Bayesian Modeling of Partially Observed Epidemic Count Data

Jonathan Fintzi

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Vladimir Minin, Chair

Jon Wakefield, Chair

M. Elizabeth Halloran

James Hughes

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Bayesian Modeling of Partially Observed Epidemic Count Data

Jonathan Fintzi

Co-Chairs of the Supervisory Committee:
Co-chair Vladimir Minin

Co-chair Jon Wakefield

An incredible abstract with all the best words will appear here.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

ACF: Autocorrelation function.

CDF: Cumulative distribution function.

CLE: Chemical Langevin equation.

CP: Centered parameterization.

CTMC: Continuous–time Markov chain.

DA: Data augmentation.

ELIPTSS: Elliptical slice sampling.

ESS: Effective sample size.

GSS: Gaussian slice sampler.

ILI: Influenza–like illness.

LNA: Linear noise approximation.

MJP: Markov jump process.

MMTL: Multinomial modification of the τ–leaping algorithm.

PACF: Partial autocorrelation function.

PMCMC: Particle Markov chain Monte Carlo.

PMMH: Particle marginal Metropolis–Hastings.

PSRF: Potential scale reduction factor.

NCP: Non–centered parameterization.

SDE: Stochastic differential equation.

SEM: Stochastic epidemic model.

ACKNOWLEDGMENTS

Very grateful to many people.

DEDICATION

Dedication to important people.

Chapter 1

INTRODUCTION AND DATA SETTING

1.1 Motivating examples

1.1.1 Influenza in a British boarding school

1.1.2 Ebola in West Africa

1.1.3 Pandemic A(H1N1) influenza in Finland

1.2 Organization of this dissertation

Chapter 2

BACKGROUND

2.1 Models for the Spread of Infectious Disease

2.1.1 Deterministic Representations

2.1.2 Stochastic Representations

Agent–based models

Population–level models

2.1.3 Large–Population Approximations

Diffusion approximations of Markov jump processes

Linear noise approximation

2.2 Computational Approaches to Fitting Stochastic Epidemic Models

2.3 Bayesian Computation and Markov Chain Monte Carlo

2.3.1 Markov Chain Monte Carlo

Bayesian Data Augmentation

Slice sampling for model parameters

Elliptical slice sampling

Chapter 3

AGENT–BASED DATA AUGMENTATION FOR FITTING STOCHASTIC EPIDEMIC MODELS TO PREVALENCE DATA

Chapter 4

APPROXIMATE INFERENCE FOR STOCHASTIC EPIDEMIC MODELS OF OUTBREAKS IN LARGE POPULATIONS

4.1 Overview

Surveillance and outbreak response systems often report incidence counts of new cases detected in each inter–observation time interval. Analyzing this type of time series data is challenging since we must overcome many of the same challenges that we face in modeling the transmission dynamics of infectious diseases in small population settings with prevalence data — discrete snapshots of a continuously evolving epidemic process, detecting a fraction of the new cases, and often directly observing only one aspect of the disease process. Furthermore, our task is made more difficult by the additional computational burden that results from repeated evaluation of CTMC likelihoods; the products of exponential waiting time distributions consist of polynomially increasing numbers of terms, and agent–based data augmentation (DA) MCMC algorithms become unwieldy as the numbers of subject–path proposals required to meaningfully perturb the CTMC likelihood get large [20].

In this chapter, we show how the LNA of Section 2.1.3 can be adapted to obtain approximate inference for SEMs fit to epidemic count data in large populations. Our contributions are threefold: First, we demonstrate how the SEM dynamics should be reparameterized so that the LNA can be used to approximate transition densities of the counting processes for disease state transition events. Second, we fold the LNA into a Bayesian DA framework in which latent LNA paths are sampled using the elliptical slice sampling (EliptSS) algorithm of [37]. This provides us with general machinery for jointly updating the latent paths while absolving us of the *de facto* requirement that the emission probability distribution to be Gaussian in order to preserve computational efficiency as in [17, 34], lest we resort

to computationally intensive particle filter methods for non–Gaussian emission distributions as in [27]. Finally, we introduce a non–centered parameterization (NCP) for the LNA that massively improves the efficiency of our DA MCMC framework and makes it tractable for fitting complex models.

4.2 Fitting Stochastic Epidemic Models via the Linear Noise Approximation

For clarity, we will present the algorithm for fitting SEMs via the LNA in the context of fitting the susceptible–infected–recovered (SIR) model to negative binomial distributed incidence counts. We will, however, provide notation where appropriate so that the generality of the algorithm should be apparent. The SIR model is an abstraction of the transmission dynamics of an outbreak as a closed, homogeneously mixing population of N exchangeable individuals who are either susceptible (S), infected, and hence infectious, (I), or recovered (R). It is important to note that the model compartments refer to disease states as they relate to the transmission dynamics, not the disease process. Thus, an individual is considered to be recovered when she no longer has infectious contact with other individuals in the population, not when she clears disease carriage. As another example, in the susceptible–exposed–infected–recovered (SEIR) type models that we will consider later, the latent period in which an individual is exposed, but not yet infectious, should be understood as possibly varying in population with different contact dynamics, even when the incubation period of the pathogen should arguably be consistent across groups.

4.2.1 Measurement Process and Data

Incidence data, $\mathbf{Y} = \{Y_1, \ldots, Y_L\}$, arise as increments of the numbers of new cases accumulated in a set of time intervals, $\mathcal{I} = \{\mathcal{I}_1, \ldots, \mathcal{I}_L : \mathcal{I}_\ell = (t_{\ell-1}, t_\ell]\}$. In outbreak or surveillance settings, we do not typically believe that every case is detected since individuals may be asymptomatic or may escape detection. Let $\mathbf{N}^c = (N^c_{SI}, N^c_{IR})$ denote the counting process for the cumulative numbers of infections ($S \to I$ transitions) and recoveries ($I \to R$ transitions), and let $\Delta\mathbf{N}^c(t_\ell) = \mathbf{N}^c(t_\ell) - \mathbf{N}^c(t_{\ell-1})$ denote the change in cumulative numbers of

transitions over $\mathcal{I}_\ell$; so, $\Delta N^c_{SI}(t_\ell)$ is the incidence over $(t_{\ell-1}, t_\ell]$. We might choose to model the number of observed cases as a negative binomial sample of the true incidence with detection rate ρ and over–dispersion parameter ϕ. Thus,

$$Y_\ell | \Delta N^c_{SI}(t_\ell), \rho \sim \text{Neg.Binom.}(\mu = \rho \Delta N^c_{SI}(t_\ell),\ \sigma^2 = \mu + \mu^2/\phi). \tag{4.1}$$

There are two minor points that we wish to make before proceeding. First, we have allowed for the possibility that cases are over–reported. This is not a necessary assumption for any of the subsequent results. It is also not unreasonable when studying outbreaks in large populations where the "fog of war" might lead to inflation of reported incidence or misclassification of individuals whose symptoms are similar to the disease of interest. Allowing for the possibility of over–reporting is also not particularly problematic when the detection probability is low since the emission densities will have negligible mass above the true incidence. The second point is that we are making this modeling choice with an eye on the compatibility of the emission distribution with the eventual LNA approximation, which takes real, not integer, values. The negative binomial distribution is well defined for non–integer values of its mean parameter.

4.2.2 Latent Epidemic Process

The SIR model is often expressed in terms of compartment counts, $\mathbf{X}^c = \{S^c, I^c, R^c\}$, that evolve in continuous time on state space $\mathcal{S}^c_X = \{\mathcal{C}_{lmn} : l, m, n \in \{0, \ldots, N\},\ l + m + n = P\}$. We will make the (not particularly limiting) modeling choice to express the waiting times between disease state transitions as being exponentially distributed. Thus, $\mathbf{X}$ evolves according to a Markov jump process (MJP). If our data had consisted of prevalence counts, which arise as partial observations of infected individuals, we might have chosen to approximate transition densities of the MJP for $\mathbf{X}$ in the usual way that appears in [34, 17].

However, incidence data are discretely observed, partial realizations of the increments of counting processes that evolve continuously in time as individuals transition among disease states. The emission probabilities for incidence data, e.g., (4.1), depend on the change in

N^c_{SI} over the time interval $(t_{\ell-1}, t_\ell]$, not on the change in I over the interval. It would be incorrect to treat incidence as simply the difference in prevalence. We could easily conjure up a scenario where there are positive numbers of infections, but where the prevalence, discretely observed, does not appear to change due to an equal number of recoveries. We need to construct the LNA that approximates transition densities of $\mathbf{N}$ if we are to write down correctly specified emission probabilities.

The cumulative incidence process for infections and recoveries, $\mathbf{N}^c$, is a Markov jump process with state space $\mathcal{S}^c_N = \{\mathcal{C}_{jk} : j, k \in \{0, \ldots, N\}\}$. Let β denote the per–contact infection rate, and μ denote the rate at which each infected individual recovers. The rate at which $\mathbf{N}^c$ transitions from state $\mathbf{n}$ to $\mathbf{n}'$ is

$$\boldsymbol{\lambda}_{\mathbf{n},\mathbf{n}'} = \begin{cases} \lambda_{SI} = \beta SI, & \mathbf{n} = (n_{SI}, n_{IR}),\ \mathbf{n}' = (n_{SI}+1, n_{IR}),\ \text{and } n_{SI}+1 \leq P, \\ \lambda_{IR} = \mu I, & \mathbf{n} = (n_{SI}, n_{IR}),\ \mathbf{n}' = (n_{SI}, n_{IR}+1),\ \text{and } n_{IR}+1 \leq P, \\ 0, & \text{for all other } \mathbf{n} \text{ and } \mathbf{n}'. \end{cases} \tag{4.2}$$

4.2.3 Tractable Approximations for Intractable Likelihoods

We would like to make inferences about the posterior distribution of the parameters, e.g., $\boldsymbol{\theta} = (\beta, \mu, \mathbf{X}_0, \rho)$, that govern the latent epidemic process and sampling distribution,

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{Y}) \propto \pi(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) &= \int L(\mathbf{Y}|\mathbf{N}^c, \boldsymbol{\theta})\pi(\mathbf{N}^c|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\pi(\mathbf{N}^c) \\ &= \int_{\mathcal{S}^c} \prod_{\ell=1}^{L} \Pr\left(\mathbf{Y}_\ell|\Delta\mathbf{N}^c_{SI}(t_\ell), \boldsymbol{\theta}\right) \pi\left(\mathbf{N}^c(t_\ell)|\mathbf{n}^c(t_{\ell-1}), \boldsymbol{\theta}\right) \pi(\boldsymbol{\theta})\mathrm{d}\pi(\mathbf{N}^c) \end{aligned} \tag{4.3}$$

where $\pi(\boldsymbol{\theta})$ specifies the prior density of the model parameters. However, this integral is analytically intractable and is challenging to compute numerically due to the size of the state space of $\mathbf{N}^c$. In the following subsections, we will obtain the LNA for transition densities of $\mathbf{N}^c$, turning (4.3) into an integral over a much more computationally tractable product of Gaussian densities and non–Gaussian emission probabilities. As we shall see, approximating the complete data likelihood in the posterior $\pi(\boldsymbol{\theta}, \mathbf{N}^c|\boldsymbol{\theta})$ with a Gaussian state space model will facilitate the use of efficient algorithms for sampling from the approximate posterior.

