methods for approximating the so-called filtering distributions and marginal likelihoods $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, $p(\mathbf{y}_{1:t})$ and $p(\mathbf{Z}_t|\mathbf{Y}_{1:t})$, $p(\mathbf{Y}_{1:t})$ under respectively the observation models of sections 3.2 and 3.3. These quantities are at the core of smoothing and parameter estimation techniques demonstrated in section 5 and detailed in the supplementary materials.

For the observation model of section 3.2, note that $(\mathbf{x}_t, \mathbf{y}_t)_{t \geq 1}$ is a hidden Markov model, and in principle the filtering distributions can be computed through a two-step recursion:

$$p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) \overset{\text{prediction}}{\longrightarrow} p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \overset{\text{update}}{\longrightarrow} p(\mathbf{x}_t|\mathbf{y}_{1:t}). \quad (6)$$

However in practice, the summations involved in the 'prediction' and 'update' operations are prohibitively expensive since they involve summing over all possible values of $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$.

4.1 Approximating the prediction operation

For each $\mathbf{x} = [x^{(1)} \cdots x^{(m)}]^{\mathrm{T}} \in \mathcal{S}_{m,n}$ and length-m probability vector $\boldsymbol{\eta}$, let $M_t(\mathbf{x}, \boldsymbol{\eta}, \cdot)$ be the probability mass function on $\mathcal{S}_{m,n}$ of $(\mathbf{1}_m^{\mathrm{T}}\mathbf{Z})^{\mathrm{T}}$, where $\mathbf{Z}$ is a random $m \times m$ matrix whose rows are independent, and whose ith row has distribution $\mathrm{Mult}(x^{(i)}, \mathbf{K}_{t,\boldsymbol{\eta}}^{(i,\cdot)})$. So by construction $M_t(\mathbf{x}_{t-1}, \boldsymbol{\eta}(\mathbf{x}_{t-1}), \mathbf{x}_t)$ is the probability transition function for the Markov chain $(\mathbf{x}_t)_{t \geq 0}$ defined in section 3.1. Thus the prediction operation in (6) can be written in terms of M_t:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \sum_{\mathbf{x}_{t-1} \in \mathcal{S}_{m,n}} p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) p(\mathbf{x}_t|\mathbf{x}_{t-1})$$
$$= \sum_{\mathbf{x}_{t-1} \in \mathcal{S}_{m,n}} p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) M_t(\mathbf{x}_{t-1}, \boldsymbol{\eta}(\mathbf{x}_{t-1}), \mathbf{x}_t).$$
$$(7)$$

Our approximation to this operation is as follows: assuming we have already obtained a multinomial distribution approximation to $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$, then in (7) we replace $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ by this multinomial distribution, and replace the vector $\boldsymbol{\eta}(\mathbf{x}_{t-1})$ by its expectation under this multinomial distribution. This results in a multinomial distribution approximation to $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$. The following lemma formalizes this recipe.

Lemma 1. *If for a given length-m probability vector $\boldsymbol{\pi}$, $\mu(\cdot)$ is the probability mass function on $\mathcal{S}_{m,n}$ associated with $\mathrm{Mult}(n, \boldsymbol{\pi})$ and $\mathbb{E}_\mu[\boldsymbol{\eta}(\mathbf{x})]$ is the expected value of $\boldsymbol{\eta}(\mathbf{x})$ when $\mathbf{x} \sim \mu$, then $\sum_{\mathbf{x} \in \mathcal{S}_{m,n}} \mu(\mathbf{x}) M_t(\mathbf{x}, \mathbb{E}_\mu[\boldsymbol{\eta}(\mathbf{x})], \cdot)$ is the probability mass function associated with $\mathrm{Mult}(n, \boldsymbol{\pi}^{\mathrm{T}}\mathbf{K}_{t,\boldsymbol{\pi}})$.*

The proof is given in the supplementary materials.

4.2 Approximating the update operation

The update operation in (6) is:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})},$$
$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \sum_{\mathbf{x}_t \in \mathcal{S}_{m,n}} p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t-1}),$$
$$(8)$$

which has the interpretation of a Bayes' rule update applied to $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$. Assuming we have already obtained a multinomial distribution approximation to $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$, our approximation to the update operation is to substitute this multinomial distribution in place of $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ in (8), resulting in a shifted-multinomial distribution whose mean vector is used

to define a multinomial distribution approximation to $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. The following lemma formalizes this recipe.