4.2.4 Diffusion Approximation

As outlined in Section 2.1.3, there are a variety of methods for arriving at a diffusion approximation for a Markov jump process, which under certain conditions yield equivalent results (for a comprehensive reference, see [21]). In the interest of clarity, we follow [17, 26, 27, 46] and appeal to an intuitive, though somewhat informal, construction of the CLE by matching its drift and diffusion with the approximate moments of increments of the MJP path in infinitesimal time intervals. For more detailed presentations see [21, 25, 45].

Suppose that, at the current time, the compartment counts are given by $\mathbf{X}^c(t) = \mathbf{x}_t^c$. We are interested in approximating the numbers of infections and recoveries in a small time interval, $(t, t + \mathrm{d}t]$, i.e., $\mathbf{N}^c(t + \mathrm{d}t) - \mathbf{N}(t)$. Now, suppose that we can choose $\mathrm{d}t$ such that the following two *leap* conditions hold:

1. $\mathrm{d}t$ is sufficiently *small* that the $\mathbf{X}^c$ is essentially unchanged over $(t, t + \mathrm{d}t]$, so that the rates of infections and recoveries are approximately constant:

$$\boldsymbol{\lambda}(\mathbf{X}^c(t')) \approx \boldsymbol{\lambda}(\mathbf{x}^c(t)),\ \forall t' \in (t, t + \mathrm{d}t]. \tag{4.4}$$

2. $\mathrm{d}t$ is sufficiently *large* that we can expect many disease state transitions of each type:

$$\boldsymbol{\lambda}(\mathbf{x}^c(t)) \gg \mathbf{1}. \tag{4.5}$$

Condition (4.4), which can be trivially satisfied just by choosing $\mathrm{d}t$ to be infinitesimally small, implies that the numbers of infections and recoveries in $(t, t + \mathrm{d}t]$ are essentially independent of one another since the rates at which they occur are approximately constant within the interval [25]. This condition also carries the stronger implication that the numbers of infections and recoveries in the interval are independent Poisson random variables with rates $\boldsymbol{\lambda}(\mathbf{x}^c(t)\mathrm{d}t)$, i.e., $N_{SI}^c(\mathrm{d}t) \sim \text{Poisson}(\beta S(t)I(t)\mathrm{d}t)$ and $N_{IR}^c(t + \mathrm{d}t) \sim \text{Poisson}(\mu I(t)\mathrm{d}t)$. Condition (4.5), which we can reasonably expect to be satisfied in large populations where transmission dynamics are near their deterministic ODE limits [45], implies that the Poisson distributed increments can be well approximated by independent Gaussian random variables.

When (4.4) and (4.5) are satisfied, we can approximate the integer–valued processes, $\mathbf{X}^c$ and $\mathbf{N}^c$, with the real–valued processes, $\mathbf{X}$ and $\mathbf{N}$. For the SIR model, the state space of $\mathbf{X}$ is $\mathcal{S}_X^R = \{\mathcal{V}_{lmn} : l, m, n \in [0, N],\ l + m + n = P\}$, and the state space of $\mathbf{N}$ is $\mathcal{S}_N^R = \{\mathcal{V}_{jk} : j, k \in [0, N],\ j \geq k\}$. More generally, the state space of $\mathbf{X}$ will be the set of compartment volumes that are non–negative and that sum to the population size, while the state space of $\mathbf{N}$ is the set of non–decreasing and non–negative incidence paths, constrained so that they do not lead to invalid prevalence paths (e.g., if at some point there are more recoveries than infections, which would lead to a negative number of infected individuals). For now, we will ignore the constraints on $\mathcal{S}_N^R$ and $\mathcal{S}_X^R$, and approximate the changes in cumulative incidence of infections and recoveries in an infinitesimal time step as

$$\mathbf{N}(t + \mathrm{d}t) - \mathbf{N}(t) \approx \boldsymbol{\lambda}(\mathbf{X}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{X}(t))^{1/2}\mathrm{d}t^{1/2}\mathbf{Z}, \tag{4.6}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\,(\boldsymbol{\lambda}(\mathbf{X}))$ and $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$. This implies the equivalent CLE,

$$\mathrm{d}\mathbf{N}(t) = \boldsymbol{\lambda}(\mathbf{X}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{X}(t))^{1/2}\mathrm{d}\mathbf{W}_t, \tag{4.7}$$

where $\mathbf{W}_t$ is a vector of independent Brownian motion and $\boldsymbol{\Lambda}(\mathbf{X}(t))^{1/2}$ denotes the matrix square root of $\boldsymbol{\Lambda}(\mathbf{X}(t))$.

Reparameterizating the CLE in terms of incidence

The LNA of (4.7) will involve derivatives of the rates, $\boldsymbol{\lambda}$, with respect to the incidence process, $\mathbf{N}$. In order to enable us to compute these derivatives, we borrow from [9, 28] a reparameterization for $\mathbf{X}(t)$ in terms of $\mathbf{N}(t)$, conditional on the initial conditions $\mathbf{X}(t) = \mathbf{X}_0$ and $\mathbf{N}(t) = \mathbf{0}$. Let $\mathbf{A}$ denote the matrix whose rows specify changes in counts of susceptible, infected, and recovered individuals corresponding to one infection or recovery event:

$$\mathbf{A} = \begin{array}{c} \\ S \to I \\ I \to R \end{array} \begin{array}{c} \begin{array}{ccc} S & I & R \end{array} \\ \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \end{array}. \tag{4.8}$$

Now, $\mathbf{X}$ is coupled to $\mathbf{N}$ via the relationship,

$$\mathbf{X}(t) = \mathbf{X}_0 + \mathbf{A}^T\mathbf{N}(t). \tag{4.9}$$

For the SIR model,

$$\begin{pmatrix} S(t) \\ I(t) \\ R(t) \end{pmatrix} = \begin{pmatrix} S_0 - N_{SI}(t) \\ I_0 + N_{SI}(t) - N_{IR}(t) \\ R_0 + N_{IR}(t) \end{pmatrix}, \tag{4.10}$$

which enables us to rewrite (4.7) as

$$\begin{aligned} \mathrm{d}\mathbf{N}(t) &= \boldsymbol{\lambda}(\mathbf{N}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{N}(t))^{1/2}\mathrm{d}\mathbf{W}_t \qquad (4.11)\\ &= \begin{pmatrix} \beta(S_0 - N_{SI}(t))(I_0 + N_{SI}(t) - N_{IR}(t)) \\ \mu(I_0 + N_{IR}(t))\mathrm{d}t \end{pmatrix} \mathrm{d}t + \\ &\qquad \begin{pmatrix} \beta(S_0 - N_{SI}(t))(I_0 + N_{SI}(t) - N_{IR}(t)) & 0 \\ 0 & \mu(I_0 + N_{IR}(t)) \end{pmatrix}^{1/2} \mathrm{d}\mathbf{W}_t. \end{aligned}$$

Log transforming the CLE

Changes in compartment volumes affect the rates, and hence increments in the incidence process, multiplicatively. Therefore, from a scientific perspective, we would like for perturbations about the drift in (4.11) to be symmetric on a multiplicative, not an additive scale. Hence, we log transform (4.11). Let $\widetilde{\mathbf{N}} = \log(\mathbf{N} + \mathbf{1}) \implies \mathbf{N} = \exp(\widetilde{\mathbf{N}}) - \mathbf{1}$. By Itô's lemma [39], the corresponding SDE for $\widetilde{\mathbf{N}}$ is

$$\begin{aligned} \mathrm{d}\widetilde{\mathbf{N}}(t) = \mathrm{diag}\left(\exp(-\widetilde{\mathbf{N}}(t)) - 0.5\exp(-2\widetilde{\mathbf{N}}(t))\right)\boldsymbol{\lambda}\left(\exp(\widetilde{\mathbf{N}}(t)) - \mathbf{1}\right)\mathrm{d}t + \\ \mathrm{diag}\left(\exp(-\widetilde{\mathbf{N}}(t))\right)\boldsymbol{\Lambda}\left(\exp(\widetilde{\mathbf{N}}(t)) - 1\right)^{1/2}\mathrm{d}\mathbf{W}_t \qquad (4.12)\\ = \boldsymbol{\eta}(\widetilde{\mathbf{N}}(t))\mathrm{d}t + \boldsymbol{\Phi}(\widetilde{\mathbf{N}}(t))^{1/2}\mathrm{d}\mathbf{W}_t \qquad (4.13) \end{aligned}$$

4.2.5 Linear Noise Approximation

In Section 2.1.3, we followed [17, 26, 27] in obtaining the LNA for SDEs of the same form as (4.13). Briefly, the derivation proceeded as follows: we first decomposed $\widetilde{\mathbf{N}}$ into its

deterministic ODE limit and a stochastic residual. The SDE corresponding to (4.12) was then Taylor expanded around its deterministic limit, discarding higher order terms, to obtain a linear SDE for the residual. This linear SDE had an explicit solution as a Gaussian random variable. As noted in [45], the LNA can reasonably approximate the stochastic aspects of a density dependent MJP when conditions (4.4) and (4.5) are satisfied, at least over short time horizons. Over longer time periods the approximation may deteriorate as departures from the deterministic behavior of the system, which is determined by its initial conditions, accumulate. One solution, proposed in [17] and that we will adopt here, is to restart the LNA approximation at the beginning of each inter–observation interval.

The restarting LNA of (4.13) over a time interval, $(t_{\ell-1}, t_\ell]$, was seen to be a Gaussian approximation of the transition density of $\widetilde{\mathbf{N}}$,

$$\widetilde{\mathbf{N}}(t_\ell)|\widetilde{\mathbf{n}}(t_{\ell-1}), \mathbf{x}(t_{\ell-1}), \boldsymbol{\theta} \sim MVN\left(\boldsymbol{\mu}(t_\ell) + \mathbf{m}(\widetilde{\mathbf{n}}(t_{\ell-1}) - \boldsymbol{\mu}(t_{\ell-1})), \boldsymbol{\Sigma}(t_\ell)\right), \tag{4.14}$$

where $\boldsymbol{\mu}(\cdot)$, $\mathbf{m}(\cdot)$, and $\boldsymbol{\Sigma}(\cdot)$ are solutions to the coupled, non–autonomous system of ODEs,

$$\frac{\mathrm{d}\boldsymbol{\mu}(t)}{\mathrm{d}t} = \boldsymbol{\eta}(\boldsymbol{\mu}(t)), \tag{4.15}$$

$$\frac{\mathrm{d}\mathbf{m}(t)}{\mathrm{d}t} = \mathbf{F}(t)\mathbf{m}(t), \tag{4.16}$$

$$\frac{\mathrm{d}\boldsymbol{\Sigma}(t)}{\mathrm{d}t} = \mathbf{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\mathbf{F}(t)^T + \boldsymbol{\Phi}(t), \tag{4.17}$$

with respect to initial conditions $\mathbf{N}(t_{\ell-1}) = \mathbf{0}$, $\mathbf{X}(t_{\ell-1}) = \mathbf{x}(t_{\ell-1})$, $\mathbf{m}(t_{\ell-1}) = \mathbf{0}$, and $\boldsymbol{\Sigma}(t_{\ell-1}) = \mathbf{0}$, and where $\mathbf{F}(t)$ is the Jacobian $\left(\frac{\partial \boldsymbol{\eta}_i(\boldsymbol{\mu}(t))}{\partial \boldsymbol{\mu}_j(t)}\right)_{i,j\in 1,\ldots,|\widetilde{\mathbf{N}}|}$ evaluated along the solution to (4.15). Note that we need never actually solve (4.16) since $\mathbf{m}(t_{\ell-1}) = \mathbf{0}$ implies that $\mathbf{m}(t_\ell) = \mathbf{0}\ \forall\ l = 0, \ldots, L-1$.

Approximating the transition densities of $\mathbf{N}$ using the LNA, (4.14), enables us to approximate the observed data likelihood in (4.3) with a Gaussian state space model. The

augmented approximate posterior is

$$\begin{aligned}\pi(\widetilde{\mathbf{N}}, \boldsymbol{\theta}|\mathbf{Y}) &\propto L(\mathbf{Y}|\widetilde{\mathbf{N}}, \boldsymbol{\theta}) \mathbb{1}_{\left\{\mathbf{N} \in \mathcal{S}_N^R\right\}} \mathbb{1}_{\left\{\mathbf{X} \in \mathcal{S}_X^R\right\}} \pi(\widetilde{\mathbf{N}}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\ &= \prod_{\ell=1}^{L} \Pr(\mathbf{Y}_\ell|\Delta\widetilde{\mathbf{N}}(t_\ell), \boldsymbol{\theta}) \mathbb{1}_{\left\{\mathbf{N}(t_\ell) \in \mathcal{S}_N^R\right\}} \mathbb{1}_{\left\{\mathbf{X}(t_\ell) \in \mathcal{S}_X^R\right\}} \pi(\widetilde{\mathbf{N}}(t_\ell)|\widetilde{\mathbf{n}}(t_{\ell-1}), \mathbf{x}(t_{\ell-1}), \boldsymbol{\theta}) \pi(\boldsymbol{\theta}).\end{aligned} \tag{4.18}$$

Note that the emission probabilities in (4.18) depend on the incidence, not the log–incidence, but that this just requires a simple reparameterization of the emission distribution. In our example, the observed incidence is a negative binomial sample of the true incidence. We also explicitly include indicators for whether the LNA path respects the positivity and monotonicity constraints of the original MJP. We do this for two reasons: We wish to more faithfully approximate the MJP. We also wish to avoid numerical instabilities that arise when $\mathbf{N}$ or $\mathbf{X}$ become negative and that can cause routines for numerically integrating the LNA ODEs to fail.

4.2.6 Inference via the Linear Noise Approximation

To this point, we have discussed how to approximate transition densities of a MJP via the LNA. However, this is only half the battle since we must also address the computational aspects of sampling from the augmented approximate posterior, (4.18). A central computation challenge that plagues DA MCMC is that MCMC chains may suffer from severe autocorrelation when the algorithm alternately updates the latent variables given the parameters, and parameters given the latent variables, see e.g., [7, 41, 42, 47]. As we can see in Figure, a DA MCMC algorithm that alternates between updates LNA paths and model parameters is no exception.

Non–centered Parameterization

We can improve the mixing of our MCMC chains by reparameterizing the log–incidence process as a deterministic mapping of standard normal random variables, $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$,

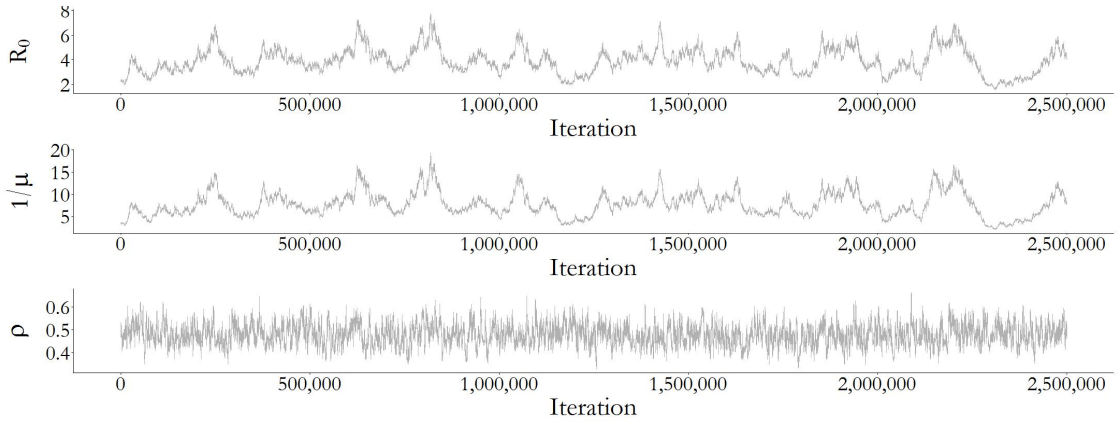

Figure 4.1: Posterior traceplots for parameters of interest for a single MCMC chain of an SIR model fit to negative binomial distributed incidence data. MCMC targeted the posterior, 4.18, alternately updating the non–restarting LNA for $\widetilde{\mathbf{N}}|\boldsymbol{\theta}, \mathbf{Y}$ via elliptical slice sampling, and $\boldsymbol{\theta}|\widetilde{\mathbf{N}}, \mathbf{Y}$ via a multivariate random walk Metropolis algorithm. $R_0 = \beta N/\mu$ is the basic reproductive number, $1/\mu$ is the mean infectious period duration, and ρ is the mean case detection rate. The true values of R_0, $1/\mu$, and ρ were 3.5, 7, and 0.5, respectively.

which are *a priori* independent of the model parameters. Let $\widetilde{\mathbf{N}}(t_\ell) \sim MVN(\boldsymbol{\mu}(t_\ell), \boldsymbol{\Sigma}(t_\ell)$ and $\mathbf{Z}(t_\ell) \sim MVN(\mathbf{0}, \mathbf{I})$. Then the NCP is given by $(\boldsymbol{\theta}, \mathbf{Z})$ and maps onto the CP via

$$\widetilde{\mathbf{N}}(t_\ell) \stackrel{\mathcal{L}}{=} \widetilde{\mathbf{W}}(t_\ell), \;\; \widetilde{\mathbf{W}}(t_\ell) = \boldsymbol{\mu}(t_\ell) + \boldsymbol{\Sigma}(t_\ell)^{1/2}\mathbf{Z}(t_\ell). \tag{4.19}$$

We now target the joint posterior of the model parameters and the non–centered LNA draws,

$$\pi(\boldsymbol{\theta}, \mathbf{Z}|\mathbf{Y}) \propto L(\mathbf{Y}|\texttt{doLNA}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I}))\mathbb{1}_{\left\{\mathbf{N}(\mathbf{Z},\boldsymbol{\theta},\mathcal{I})\in\mathcal{S}_N^R\right\}}\mathbb{1}_{\left\{\mathbf{X}(\mathbf{Z},\boldsymbol{\theta},\mathcal{I})\in\mathcal{S}_X^R\right\}}\pi(\mathbf{Z})\pi(\boldsymbol{\theta}). \tag{4.20}$$

We will denote by $\mathbf{N}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I})$ and $\mathbf{X}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I})$ the incidence and prevalence sample paths that are output by the `doLNA` procedure. The procedure for this mapping, denoted `doLNA`, is presented in Algorithm (1).

The NCP of the log–incidence process substantially improves the mixing of MCMC chains that alternate between updates to $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ and $\boldsymbol{\theta}|\mathbf{Z}, \mathbf{Y}$. Figure 4.1 shows traceplots of model parameters for one of MCMC chains for an SIR model fit to Poisson distributed incidence data using the centered parameterization (CP) of the LNA transition density. MCMC was

Algorithm 1 Mapping standard normal draws onto LNA sample paths.

procedure DOLNA($\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I}$)
  initialize: $\mathbf{X}(t_0) \leftarrow \mathbf{X}_0,\ \mathbf{N}(t_0) \leftarrow \mathbf{0},\ \widetilde{\mathbf{N}}(t_0) \leftarrow \mathbf{0},\ \boldsymbol{\mu}(t_0) \leftarrow \mathbf{0},\ \boldsymbol{\Sigma}(t_0) \leftarrow \mathbf{0}$
  for $\ell = 1, \ldots, L$ **do**
    $\boldsymbol{\mu}(t_\ell),\ \boldsymbol{\Sigma}(t_\ell) \leftarrow$ solutions to (4.15) and (4.17) over $(t_{\ell-1}, t_\ell]$
    $\widetilde{\mathbf{N}}(t_\ell) \leftarrow \boldsymbol{\mu}(t_\ell) + \boldsymbol{\Sigma}(t_\ell)^{1/2}\mathbf{Z}(t_\ell)$ ▷ non–centered parameterization
    $\mathbf{N}(t_\ell) \leftarrow \mathbf{N}(t_{\ell-1}) + \exp(\widetilde{\mathbf{N}}(t_\ell)) - \mathbf{1}$
    restart initial conditions:
      $\mathbf{X}(t_\ell) \leftarrow \mathbf{X}(t_{\ell-1}) + \mathbf{A}^T(\mathbf{N}(t_\ell) - \mathbf{N}(t_{\ell-1})),\ \widetilde{\mathbf{N}}(t_\ell) \leftarrow \mathbf{0},\ \boldsymbol{\mu}(t_\ell) \leftarrow \mathbf{0},\ \boldsymbol{\Sigma}(t_\ell) \leftarrow \mathbf{0}$
return ▷ return incidence and/or prevalence sample paths
  $\mathbf{N} = \{\mathbf{N}(t_0), \mathbf{N}(t_1), \ldots, \mathbf{N}(t_L)\},\ \mathbf{X} = \{\mathbf{X}(t_0),\ \mathbf{X}(t_1), \ldots, \mathbf{X}(t_\ell)\}$

run for 2.5 million iterations, following a tuning run of equal length, but each chain only manages to yield a effective sample sizes for R_0 and the infectious period duration in the low double digits. In contrast, the NCP yields effective sample sizes per–chain of between 500–700 for each of the model parameters in only 50,000 iterations, following a short tuning run of equal length. Figure 4.2 shows the traceplot for one of the MCMC chains, which clearly mixes better.