Lemma 2. *Suppose that $\mathbf{x} \sim \mathrm{Mult}(n, \boldsymbol{\pi})$ for a given length-m probability vector $\boldsymbol{\pi}$, and assume that given $\mathbf{x}$, $\mathbf{y}$ is a vector with conditionally independent elements distributed: $y^{(i)} \sim \mathrm{Bin}(x^{(i)}, q^{(i)})$. Then the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ is equal to that of $\mathbf{y} + \mathbf{x}^\star$, where*

$$\mathbf{x}^\star \sim \mathrm{Mult}\left(n - \mathbf{1}_m^{\mathrm{T}}\mathbf{y}, \frac{\boldsymbol{\pi} \circ (\mathbf{1}_m - \mathbf{q})}{1 - \boldsymbol{\pi}^{\mathrm{T}}\mathbf{q}}\right) \quad (9)$$

with $\mathbf{q} = [q^{(1)} \cdots q^{(m)}]^{\mathrm{T}}$, and the conditional mean of $\mathbf{x}$ given $\mathbf{y}$ is:

$$\mathbb{E}[\mathbf{x}|\mathbf{y}] = y + (n - \mathbf{1}_m^{\mathrm{T}}\mathbf{y})\left(\frac{\boldsymbol{\pi} \circ (\mathbf{1}_m - \mathbf{q})}{1 - \boldsymbol{\pi}^{\mathrm{T}}\mathbf{q}}\right). \quad (10)$$

Moreover, the marginal distribution of $\mathbf{y}$ has probability mass function given by:

$$\log p(\mathbf{y}) = \log(n!) + \mathbf{y}^{\mathrm{T}}(\log \boldsymbol{\pi} + \log \mathbf{q}) - \mathbf{1}_m^{\mathrm{T}} \log(\mathbf{y}!)$$
$$+ (n - \mathbf{1}_m^{\mathrm{T}}\mathbf{y}) \log(1 - \boldsymbol{\pi}^{\mathrm{T}}\mathbf{q}) - \log((n - \mathbf{1}_m^{\mathrm{T}}\mathbf{y})!),$$
$$(11)$$

with the convention $0 \log 0 \equiv 0$.

The proof is given in the supplementary materials.

4.3 Multinomial filtering

Putting together the results of lemma 1 and lemma 2 in a recursive fashion leads us to algorithm 1; line 3 is motivated by lemma 1, line 4 is motivated by (9)-(10).

Algorithm 1 Multinomial filtering with observations derived from $(\mathbf{x}_t)_{t \geq 1}$

1: **initialize** $\boldsymbol{\pi}_{0|0} \leftarrow \boldsymbol{\pi}_0$
2: **for** $t \geq 1$ **do**
3: $\quad \boldsymbol{\pi}_{t|t-1} \leftarrow (\boldsymbol{\pi}_{t-1|t-1}^{\mathrm{T}} \mathbf{K}_{t,\boldsymbol{\pi}_{t-1|t-1}})^{\mathrm{T}}$
4: $\quad \boldsymbol{\pi}_{t|t} \leftarrow \dfrac{\mathbf{y}_t}{n} + \left(1 - \dfrac{\mathbf{1}_m^{\mathrm{T}}\mathbf{y}_t}{n}\right) \dfrac{\boldsymbol{\pi}_{t|t-1} \circ (\mathbf{1}_m - \mathbf{q}_t)}{1 - \boldsymbol{\pi}_{t|t-1}^{\mathrm{T}}\mathbf{q}_t}$
5: $\quad \log w_t \leftarrow \log(n!) + \mathbf{y}_t^{\mathrm{T}}(\log \boldsymbol{\pi}_{t|t-1} + \log \mathbf{q}_t) - \mathbf{1}_m^{\mathrm{T}} \log(\mathbf{y}_t!) + (n - \mathbf{1}_m^{\mathrm{T}}\mathbf{y}_t) \log(1 - \boldsymbol{\pi}_{t|t-1}^{\mathrm{T}}\mathbf{q}_t) - \log((n - \mathbf{1}_m^{\mathrm{T}}\mathbf{y}_t)!)$
6: **end for**

One may take as output from algorithm 1 the approximations:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \approx \mathrm{Mult}(n, \boldsymbol{\pi}_{t|t-1}),$$
$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \overset{d}{\approx} \mathbf{y}_t + \mathbf{x}_t^\star,$$
$$(12)$$

where the $\overset{d}{\approx}$ term indicates approximation of $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ by the distribution of the sum of $\mathbf{y}_t$ (regarded as a

constant) and a random variable $\mathbf{x}_t^\star$ which is defined to have distribution:

$$\mathbf{x}_t^\star \sim \text{Mult}\left(n - \mathbf{1}_m^\text{T}\mathbf{y}_t, \frac{\boldsymbol{\pi}_{t|t-1} \circ (\mathbf{1}_m - \mathbf{q}_t)}{1 - \boldsymbol{\pi}_{t|t-1}^\text{T}\mathbf{q}_t}\right). \quad (13)$$