In each iteration of a DA MCMC algorithm, we alternate between updates to the latent path, conditional on the model parameters, and updates to the parameters, conditional on the latent path. Figure 4.3, which depicts the CP and NCP representations of an LNA path, provides some insight into why the NCP improves MCMC mixing. Under the CP (top plot), updates to $\boldsymbol{\theta}|\widetilde{\mathbf{N}}, \mathbf{Y}$ are made conditionally on a *fixed* LNA path. Therefore, proposed parameter values are accepted depending on whether they are concordant with the data *and* the current path. Even small perturbations to model parameters can result in shifts of the LNA transition densities (grey densities) that would render the current path (red points) unlikely under the proposal. In contrast, perturbations to parameters implicitly perturb the LNA path even as the LNA draws, $\mathbf{Z}$, are clamped to their current values.

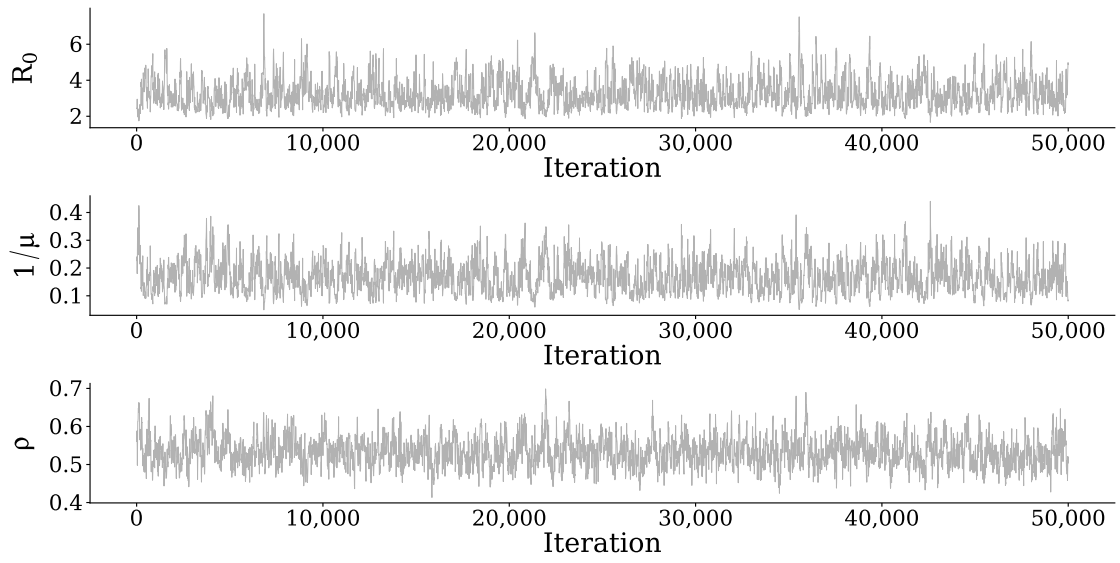

Figure 4.2: Posterior traceplots of parameters of interest sampled by a single MCMC chain of an SIR model fit to Poisson distributed incidence data. MCMC targeted the posterior, 4.20, alternately updating $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ via elliptical slice sampling, and $\boldsymbol{\theta}|\mathbf{Z}, \mathbf{Y}$ via a multivariate random walk Metropolis algorithm. $R_0 = \beta N/\mu$ is the basic reproductive number, $1/\mu$ is the mean infectious period duration, and ρ is the mean case detection rate. The true values of R_0, $1/\mu$, and ρ were 3.5, 7, and 0.5, respectively.

The NCP of the LNA also plays an important role in facilitating efficient updates of $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ via the elliptical slice sampling (EllipSS) algorithm of [37], which was detailed in Section 2.3.1 and is presented in Algorithm 2. EllipSS is an efficient and easy to implement MCMC algorithm for sampling latent Gaussian random variables, $\mathbf{Z}$, in models where the posterior of can be decomposed as the Gaussian prior for $\mathbf{Z}$ and an arbitrary likelihood, $L(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta})$, i.e.,

$$\pi(\boldsymbol{\theta}, \mathbf{Z}|\mathbf{Y}) \propto L(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}) MVN(\mathbf{Z}; \mu_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}}). \tag{4.21}$$

The target posterior under the LNA NCP, (4.20), is of this form, regardless of whether the LNA is restarted at the beginning of each inter–observation interval, as in [17], or the non–restarting version is used as in [34]. Note that the CP cannot be expressed as a jointly Gaussian collection of random variables with complete data likelihood of the form (4.21)

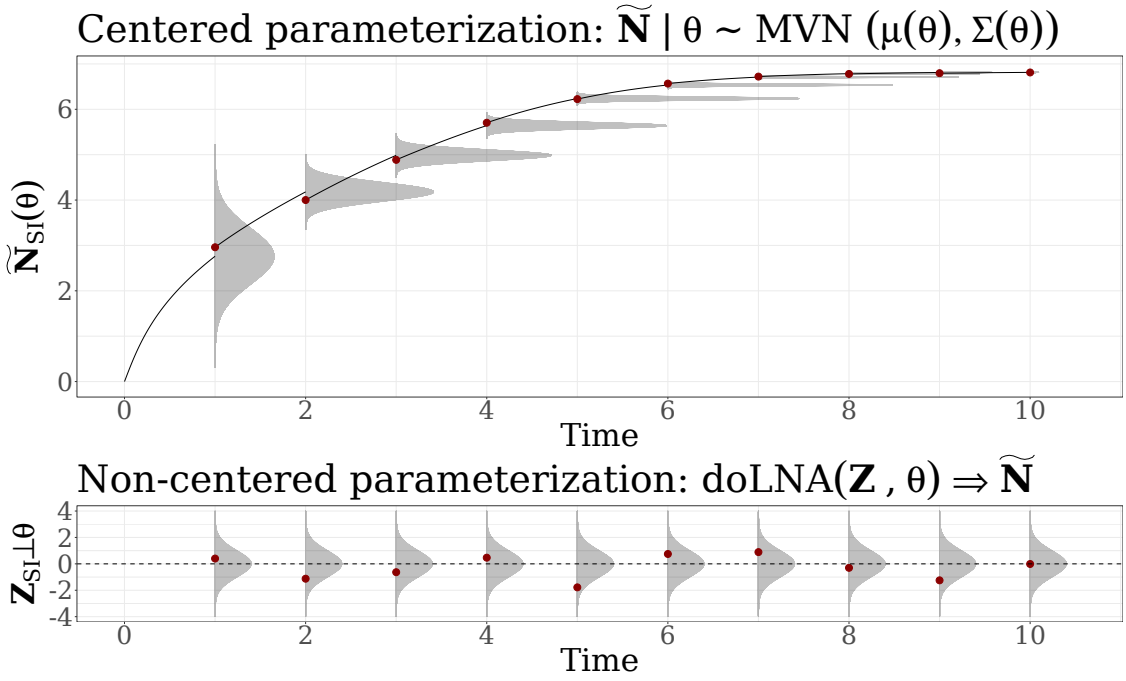

Figure 4.3: Centered (top) and non–centered (bottom) parameterizations of an LNA incidence path. In the CP, the log–incidence is normally distributed with mean and covariance obtained by solving the LNA ODEs, (4.15) and (4.17). In the NCP, the log–incidence is a draw from a standard normal distribution that is deterministically mapped to a sample path via the `doLNA` algorithm. In both the CP and NCP, state at the end of each interval determines the initial conditions of the LNA ODEs for the next interval. Plots of CP LNA transition densities are rescaled for clarity.

when we use the restarting version of the LNA. Although each transition density, (4.14), is itself Gaussian, the joint LNA path, $\mathbf{N}$, is not *a priori* Gaussian when the LNA ODEs are restarted since the mean of $\mathbf{N}(t_\ell)$ depends non–linearly on the value of $\mathbf{N}(t_{\ell-1})$. The quality of the LNA approximation is known to degenerate over long time intervals. Restarting the LNA ODEs has been established to improve the approximation when analyzing time series data of non–negligible length [17, 23]. Hence, use of the NCP is critical to enabling the use of EllipticSS for jointly updating of $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ when using the restarting version of the LNA.

We note that the elliptical slice sampling Algorithm 2 differs slightly from, but is equivalent to, the algorithm in [37] regarding how the initial proposal is made. The original

algorithm was modified so that the distribution of angles for accepted proposals would be symmetric about zero in order to facilitate tuning of the initial bracket width. We have found that shrinking the initial bracket width often improves computational efficiency when fitting more complex models. This is discussed further in Section A.1.

Algorithm 2 Sampling LNA draws via elliptical slice sampling.

procedure doElliptSS($\mathbf{Z}_{cur}, \boldsymbol{\theta}, \mathbf{Y}, \mathcal{I}, \omega = 2\pi$)

Sample ellipse: $\mathbf{Z}_{prop} \sim N(\mathbf{0}, \mathbf{I})$

Sample threshold: $u|\mathbf{x} \sim \text{Unif}(0, L(\mathbf{Y}|\texttt{doLNA}(\mathbf{Z}_{cur}, \boldsymbol{\theta}, \mathcal{I})))$

Position the bracket and make initial proposal:

$$\psi \sim \text{Unif}(0, \omega)$$

$$L_\psi \leftarrow -\psi;\ R_\psi \leftarrow L_\psi + \psi$$

$$\phi \sim \text{Unif}(L_\psi, R_\psi)$$

Set $\mathbf{Z}' \leftarrow \mathbf{Z}_{cur}\cos(\phi) + \mathbf{Z}_{prop}\sin(\phi)$.

if $L(\mathbf{Y}|\texttt{doLNA}(\mathbf{Z}', \boldsymbol{\theta}, \mathcal{I})) > u$ **then** accept $\mathbf{Z}'$

    return $\mathbf{Z}'$

else

    Shrink bracket and try a new angle:

    If: $\phi < 0$ **then:** $L_\phi \leftarrow \phi$ **else:** $R_\phi \leftarrow \phi$

    $\phi \sim \text{Unif}(L_\phi, R_\phi)$

    GoTo: 5

Initializing the LNA Draws

In simple models, reasonable parameter values will generally lead to valid LNA paths for initial $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$, i.e., paths that satisfy the monotonicity and positivity conditions, and thus have non–zero likelihood. However, this is not necessarily the case for complex models with many types of transition events, or when the time–series of incidence counts

is long. One option is to include a re–sampling step after line 6 in Algorithm 1, in which $\mathbf{Z}(t_\ell)$ is redrawn in place until the conditions for a valid path over the interval are met. It is important to note that such a procedure does not sample from the correct distribution since $\mathbf{Z}$ is not actually a truncated multivariate Gaussian. To correct for this, we will "warm–up" the LNA path with an initial run of `doElliptSS` iterations in which the likelihood only consists of the indicators for whether the path is valid. Finally, note that ElliptSS, or any valid MCMC algorithm for updating $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ for that matter, will never lead to an invalid LNA path being accepted if the current LNA draws and model parameters correspond to a valid path. Similarly, updates to model parameters conditional on the LNA draws will also preserve the validity of LNA paths.