In view of (11), the quantities w_t computed in algorithm 1 can be used to approximate the marginal likelihood as follows:

$$p(\mathbf{y}_{1:t}) = p(\mathbf{y}_1) \prod_{s=2}^{t} p(\mathbf{y}_s|\mathbf{y}_{1:s-1}) \approx \prod_{s=1}^{t} w_s. \quad (14)$$

Now turning to the observation model from section 3.3 and noting that $(\mathbf{Z}_t, \mathbf{Y}_t)_{t\geq 1}$ is a hidden Markov model, we approximate the recursion:

$$p(\mathbf{Z}_{t-1}|\mathbf{Y}_{1:t-1}) \xrightarrow{\text{prediction}} p(\mathbf{Z}_t|\mathbf{Y}_{1:t-1}) \xrightarrow{\text{update}} p(\mathbf{Z}_t|\mathbf{Y}_{1:t}). \quad (15)$$

Many details are similar to those above so are given in the supplementary materials. The counterpart of algorithm 1 is algorithm 2, from which one may take the approximations:

$$\begin{aligned} p(\mathbf{Z}_t|\mathbf{Y}_{1:t-1}) &\approx \text{Mult}(n, \mathbf{P}_{t|t-1}), \\ p(\mathbf{Z}_t|\mathbf{Y}_{1:t}) &\stackrel{d}{\approx} \mathbf{Y}_t + \mathbf{Z}_t^\star, \end{aligned} \quad (16)$$

where

$$\mathbf{Z}_t^\star \sim \text{Mult}\left(n - \mathbf{1}_m^\text{T}\mathbf{Y}_t\mathbf{1}_m, \frac{\mathbf{P}_{t|t-1} \circ (\mathbf{1}_m \otimes \mathbf{1}_m - \mathbf{Q}_t)}{1 - \mathbf{1}_m^\text{T}(\mathbf{P}_{t|t-1} \circ \mathbf{Q}_t)\mathbf{1}_m}\right). \quad (17)$$

The marginal likelihood is approximated using the same formula as in (14) but with the w_t's computed as per algorithm 2.

Algorithm 2 Multinomial filtering with observations derived from $(\mathbf{Z}_t)_{t\geq 1}$

1: **initialize** $\boldsymbol{\pi}_{0|0} \leftarrow \boldsymbol{\pi}_0$
2: **for** $t \geq 1$ **do**
3:      $\mathbf{P}_{t|t-1} \leftarrow (\boldsymbol{\pi}_{t-1|t-1} \otimes \mathbf{1}_m) \circ \mathbf{K}_{t,\boldsymbol{\pi}_{t-1|t-1}}$
4:      $\mathbf{P}_{t|t} \leftarrow \dfrac{\mathbf{Y}_t}{n} + \dfrac{\mathbf{P}_{t|t-1} \circ (\mathbf{1}_m \otimes \mathbf{1}_m - \mathbf{Q}_t)}{1 - \mathbf{1}_m^\text{T}(\mathbf{P}_{t|t-1} \circ \mathbf{Q}_t)\mathbf{1}_m} -$
     $\left(\dfrac{\mathbf{1}_m^\text{T}\mathbf{Y}_t\mathbf{1}_m}{n}\right) \dfrac{\mathbf{P}_{t|t-1} \circ (\mathbf{1}_m \otimes \mathbf{1}_m - \mathbf{Q}_t)}{1 - \mathbf{1}_m^\text{T}(\mathbf{P}_{t|t-1} \circ \mathbf{Q}_t)\mathbf{1}_m}$
5:      $\log w_t \leftarrow \log(n!) + \mathbf{1}_m^\text{T}(\mathbf{Y}_t \circ \log \mathbf{P}_{t|t-1})\mathbf{1}_m + \mathbf{1}_m^\text{T}(\mathbf{Y}_t \circ \log \mathbf{Q}_t)\mathbf{1}_m - \mathbf{1}_m^\text{T}\log(\mathbf{Y}_t!)\mathbf{1}_m + (n - \mathbf{1}_m^\text{T}\mathbf{Y}_t\mathbf{1}_m)\log(1 - \mathbf{1}_m^\text{T}(\mathbf{P}_{t|t-1} \circ \mathbf{Q}_t)\mathbf{1}_m) - \log((n - \mathbf{1}_m^\text{T}\mathbf{Y}_t\mathbf{1}_m)!)$
6:      $\boldsymbol{\pi}_{t|t} \leftarrow (\mathbf{1}_m^\text{T}\mathbf{P}_{t|t})^\text{T}$
7: **end for**