Parameter Updates

Each MCMC iteration will consist of a number of ElliptSS updates, typically one but possibly 2–3 for complex models, followed by a set of parameter updates. We will generally use either a global adaptive random walk Metropolis algorithm (Algorithm 4 in [5]) or the adaptive non–isotropic Gaussian slice sampler (GSS) presented in Section 2.3.1. In models where the initial state, $\mathbf{X}_0$, is not fixed, we will assign the prior $\mathbf{X}_0 \sim TMVN_{\mathcal{S}_X^R}\left(N\mathbf{p}, N(\text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^T)\right)$, which is a truncated multivariate normal approximation to a multinomial with initial state probabilities, $\mathbf{p}$, constrained to the state space of compartment volumes. We update $\mathbf{X}_0$ jointly with the LNA path via ElliptSS. Additional details are presented in Section A.2.

We have found it helpful, for the purpose of assigning sensible priors to model parameters and for improving MCMC mixing and convergence, to parameterize the estimation scale on which the MCMC explores the parameter space in terms of how the parameters directly affect the model dynamics. For the SIR model, this might mean re-expressing the model parameters in terms of the basic reproductive number of an outbreak, $R0 = \beta N/\mu$, and the recovery rate, μ. Additionally, we would like our estimation scale to be unconstrained and therefore sample (and either accept or reject) values for $\log(R0)$ and $\log(\mu)$. The importance of appropriately parameterizing the estimation scale is discussed at length in Section A.3.

4.2.7 Implementation

The algorithms for approximate inference via the LNA and ODE models are implemented in the `stemr` `R` package, which can be available, along with vignettes for reproducing our results, from the following stable GitHub repository: `https://github.com/fintzij/stemr`. The implementation is flexible and provides facilities for specification of arbitrary SEM dynamics, a variety of emission probability distributions, and capabilities for accommodating time–varying covariates, time–varying parameters, and deterministic forcings. Computationally intensive operations are implemented in `C++` via `Rcpp` and `RcppArmadillo` [15, 16]. ODE integration functions are dynamically compiled in `C++` with the help of the `odeintr` R package [30] and ODEs can be integrated using a variety of methods available in the `Odeint C++` library [1]. Additional aspects of the implementation are discussed in Section A.6.

4.3 Simulations

4.3.1 Motivating Use of the LNA — Comparison with Common SEM Approximations

The LNA is by no means the only approximation of the transition density of the MJP representation of a SEM. In the following subsection, we will illustrate why the LNA is an attractive choice, balancing computational cost with fidelity of the MJP approximation. We benchmark the LNA against two commonly used approximations of the MJP: the deterministic approximation given by a system of deterministic ODEs that are the functional infinite population limit of the MJP [21], and a discrete–time approximation of the MJP using a multinomial modification of the τ–leaping algorithm (MMTL) [9] to simulate epidemic paths within a particle marginal Metropolis–Hastings (PMMH) framework [4]. The ODE approximation was chosen because of its ubiquity in the study of epidemic modeling, while the MMTL approximation in combination with PMMH was chosen because of a straightforward and general implementation in the popular `pomp` package in `R` [32]. Arguably, the MMTL approximation is somewhat closer to the original MJP than the LNA since it preserves the discreteness of the latent state space, while the ODE approximation, being deterministic, is

further removed from the MJP.

The fidelity of each approximation to the original MJP depends on the population size and the epidemic dynamics. In relative terms, outbreaks with explosive dynamics in large populations will tend to deviate less from their infinite population deterministic limits than outbreaks that occur in small populations, that are less contagious, or that are characterized by uncertainty in the probability and timing of a major outbreak. We fit SIR models to 500 datasets simulated under a range of SIR dynamics. Each dataset is simulated by drawing the model parameters from a set of prior distributions, simulating an outbreak via Gillespie's direct algorithm [24], and finally simulating the dataset as a negative binomial sample of the true incidence. Datasets arising from outbreaks that died off immediately were discarded and re–simulated, while datasets arising from outbreaks lasting longer than 50 epochs were truncated at 50 observations. SIR models were then fit via the LNA, ODE, and MMTL approximations under the priors from which the parameters were drawn. The simulation was repeated under three different regimes for the population size and the initial number of infected individuals, reflecting different levels of initial stochasticity in the epidemic trajectory. The priors and population sizes (Table 4.3.1) were chosen because they were typical of settings in which the methods might reasonably be applied, e.g., the population sizes are not so big that the outbreaks would evolve deterministically, nor so small that the approximations would be unreasonable. All individuals who were not initially infected were susceptible at the start of each outbreak (i.e., no individuals with pre-existing immunity). The population sizes and initial conditions were fixed at their true values. Hence, the only model misspecification was in the approximation used for the latent epidemic process. Additional results and details about the simulation setup are provided in Section A.4. We also performed four analogous supplementary simulations, with similar results, where we generated datasets under fixed parameter regimes (presented in Section A.5).

Table 4.1: Population sizes, initial conditions, and priors under which datasets were simulated. Five hundred datasets were simulated for each of the population size regimesEach outbreak was simulated from a MJP with SIR dynamics. The observed incidence was a negative binomial sample of the true incidence in each inter–observation interval.

	Regime 1	**Regime 2**	**Regime 3**
Population size (N)	10,000	50,000	250,000
Initial infecteds (I_0)	1	5	25

Parameter	**Interpretation**	**Prior**	**Median (95% Interval)**
$R0 - 1$	Basic reproduction # - 1	LogNormal(0, 0.25)	$\implies R0 = 2.00$ (1.38, 3.66)
$1/\mu$	Mean infectious period	LogNormal(-0.7, 0.35)	1.43 (0.72, 2.84)
$\rho/(1-\rho)$	Odds of case detection	LogNormal(0, 1)	$\implies \rho = 0.5$ (0.12, 0.88)
ϕ	Neg.Binom. overdispersion	Exponential(0.1)	6.93 (0.25, 36.89)

Results

This simulation was designed to be generous to the approximations that were used in fitting SEMs to the simulated data. The initial compartment counts and true population sizes were known, and there was no misspecification with respect to either the sampling model or the epidemic dynamics. Despite this, the ODE models struggle to reliably recover the true parameters, particularly those governing the sampling process. As shown in Figure 4.4, coverage of credible intervals for ODE models was low for all model parameters, and this was only somewhat mitigated as the population size increased. Coverage of credible intervals for models fit via the LNA and via MMTL was close to the nominal 95% level for all model parameters in all three population size regimes. Further inspection of the posterior median errors (middle row of Figure 4.4) and the widths of 95% credible intervals (bottom row of Figure 4.4) provides intuition for why the ODE performs so poorly. Estimates of the case detection probability tend to be high and estimates of the negative binomial over–dispersion parameter are low (corresponding to large variances in the conditional distribution

of observed incidence). Furthermore, credible intervals for the basic reproductive number and recovery rate obtained via the ODE approximation tend to be narrower than LNA and MMTL credible intervals. This is in agreement with findings by other authors who have found that ODE models tend to underestimate uncertainty in epidemic dynamics (see e.g., [31]). Taken together, these results suggest that the LNA is, at least in this simple example, about as good at approximating the original MJP as is the more exact MMTL.

We note that these results are not intended to suggest that there is no place for ODE models in the computational toolbox of disease modelers. To the contrary, when time is of the essence, as in an outbreak setting, crude estimates via the ODE may be obtained quickly. Average ODE run times were substantially shorter than LNA and MMTL run times and required far less CPU time per effective sample (see Table 4.3.1). ODE models are also appealing because they lend themselves to analytic characterizations of various aspects of the outbreak dynamics, e.g., relating the final outbreak size to the basic reproductive number (see, e.g., [3, 10, 29]).

In this simple simulation, the LNA and MMTL approximations had comparable computational performance, with the LNA perhaps being somewhat faster, but also with the caveat that comparing the ODE/LNA approximations with the MMTL approximation on the basis of computational performance is a bit misleading since the comparison would have turned out differently had we made other choices for the LNA and MCMC settings (e.g., timestep of MMTL, number of particles in PMMH, or tuning the initial EllipSS bracket width for the LNA). The important point to make regarding computational performance is that as model dynamics get more complex and the time series get longer, approximations, such as MMTL, that are used within a particle filter framework, such as PMMH, will become computationally infeasible. In many cases, the lack of an adequate model from which to simulate particle paths will lead to issues of particle degeneracy and an inability to fit even simple models (see [20] for an example). Indeed, PMMH was abandoned as a computational strategy for analyzing Ebola data in later sections because of difficulty fitting SEMs with reasonable effort. However, as we shall see in the following sections, the LNA remains performant even

as the model dynamics increase in complexity.

Table 4.2: Run times, effective sample sizes, and relative geometric mean (GM) log–posterior effective sample size (ESS) per CPU time for models fit via the ODE, LNA, and MMTL approximations. Run times and ESS are computed over all chains. The GM log–posterior ESS/CPU time was computed over the five chains for each model and divided by the corresponding GM ESS/CPU time for the MMTL model. We report 50% (2.5%, 97.5%) quantiles of the CPU time, ESS, and relative GM ESS/CPU time.

	ODE	**LNA**	**MMTL**
CPU time (minutes)	0.42 (0.23, 0.64)	27.78 (12.03, 56.25)	86.96 (40.47, 159.68)
Effective sample size	5745 (4557, 6616)	4067 (346, 11313)	6834 (3764, 11879)
GM ESS/CPU time vs. MMTL	180 (90, 350)	1.87 (0.14, 8.84)	—

4.3.2 Assessing Model Fit

Although model assessment is not the focus of this work, we emphasize the importance of diagnostics and briefly present several tools that can be used to interrogate SEMs fit via the LNA. We highlight the following diagnostics because they are easily implemented post-hoc and require little additional effort beyond caching random quantities throughout the MCMC run.

One of the central objectives in infectious disease modeling is to estimate the severity and duration of an outbreak [36]. However, the discrete and partial nature of the data makes it difficult to ascertain whether our estimates of the latent epidemic process reconstruct the true path of the outbreak (left plot, Figure 4.5). A standard alternative is to check whether the observed incidence is captured in the posterior predictive distribution.

The partial posterior predictive distribution, shown in the middle plot of Figure 4.5, is the predicted observed incidence, conditional on the outbreak being distributed according to the epidemic paths that we believe to be likely given the data at hand (i.e. the latent posterior). Here, new data are sampled using the posterior distribution of the observation

model conditionally on the paths in the latent posterior sample. Hence, for replication j, we simulate

$$Y_\ell^{(j)}|\mathbf{N}_{post}^{(j)}, \boldsymbol{\theta}_{post}^{(j)} \sim \text{Neg.Binom.}\left(\mu = \rho^{(j)}\left(N_{SI}^{(j)}(t_\ell) - N_{SI}^{(j)}(t_{\ell-1})\right), \sigma^2 = \mu + \mu^2/\phi^{(j)}\right), \tag{4.22}$$

where $\mathbf{N}_{post}^{(j)}$ and $\boldsymbol{\theta}_{post}$ are the j^{th} posterior samples for the latent path and model parameters.

The full posterior predictive distribution, shown in the right plot of Figure 4.5, is the predicted observed incidence, integrated over the predicted latent incidence given the current data. Under the full posterior predictive distribution, new data are sampled by simulating a new latent path under the posterior distribution of the epidemic model parameters, and then simulating a new dataset conditionally on the simulated path. So, for replication j, we simulate

$$\mathbf{Z}_{rep}^{(j)} \overset{i.i.d.}{\sim} N(0,1),\ \mathbf{N}_{rep}^{(j)} = \texttt{doLNA}(\mathbf{Z}_{rep}^{(j)}, \boldsymbol{\theta}_{post}^{(j)}, \mathcal{I}), \tag{4.23}$$

$$Y_\ell^{(j)}|\mathbf{N}_{rep}^{(j)}, \boldsymbol{\theta}_{post}^{(j)} \sim \text{Neg.Binom.}\left(\mu = \rho^{(j)}\left(N_{SI}^{(j)}(t_\ell) - N_{SI}^{(j)}(t_{\ell-1})\right), \sigma^2 = \mu + \mu^2/\phi^{(j)}\right), \tag{4.24}$$

where $\mathbf{Z}_{rep}^{(j)}$ is a new vector of LNA draws.

The full posterior predictive distribution provides a diagnostic for whether the joint model for the latent epidemic process and the observation process is reasonable as a data generating mechanism. In contrast, the partial posterior predictive distribution is useful for qualitatively assessing the adequacy of the emission distribution. Note that, when simulating from the full posterior predictive distribution, the replicated paths must still satisfy the constraints on the state space of $\mathbf{N}$ and $\mathbf{X}$. We can sample from the approximate posterior predictive distribution using the procedure described in Section 4.2.6 to avoid rejection sampling.

The second diagnostic, motivated by ideas presented in [35], leverages the structure of LNA draws as i.i.d. standard normal variates to identify time intervals over which the observed incidence suggests that dynamics of the latent process are not well approximated by the LNA prior. The normal CDF values of the LNA draws, $\mathbf{U}_\mathbf{Z} \equiv \Phi(\mathbf{Z}) \overset{i.i.d.}{\sim} \text{Unif}(0,1)$, can be understood as latent recursive residuals in the sense of [14] since each LNA increment is a deterministic function of the model parameters and the LNA draws, $\mathbf{Z}$. Suppose that

the emission distribution reasonably describes the sampling mechanism by which we the observed incidence is related to the true incidence. Then, the latent residuals for intervals where the latent process is well approximated by the LNA will be approximately uniformly distributed, whereas we would expect to see systematic deviations from uniformity when the LNA is a poor approximation. This is clearly the case in Figure 4.6, where quantiles of the latent residuals are clearly not uniform towards the start and end of the outbreak. This is not unexpected since the number of MJP transitions during these time periods are relatively low. The interpretation of this diagnostic as an indicator for misspecification in the latent epidemic process depends on the sampling model being approximately correct, and it is thus most useful in conjunction with other model diagnostics. Still, even when the emission distribution is unreasonable, e.g., if the case detection probability varies over time, we might be able to use the latent residuals to identify time intervals where the joint model is misspecified.

Finally, we can interrogate the validity of the SEM and whether it reflects the temporal dependencies in the data by computing summary statistics for the observed incidence and posterior predictive incidence. A variety of summary statistics are presented in the supplement of [31], which are computed and compared for the observed and predicted data. These include the lag 1 autocorrelation (ACF) and the detrended autocorrelations at lags 1, 2, and 3. We will also consider the partial autocorrelation function (PACF) [11] of the log–transformed data. We compute the PACF of $\widetilde{\mathbf{Y}} = \log(\mathbf{Y} + 0.5)$ at lag k as

$$\alpha(1) = \text{Cor}\left(\widetilde{\mathbf{Y}}_{(t)}, \widetilde{\mathbf{Y}}_{(t+1)}\right), \tag{4.25}$$

$$\alpha(k) = \text{Cor}\left(\widetilde{\mathbf{Y}}_{(t+k)} - \text{Proj}_{(t,k)}\left(\widetilde{\mathbf{Y}}_{(t+k)}\right),\ \widetilde{\mathbf{Y}}_{(t)} - \text{Proj}_{(t,k)}\left(\widetilde{\mathbf{Y}}_{(t)}\right)\right), \tag{4.26}$$

where the constant 0.5 is added to stabilize the logarithm for zero counts and $\text{Proj}_{(t,k)}(\widetilde{\mathbf{Y}}_{(j)})$ is the projection of $\widetilde{\mathbf{Y}}_{(j)}$ onto the span of $\widetilde{\mathbf{Y}}_{(t+1)}, \ldots, \mathbf{Y}_{(t+k-1)}$ (N.B. this is equivalent to regressing $\widetilde{\mathbf{Y}}_{(j)}$ on $\widetilde{\mathbf{Y}}_{(t+1)}, \ldots, \mathbf{Y}_{(t+k-1)}$). The PACF is a measure of the residual autocorrelation at a particular lag, adjusting for the linear dependencies on the intermediate variables. We should expect the PACF of the predicted data in the full posterior predictive distribution

to be distributed around zero for lags greater than one since the SEM is a first order Markov process. If the distributions of PACFs for predicted datasets in the partial predictive distribution are pulled away from zero for higher order lags, this may be indicative of long range correlation in the latent process that is unaccounted for by the SEM dynamics. Furthermore, we can compare the PACF for the observed and predicted data to see whether the model reflects the observed temporal dependence. Plots of the observed and predicted PACFs in Figure 4.7 do not suggest any model misspecification (as expected, since there was none).

4.3.3 A Stratified SEIR Model for a Simulated Outbreak

In the next section, we will jointly model the spread of Ebola in Guinea, Liberia, and Sierra Leone during the 2014–2015 outbreak. To do so, we will fit a stratified SEIR model allowing for country specific epidemic dynamics as well as interactions between the countries. It goes without saying that the model, despite being fully stochastic in all aspects of the measurement and epidemic process, is still an overly simplistic abstraction of the outbreak with respect to some obvious and potentially important complexities, e.g., spatial heterogeneity and time varying dynamics. However, it is a good starting point from a modeling perspective, and is highly non–trivial to fit due to the number of parameters and size of the latent space. Due to the myriad challenges involved in fitting the model, we will first assess its performance in recovering the true parameters and latent path for a simulated outbreak.

With an eye on jointly modeling

4.4 Application: Modeling the Spread of Ebola

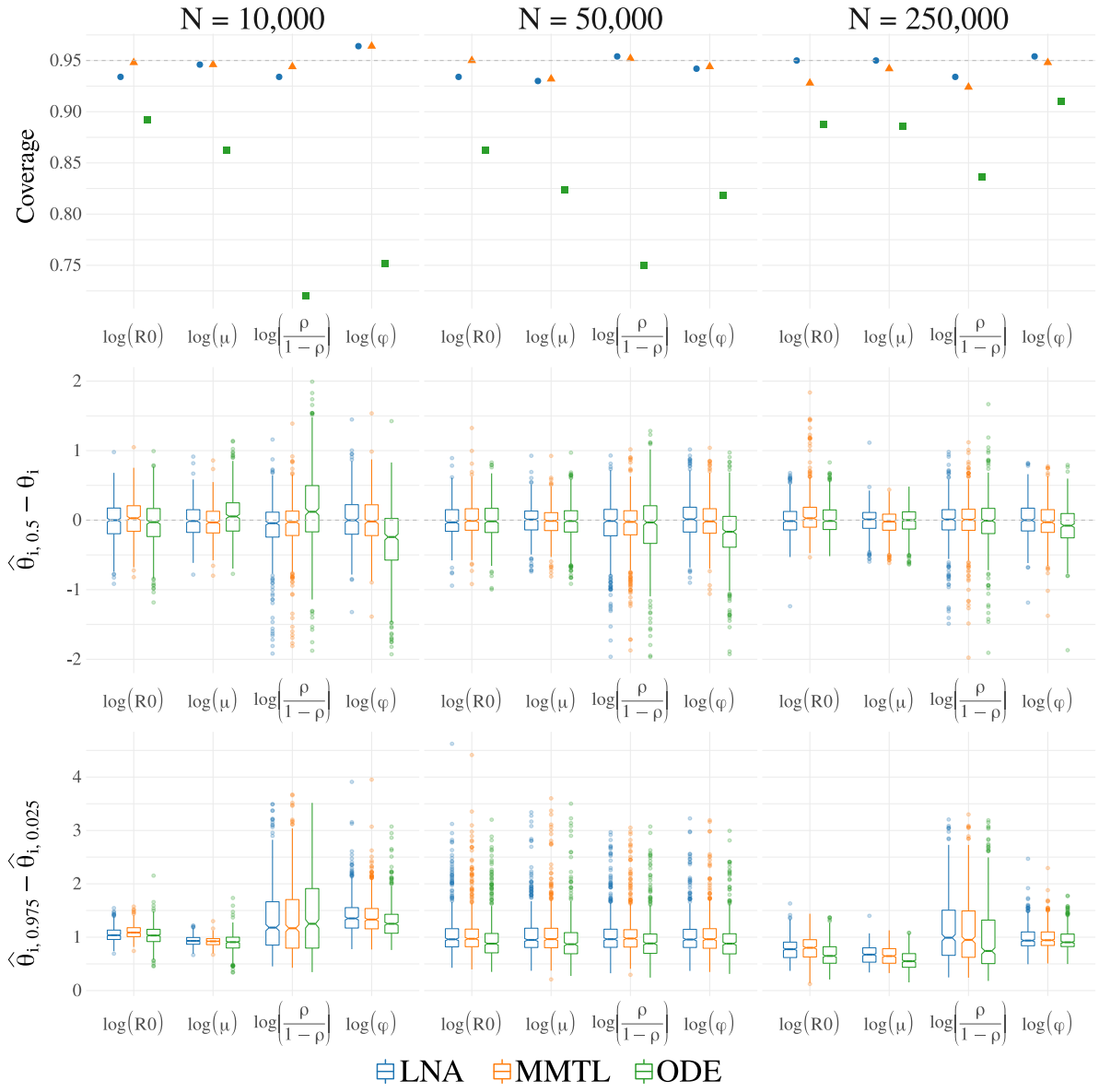

Figure 4.4: Comparison of SIR models fit to 500 datasets simulated in populations of three different sizes. Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over–dispersion parameter. The rows correspond to the proportion of runs where the 95% Bayesian credible interval covered the true parameter values, the posterior median deviation, and the widths of 95% Bayesian credible intervals. The simulation was repeated for three regimes of population sizes and initially infected individuals (columns).