4.4 Computational cost

The computational cost of algorithms 1 and 2 is independent of the overall population size n, except through factorial terms such as $\log(n!)$ and $\log((n - \mathbf{1}_m^\text{T}\mathbf{y})!)$. However these terms do not depend on the model parameters $\mathbf{K}_{t,\boldsymbol{\eta}}$, $\mathbf{q}_t$ etc., so can be pre-computed or even not computed at all if the approximate marginal likelihood needs to be evaluated only up to a constant of proportionality independent of model parameters. Leaving these factorial terms out the worst case costs of algorithms 1 and 2 are therefore respectively $\mathcal{O}(tm^2)$ and $\mathcal{O}(tm^3)$. Costs may be substantially lower in practice as $\mathbf{K}_{t,\boldsymbol{\eta}}$ and $\mathbf{q}_t$ are typically sparse. Similar observations hold for the smoothing algorithms.

This compares to $O(tmf(n))$ to simulate $(\mathbf{x}_t)_{t\geq 0}$ from the model where $f(n)$ is the complexity of sampling from $\text{Bin}(n, p)$, assuming no more than two non-zero entries in each row of $\mathbf{K}_{t,\boldsymbol{\eta}}$. A larger number of non-zero entries would imply a higher cost. Such a simulation is necessary (but usually not sufficient) to approximately evaluate the likelihood in ABSEIR (Brown et al., 2018). The worst case is $f(n) = O(n)$, but modest improvements are available if one accepts 'with high probability' performance measures (Farach-Colton and Tsai, 2015). The worst case time complexity of the Data Augmentation MCMC method (Lekone and Finkenstädt, 2006) is also linear in n. Whilst the wall-clock time of any given algorithm is of course heavily dependent on exactly how it is implemented, these considerations suggest that the proposed methods will have attractive computational costs in many applications, where m is often many orders of magnitude smaller than n

5 Numerical results

Additional details of models, data sources, algorithms, prior distributions, hyper-parameter settings, further numerical results and tutorials are given in the supplementary materials.

5.1 The 1995 Ebola outbreak in the Democratic Republic of Congo

We analyzed simulated and real data under a discrete-time SEIR model used by Lekone and Finkenstädt (2006) to investigate the impact of control interventions on the 1995 outbreak of Ebola in the Democratic Republic of Congo. Our experiments follow closely those in Lekone and Finkenstädt (2006) to allow comparisons with their Data Augmentation MCMC method. We also include comparisons to the ABC method from the ABSEIR R package (Brown et al., 2018), and results of least-squares fitting of an ODE model from Chowell et al. (2004) which Lekone and Finkenstädt (2006) used

as a benchmark.

The model of Lekone and Finkenstädt (2006) is the same as the SEIR model in (1) with $h = 1$, except that β is replaced by a time-varying parameter $\beta_t = \beta$ for $t < t_\star$ and $\beta_t = \beta e^{-\lambda(t - t_\star)}$ for $t \geq t_\star$ where $t_\star$ is the time at which control measures began. Thus $\mathbf{K}_{t, \boldsymbol{\eta}}$ is as in (3) but with β replaced by this β_t. Also following Lekone and Finkenstädt (2006), the data consist of daily counts of new cases (i.e. new infectives) and new deaths (i.e. new removals). In Lekone and Finkenstädt (2006) it was assumed these counts are observed directly, subject to known proportions of missing data. We consider a slightly more general observation model as per section 3.3 with $q_t^{(i,j)} = 0$ for all (i, j) except $(2, 3)$ and $(3, 4)$, and where $q_t^{(2,3)}$ and $q_t^{(3,4)}$ are treated as constant-in-t but otherwise unknown and to be estimated.

Synthetic data Using the following settings from Lekone and Finkenstädt (2006): $(\beta, \lambda, \rho, \gamma) = (0.2, 0.2, 0.2, 0.143)$, $S_0 = 5,364,500$, $E_0 = 1$, $I_0 = R_0 = 0$, $t_\star = 130$, plus $q_t^{(2,3)} = 291/316$ and $q_t^{(3,4)} = 236/316$ for all $t \geq 1$ informed by realistic proportions of non-missing data (Lekone and Finkenstädt, 2006), we simulated the epidemic from the model until extinction, which took 175 time steps. Table 1 shows MLE's from an EM algorithm which uses our approximate filtering and smoothing methods, and marginal posterior means and standard deviations estimated using a Metropolis-within-Gibbs MCMC algorithm which incorporates our approximate marginal likelihood, under three sets of prior distributions over $(\beta, \lambda, \rho, \gamma)$ labelled 'vague', 'informative' and 'noncentered' by Lekone and Finkenstädt (2006). The basic reproduction number is $R_0 = \beta/\gamma$. The results show accurate recovery of the true parameter values.