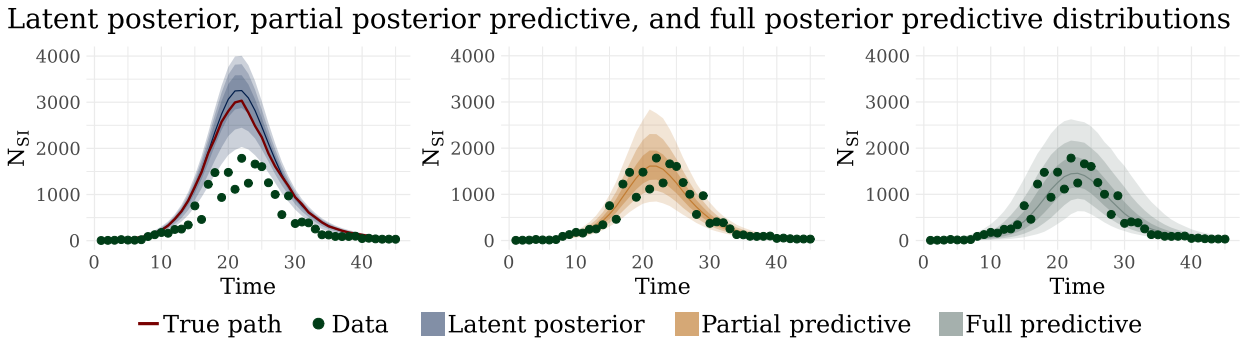

Figure 4.5: Estimated latent posterior (left), partial posterior predictive (middle), and full posterior predictive (right) distributions. The true path is the red line in the leftmost plot and the observed incidence counts are the solid green dots. The shaded bands, in order of lightest to darkest, correspond to pointwise posterior 95%, 80%, and 50% credible intervals, with the pointwise posterior median given by the corresponding colored solid line.

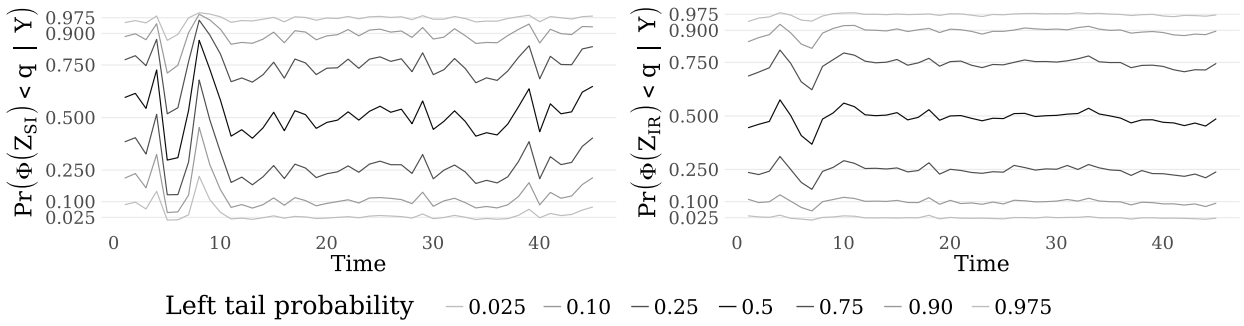

Figure 4.6: Posterior quantiles of latent recursive residuals for infections (left) and recoveries (right). Solid lines of varying shades correspond to the posterior quantiles, light horizontal lines in the background correspond to the theoretical quantiles under the prior.

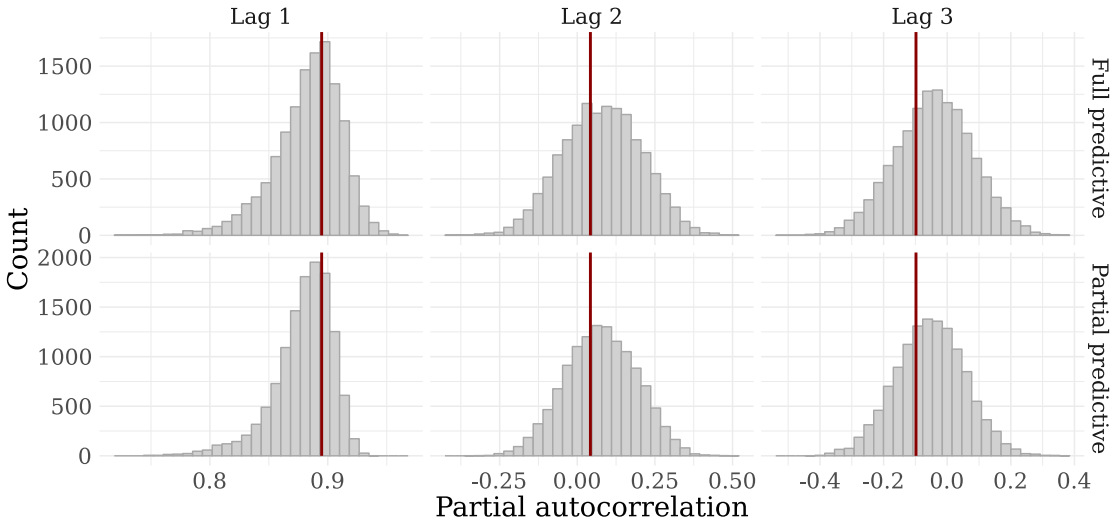

Figure 4.7: Distributions of partial autocorrelations at lags 1, 2, and 3 for datasets generated under the full and partial posterior predictive distributions. Vertical red lines are the partial autocorrelations for the observed incidence data at the respective lags.

Chapter 5

DYNAMIC TRANSMISSION MODELING OF PANDEMIC A(H1N1) INFLUENZA IN FINLAND

Chapter 6

DISCUSSION AND FUTURE WORK

BIBLIOGRAPHY

[1] K. Ahnert and M. Mulansky. Odeint–solving ordinary differential equations in C++. In *AIP Conference Proceedings*, volume 1389, pages 1586–1589. AIP, 2011.

[2] L.J.S. Allen. An introduction to stochastic epidemic models. In *Mathematical Epidemiology*, pages 81–130. Springer, New York, 2008.

[3] H. Andersson and T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis*. Lecture Notes in Statistics. Springer, New York, 2000.

[4] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72:269–342, 2010.

[5] C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Statistics and Computing*, 18:343–373, 2008.

[6] F. Bauer. Compartmental models in epidemiology. In *Mathematical Epidemiology*, pages 18–80. Springer, New York, 2008.

[7] J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and D. West. Non-centered parameterisations for hierarchical models and data augmentation. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, volume 307. Oxford University Press, USA, 2003.

[8] C.M. Bretó, D. He, E.L. Ionides, and A.A. King. Time series analysis via mechanistic models. *The Annals of Applied Statistics*, pages 319–348, 2009.

[9] C.M. Bretó and E.L. Ionides. Compound Markov counting processes and their applications to modeling infinitesimally over–dispersed systems. *Stochastic Processes and their Applications*, 121:2571–2591, 2011.

[10] T. Britton. Basic stochastic transmission models and their inference. *ArXiv e-prints*, January 2018.

[11] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2013.

[12] S.P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1998.

[13] E. Buckingham-Jeffery, V. Isham, and T. House. Gaussian process approximations for fast inference from infectious disease data. *Mathematical biosciences*, 2018.

[14] D.R. Cox and E.J. Snell. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 248–275, 1968.

[15] D. Eddelbuettel and R. François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40:1–18, 2011.

[16] D. Eddelbuettel and C. Sanderson. RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063, 2014.

[17] P. Fearnhead, V. Giagos, and C. Sherlock. Inference for reaction networks using the linear noise approximation. *Biometrics*, 70:457–466, 2014.

[18] J. Fintzi. *ECctmc: Simulation from endpoint–conditioned continuous time Markov chains*, 2016. R package, version 0.2.2.

[19] J. Fintzi. *stemr: Fit stochastic epidemic models via Bayesian data augmentation*, 2018. R package, version 0.2.1.

[20] J. Fintzi, X. Cui, J. Wakefield, and V.N. Minin. Efficient data augmentation for fitting stochastic epidemic models to prevalence data. *Journal of Computational and Graphical Statistics*, 26:918–929, 2017.

[21] C. Fuchs. *Inference for Diffusion Processes: With Applications in Life Sciences*. Springer Science & Business Media, New York, 2013.

[22] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472, 1992.

[23] V. Giagos. *Inference for auto-regulatory genetic networks using diffusion process approximations*. PhD dissertation, Lancaster University, 2010.

[24] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.

[25] D.T. Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113:297–306, 2000.

[26] A. Golightly and C.S. Gillespie. Simulation of stochastic kinetic models. In *In Silico Systems Biology*, pages 169–187. Springer, 2013.

[27] A. Golightly, D.A. Henderson, and C. Sherlock. Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing*, 25(5):1039–1055, 2015.

[28] L.S.T. Ho, F.W. Crawford, and M.A. Suchard. Direct likelihood-based inference for discretely observed stochastic compartmental models of infectious disease. *arXiv preprint arXiv:1608.06769*, 2016.

[29] M.J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, 2008.

[30] T.H. Keitt. *odeintr: C++ ODE Solvers Compiled on-Demand*, 2017. R package version 1.7.1.

[31] A.A. King, M.D. de Celles, F.M.G. Magpantay, and P. Rohani. Avoidable errors in the modeling of outbreaks of emerging pathogens, with special reference to Ebola. *Proceedings of the Royal Society, Series B*, 282:20150347, 2015.

[32] A.A. King, D. Nguyen, and E.L. Ionides. Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69:1–43, 2016.

[33] A.A. Koepke, I.M. Longini Jr., M.E. Halloran, J. Wakefield, and V.N. Minin. Predictive modeling of Cholera outbreaks in Bangladesh. *The Annals of Applied Statistics*, 10:575–595, 2016.

[34] M. Komorowski, B. Finkenstädt, C.V. Harper, and D.A. Rand. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*, 10:343, 2009.

[35] M.S.Y. Lau, G. Marion, G. Streftaris, and G.J. Gibson. New model diagnostics for spatio-temporal systems in epidemiology and ecology. *Journal of The Royal Society Interface*, 11:20131093, 2014.

[36] E.T. Lofgren, M.E. Halloran, C.M. Rivers, J.M. Drake, T.C. Porco, B. Lewis, W. Yang, A. Vespignani, J. Shaman, J.N.S. Eisenberg, M.C. Eisenberg, M. Marathe, S.V. Scarpino, K.A. Alexander, R. Meza, M.J. Ferrari, J.M. Hyman, L.A. Meyers, and S. Eubank. Opinion: Mathematical models: A key tool for outbreak response. *Proceedings of the National Academy of Sciences*, 111:18095–18096, 2014.

[37] I. Murray, R.P. Adams, and D.J.C. MacKay. Elliptical slice sampling. *JMLR: W&CP*, 9:541–548, 2010.

[38] P. Neal and G.O. Roberts. A case study in non-centering for data augmentation: stochastic epidemics. *Statistics and Computing*, 15:315–327, 2005.

[39] B. Øksendal. *Stochastic Differential Equations*. Springer, New York, 2003.