Real data We analyzed the same real Congo Ebola data as in Lekone and Finkenstädt (2006). Table 2 shows several interesting findings. 1) The results from our methods are generally closer to those from the Data Augmentation MCMC sampler of Lekone and Finkenstädt (2006) than those from the ABC method of Brown et al. (2018); the former targets the true posterior distribution whilst the latter does so only approximately. 2) Under the 'vague' prior our method finds bi-modal posteriors for β, λ, and $1/\rho$. For β, one of the modes roughly matches the posterior mean obtained using Lekone and Finkenstädt (2006) whilst the other is more similar to the least-squares estimate from Chowell et al. (2004); we conjecture that our MCMC sampler has better mixing than that of Lekone and Finkenstädt (2006), allowing it to find these two modes. 3) We can report estimates for $q^{(2,3)}$ and $q^{(3,4)}$, whilst the other methods do not. Figure 2 shows posterior and posterior-predictive distributions for the counts of new infectives each day. The former estimates for the true numbers which gave rise to the under-reported data, whilst the latter shows coverage of the data hence a good model fit (Gelman et al., 1996).

5.2 Accuracy: filtering bias and credible interval coverage

The purpose of this subsection is to study the accuracy of the approximate filtering distributions obtained from algorithm 2 when applied to the Ebola model described in subsection 5.1. The ground truth parameter values $(\beta, \lambda, \rho, \gamma)$ in the synthetic data experiment were taken together with $q_t^{(2,3)} = 291/316$, $q_t^{(3,4)} = 236/316$. We considered three population sizes $n = 5 \times 10^2, 5 \times 10^4, 5 \times 10^6$, and in each case the initial distribution was $\boldsymbol{\pi}_0 = [1 - 1/n, 1/n, 0, 0]^{\mathrm{T}}$. For each value of n, we simulated 2×10^4 data sets from the model, each over 200 time steps.

To assess accuracy we considered bias and credible-interval coverage. For the former we calculated the empirical bias associated with the mean vector of the approximation to $p(\mathbf{x}_t | \mathbf{Y}_{1:t})$ obtained from algorithm 2 as an estimator of $\mathbf{x}_t$. For the latter we calculated the empirical coverage of the nominal 95%-credible interval for the marginal over each $x_t^{(i)}$, $i = 1, 2, 3, 4$. For the true (i.e. approximation-free) filtering distributions, asymptotically in the number of simulated data sets

Parameter	β	λ	ρ	γ	$\mathbf{q^{(2,3)}}$	$\mathbf{q^{(3,4)}}$	$\mathbf{R_0}$
True value	0.2	0.2	0.2	0.143	0.92	0.75	1.40
MLE (EM-alg.)	0.20	0.18	0.21	0.139	1.00	0.81	1.44
MCMC (vague)	0.23 (0.028)	0.21 (0.080)	0.22 (0.076)	0.173 (0.024)	0.81 (0.140)	0.66 (0.119)	1.31 (0.088)
MCMC (infor.)	0.22 (0.020)	0.22 (0.065)	0.20 (0.035)	0.162 (0.017)	0.83 (0.130)	0.67 (0.112)	1.34 (0.082)
MCMC (noncent.)	0.32 (0.048)	0.35 (0.101)	0.17 (0.031)	0.256 (0.049)	0.79 (0.147)	0.64 (0.125)	1.28 (0.084)

Table 1: Parameter estimation for synthetic data under the Ebola model using our EM and MCMC methods under three sets of prior distributions specified by Lekone and Finkenstädt (2006). For the MCMC results, the posterior means is reported as the point estimate and the numbers in parentheses are posterior standard deviations.

Parameter	β	λ	$1/\rho$	$1/\gamma$	$q^{(2,3)}$	$q^{(3,4)}$	R_0
Our MCMC method vague prior	0.36 (0.049) 0.22 (0.025)	0.32 (0.140) 0.05 (0.008)	10.39 (1.554) 1.86 (0.487)	6.17 (1.042)	0.44 (0.103)	0.36 (0.088)	2.18 (0.227) 1.42 (0.102)
Our MCMC method informative prior	0.26 (0.033)	0.12 (0.064)	6.07 (1.919)	6.86 (0.834)	0.50 (0.109)	0.41 (0.093)	1.64 (0.696)
(Lekone and Finkenstädt, 2006) vague prior	0.24 (0.020)	0.16 (0.009)	9.43 (0.620)	5.71 (0.548)	-	-	1.38 (0.127)
(Lekone and Finkenstädt, 2006) informative prior	0.21 (0.017)	0.15 (0.010)	10.11 (0.713)	6.52 (0.564)	-	-	1.36 (0.128)
ODE + least squares (Chowell et al., 2004)	0.33 (0.006)	0.98 (unknown)	5.30 (0.230)	5.61 (0.190)	-	-	1.83 (0.060)
ABC ABSEIR (Brown et al., 2018)	0.3 (0.088)	0.36 (0.325)	7.91 (2.703)	15.01 (32.863)	-	-	3.66 (6.592)

Table 2: Parameter estimation for the real Ebola data. Numbers in parentheses in column 5 are standard errors, for all other columns they are posterior standard deviations. For columns 2,3,4,6 the parameter estimates are posterior means. For each of β, λ, and $1/\rho$ the pairs of estimates in column 1 were obtained from the respective bi-modal posteriors by applying k-means clustering, with $k = 2$, to the MCMC output.
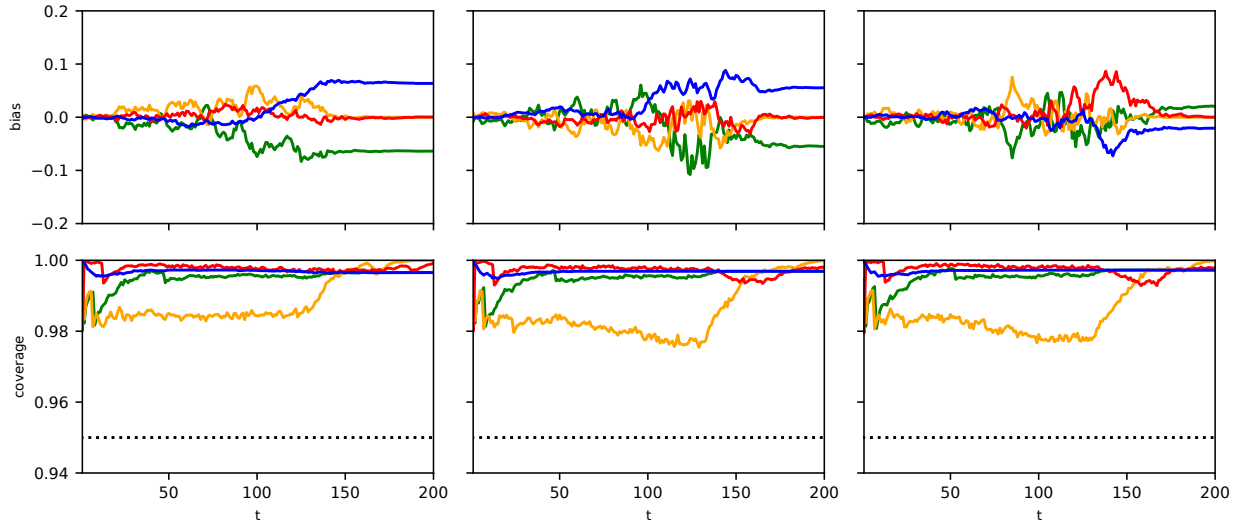

Figure 1: Empirical bias and empirical coverage of nominal 95%-credible intervals from 2×10^4 simulations over 200 time steps of the Ebola model. Columns from left to right: $n = 5 \times 10^2, 5 \times 10^4, 5 \times 10^6$. Top row: bias, bottom row: coverage. Red, yellow, blue, green correspond to $x_t^{(i)}$, $i = 1, 2, 3, 4$, i.e. susceptible, exposed, infective, recovered.

the bias would be zero and the coverage would be 95%.

Figure 1 shows that for all three values of n, the bias at every time step and for every compartment is less than 0.1 in magnitude. This shows the approximation is very accurate: the true values of $x_t^{(i)}$, $i = 1, 2, 3, 4$ are always integers, and a bias less than 0.5 in magnitude means that, on average, if the estimated number of individuals is rounded to the nearest integer, the true number of individuals is recovered. The credible interval coverage reported in figure 1 shows that the approximate filtering distributions tend to over-represent uncertainty: the empirical coverage at all time steps for all compartments of the nominal 95% interval is is between 97% and 100%. The bias and coverage appear robust to population size.

5.3 Estimating the time-varying reproduction number of COVID-19 in Wuhan, China

A compartmental model for estimating the time-varying reproduction number of COVID-19 in Wuhan, China, has recently been published in Kucharski et al. (2020). The model has 15 compartments: susceptibles in Wuhan become exposed and either stay in Wuhan or depart internationally, then in either case pass through further stages being exposed, infective, symptomatic and confirmed. The transmission rate is modelled as time-varying $(\beta_t)_{t \geq 0}$, a-priori by a geometric random walk, and β_t is considered proportional to the reproductive number R_t. Kucharski et al. (2020) proposed a Sequential Monte Carlo (SMC) algorithm to estimate $(R_t)_{t \geq 0}$ which weights samples of $(\beta_t)_{t \geq 0}$ by

the likelihood of the associated ODE solution under a Poisson observation model. Our methods can be used to replace their ODE model with a discrete-time stochastic version of the compartmental model, and with their Poisson model replaced our binomial observation model from section 3.3. Our version of the SMC algorithm weights samples of $(\beta_t)_{t \geq 0}$ by their approximate marginal likelihoods, computed using our multinomial filtering techniques. We jointly analyzed two of three data sets from Kucharski et al. (2020): daily counts of new infectives by date of symptom onset in Wuhan, and internationally exported from Wuhan.

Figure 3 shows our results in the format of Kucharski et al. (2020). Compared to results obtained using their ODE model (see supplementary material), our esti-mates of R_t are generally lower, and closer 1 for the period after travel restrictions are introduced; and our credible intervals for the in-sample plots are generally wider, reflecting the stochastic nature of our compartmental model. In the bottom two plots our posteriors are mostly concentrated on lower values than those from the method of Kucharski et al. (2020).
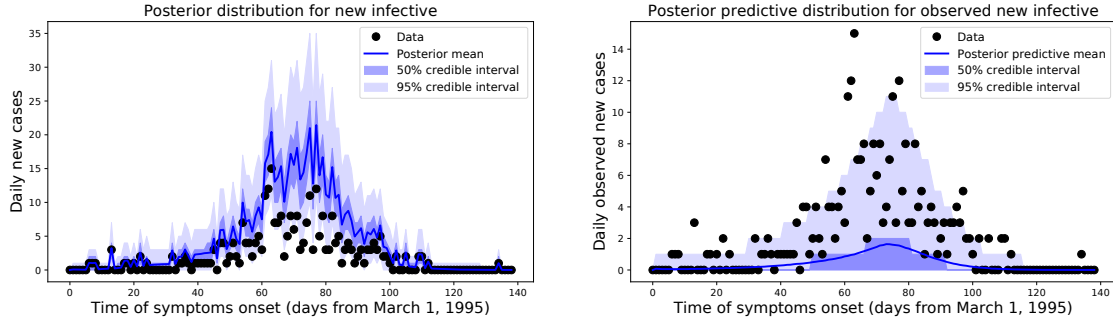
Acknowledgements

Figure 2: Analysis of real Ebola data with our method. Posterior smoothing distributions for the number of new infectives per day and posterior predictive distributions for the associated observations, i.e., subject to under-reporting. Control measures were introduced on day 70.
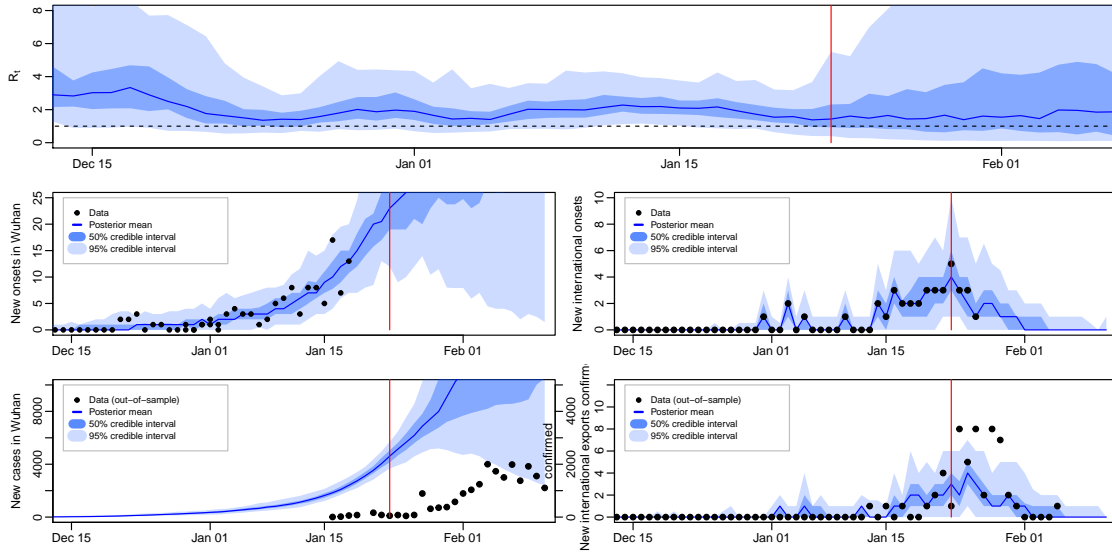

Figure 3: Results for the COVID-19 model using our methods. Red line is date at which travel restrictions were introduced. Top: estimated reproduction number. Middle row: estimated daily new confirmed cases in Wuhan (left) and internationally (right), both with in-sample data by date of symptom onset. Bottom row, left: estimated new symptomatic but possibly unconfirmed cases (left axis) and out-of-sample new confirmed cases data (right axis); right: estimated confirmed international cases by date of confirmation, and out-of-sample data.

References

Brauer, F. (2008). *Compartmental models in epidemiology*. Springer.

Bretó, C. (2018). Modeling and inference for infectious disease dynamics: a likelihood-based approach. *Statistical Science: a review journal of the Institute of Mathematical Statistics*, 33(1):57–69.

Brown, G. D., Oleson, J. J., and Porter, A. T. (2016). An empirically adjusted approach to reproductive number estimation for stochastic compartmental models: A case study of two Ebola outbreaks. *Biometrics*, 72(2):335–343.

Brown, G. D., Porter, A. T., Oleson, J. J., and Hinman, J. A. (2018). Approximate Bayesian Computation for spatial SEIR(S) epidemic models. *Spatial and Spatio-temporal Epidemiology*, 24:27–37.

Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., and Hyman, J. M. (2004). The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of Theoretical Biology*, 229(1):119–126.

de Lusignan, S., Bernal, J. L., Zambon, M., Akinyemi, O., Amirthalingam, G., Andrews, N., Borrow, R., Byford, R., Charlett, A., Dabrera, G., et al. (2020). Emergence of a novel coronavirus (COVID-19): protocol for extending surveillance used by the Royal College of general practitioners research and surveillance centre and public health England. *JMIR Public Health and Surveillance*, 6(2):e18606.

Farach-Colton, M. and Tsai, M.-T. (2015). Exact sublinear binomial sampling. *Algorithmica*, 73(4):637–651.

Fasiolo, M., Pya, N., and Wood, S. N. (2016). A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology. *Statistical Science*, 31(1):96–118.

Funk, S. and King, A. A. (2020). Choices and trade-offs in inference with infectious disease models. *Epidemics*, 30:100383.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4):733–760.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Sun, F., Jit, M., Munday, J. D., et al. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(5):553–558.

Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7(1):49–58.

Kurtz, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(2):344–356.

Kypraios, T., Neal, P., and Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. *Mathematical Biosciences*, 287:42–53.

Lekone, P. E. and Finkenstädt, B. F. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 62(4):1170–1177.

McKinley, T. J., Vernon, I., Andrianakis, I., McCreesh, N., Oakley, J. E., Nsubuga, R. N., Goldstein, M., and White, R. G. (2018). Approximate Bayesian computation and Simulation-Based Inference for Complex Stochastic Epidemic Models. *Statistical Science*, 33(1):4–18.

Murray, L., Lundén, D., Kudlicka, J., Broman, D., and Schön, T. (2018). Delayed sampling and automatic Rao-Blackwellization of probabilistic programs. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR.

O'Neill, P. D. (2010). Introduction and snapshot review: relating infectious disease transmission models to data. *Statistics in Medicine*, 29(20):2069–2077.

Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., et al. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*.

Roberts, M., Andreasen, V., Lloyd, A., and Pellis, L. (2015). Nine challenges for deterministic epidemic models. *Epidemics*, 10:49–53.

Rubrichi, S., Smoreda, Z., and Musolesi, M. (2018). A comparison of spatial-based targeted disease mitigation strategies using mobile phone data. *EPJ Data Science*, 7(1):1–15.

Stocks, T. (2019). Iterated filtering methods for Markov process epidemic models. In Held, L., Hens, N., D O'Neill, P., and Wallinga, J., editors, *Handbook of Infectious Disease Data Analysis*, chapter 11, pages 199–220. CRC Press.

Wu, J. T., Leung, K., and Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395(10225):689–697.