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

[41] O. Papaspiliopoulos, G.O. Roberts, and M. Sköld. Non-centered parameterisations for hierarchical models and data augmentation. *Bayesian Statistics*, 7:307–326, 2003.

[42] O. Papaspiliopoulos, G.O. Roberts, and M. Sköld. A general framework for the parametrization of hierarchical models. *Statistical Science*, pages 59–73, 2007.

[43] M. Plummer, N. Best, K. Cowles, and K. Vines. Coda: Convergence diagnosis and output analysis for MCMC. *R News*, 6:7–11, 2006.

[44] M.M. Tibbits, C. Groendyke, M. Haran, and J.C. Liechty. Automated factor slice sampling. *Journal of Computational and Graphical Statistics*, 23:543–563, 2014.

[45] E.W.J. Wallace, D.T. Gillespie, K.R. Sanft, and L.R. Petzold. Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET systems biology*, 6:102–115, 2012.

[46] D.J. Wilkinson. *Stochastic Modelling for Systems Biology*. CRC Press, Boca Raton, 2011.

[47] Y. Yu and X. Meng. To center or not to center: That is not the question — An Ancillarity–Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, 20:531–570, 2011.

Appendix A

APPENDIX TO CHAPTER 4

A.1 Tuning the Initial Elliptical Slice Sampling Bracket Width

When fitting SEMs with complex dynamics, e.g., when there are many strata or when the dynamics are time varying, we will be able improve the computational efficiency of our MCMC by initialing the EllipSS bracket width at $\omega < 2\pi$. This choice is motivated by the observation that when the model dynamics are complex, the EllipSS bracket will typically need to be shrunk many times before the sampler reaches a range of acceptable angles in the proposal. Each time we propose a new angle in the EllipSS algorithm we must solve the LNA ODEs in order to compute the observed data likelihood. Thus, if we can reduce the number of EllipSS steps, we will be able to shorten the time required to complete our MCMC runs.

In models where it is advantageous to shring the initial bracket width, we will typically set the initial bracket width to a constant times the standard deviation of the accepted angles in a tuning phase. Since we do not step out the EllipSS bracket, the initial width should not be so small as to induce additional autocorrelation in the latent process, and should also not be so wide that the bracket is contracted needlessly. We have found a bracket width of $\omega = 2\sqrt{2\log(10)}\sigma$, corresponding to the full width at one tenth maximum for a Gaussian with standard deviation σ, to work well in practice. Figure A.1 presents histograms of the number of contractions per EllipSS update and the accepted angles before and after contracting the initial EllipSS bracket width for the joint Ebola model of Section 4.4. In this instance, we were able to substantially reduce the number of contractions, and hence likelihood evaluations, per EllipSS update while leaving the distribution of accepted angles essentially unchanged. We like to call this a "free lunch".

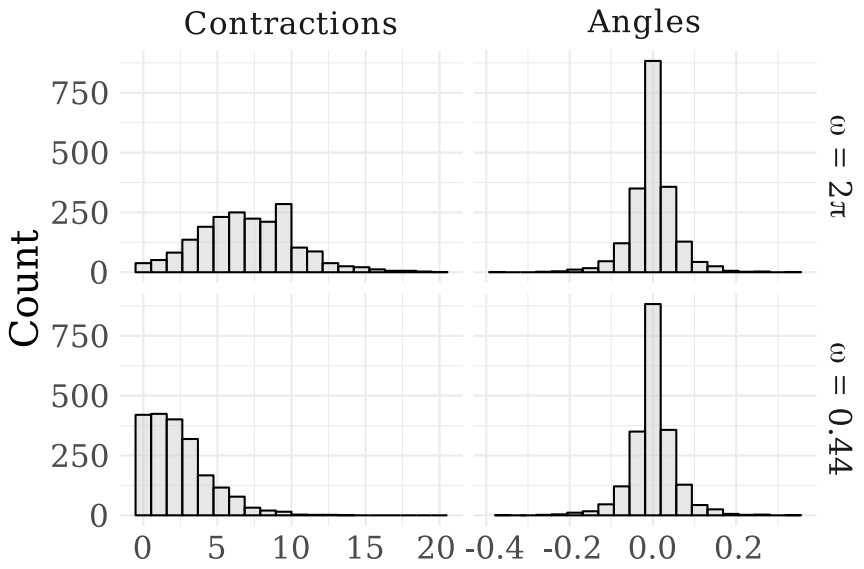

Figure A.1: Distributions of the numbers of contractions per ElliptSS update and the accepted angles for an MCMC chain for the joint Ebola model of Section 4.4 . An initial bracket width of 2π was used for the first 5,000 iterations (top row), after which the initial bracket width was set to $2\sqrt{2\log(10)}\sigma_{ElliptSS}$, where $\sigma_{ElliptSS}$ was the standard deviation of the accepted angles from the initial run (bottom row).

We mentioned in Section 4.2.6 that our EllipSS algorithm was modified slightly from that presented in [37] in order to facilitate tuning of the initial EllipSS bracket width. In both cases, the distribution of proposed states, $\mathbf{Z}_{prop}$, is centered at the current state $\mathbf{Z}_{cur}$. However, the distribution of angles of accepted states using our algorithm will be centered around 0, whereas the distribution of angles for accepted proposals using the algorithm in [37] will be bimodal with peaks at 0 and 2π. The algorithms are nevertheless equivalent due to the rotational symmetry of the proposals (a proposal made using an angle ϕ is equivalent to a proposal using $\phi + 2\pi$). It is more natural to compute the standard deviation of the accepted angles if the distribution of angles is symmetric about zero than if it is bimodal (and would thus require a rotation).

A.2 Inference for Initial Compartment Volumes

When the initial compartment volumes are included as initial parameters in the model instead of being treated as fixed, we will model them as arising from the following truncated multivariate normal distribution:

$$\mathbf{X}_0 \sim TMVN_{\mathcal{S}_X^R}(N\mathbf{p}, N(\mathbf{P} - \mathbf{p}\mathbf{p}^T)), \tag{A.1}$$

where $\mathbf{p}$ is a vector of subject–level initial state probabilities, $\mathbf{P} = \text{diag}(\mathbf{p})$, N is the population size, and the subscript $\mathcal{S}_X^R$ specifies the state space of $\mathbf{X}$ (so that the compartment volumes add up to N and each compartment volume is non–negative and less than the total population size at time t_0). Thus, the initial distribution is specified as the truncated normal approximation of a multinomial distribution with size N and probability vector $\mathbf{p}$. In models with multiple strata, we will similarly model the initial compartment volumes as having independent truncated multivariate normal distributions that are each approximations of multinomial distributions over initial compartment counts within each stratum. Notation and details are completely analogous to the single stratum case, and are therefore omitted for the sake of clarity.

Let $\mathbf{V} = N(\mathbf{P} - \mathbf{p}\mathbf{p}^T)$, and $\mathbf{V}^{1/2}$ be the matrix square root of $\mathbf{V}$, which we will compute

using the singular value decomposition $\mathbf{V} = \mathbf{U}\mathbf{D}\mathbf{U}^T \implies \mathbf{V}^{1/2} = \mathbf{U}\mathbf{D}^{1/2}$. Let $\mathbf{Z}^X$ denote the LNA draws as before, and let $\mathbf{Z}^{X_0} \sim \text{MVN}(\mathbf{0}, \mathbf{I})$ denote the vector of draws that will be mapped to $\mathbf{X}_0$. We will update the initial compartment volumes jointly with the LNA draws using elliptical slice sampling.

Algorithm 3 Sampling LNA draws and initial volumes via elliptical slice sampling.

procedure doEllipticalSS2($\mathbf{Z}^X_{cur}, \mathbf{Z}^{X_0}_{cur}, \boldsymbol{\theta}, \mathbf{Y}, \mathcal{I}, \omega = 2\pi$)

Sample ellipse: $\mathbf{Z}^X_{prop} \sim N(\mathbf{0}, \mathbf{I}),\ \mathbf{Z}^{X_0}_{prop} \sim N(\mathbf{0}, \mathbf{I})$

Sample threshold: $u|\mathbf{x} \sim \text{Unif}(0, L(\mathbf{Y}|\texttt{doLNA}(\mathbf{Z}_{cur}, \boldsymbol{\theta}, \mathcal{I})))$

Position the bracket and make initial proposal:

$$\psi \sim \text{Unif}(0, \omega)$$

$$L_\psi \leftarrow -\psi;\ R_\psi \leftarrow L_\psi + \psi$$

$$\phi \sim \text{Unif}(L_\psi, R_\psi)$$

Set $\mathbf{Z}^{X'} \leftarrow \mathbf{Z}^X_{cur}\cos(\phi) + \mathbf{Z}^X_{prop}\sin(\phi)$, $\mathbf{X}'_0 = \mathbf{V}^{1/2}\left(\mathbf{Z}^{X_0}_{cur}\cos(\phi) + \mathbf{Z}^{X_0}_{prop}\sin(\phi)\right)$

if $L(\mathbf{Y}|\texttt{doLNA}(\mathbf{Z}', \boldsymbol{\theta}', \mathcal{I})) > u$ **then** accept $\mathbf{Z}^{X'}, \mathbf{Z}^{X'_0}$

  return $\mathbf{Z}'$

else

  Shrink bracket and try a new angle:

  If: $\phi < 0$ **then:** $L_\phi \leftarrow \phi$ **else:** $R_\phi \leftarrow \phi$

  $\phi \sim \text{Unif}(L_\phi, R_\phi)$

  GoTo: 5

A.3 Choice of Estimation Scale and Implications for Mixing and Convergence

How we parameterize the MCMC estimation scale is critically important to its computational performance. If we can identify transformations of the model parameters that minimize strong correlations and non–linear relationships on the estimation scale, we will be able to substantially improve MCMC mixing. In our context, it will often be relatively straight-

forward to identify such transformations (or at least intermediate transformations that can be used in combination). As a general approach, we will try to identify transformations that reflect the ways in which model parameters jointly act on the model dynamics, and then a second set of transformations that remove any boundary conditions.

As an example, consider the single country SEIR model fit to the incidence data from Sierra Leone in Section 4.4. This model includes parameters for the external force of infection and the effective population size, which add complexity to the usual formulation of the SEIR dynamics as being entirely driven by endogenous contacts within a closed homogeneously mixing population. The model parameters on their natural scales are provided in Table A.3. Each of the model parameters has a clear marginal interpretation, but upon examining the pairwise scatterplots of the posterior (Figure A.2) it becomes obvious that the parameters interact in highly non–linear ways. We would encounter a variety of pathological computational problems if we were to naively parameterize the MCMC estimation scale without considering the ways in which the parameters interact to affect the dynamics. For example, it would be extremely difficult for any sampler that does not account for the curvature in the posterior, e.g., Hamiltonian Monte Carlo (HMC), to explore the parameter space. (An aside: we experimented with implementing the LNA in `Stan` and using HMC to sample the posterior, but repeatedly integrating the LNA ODEs along with their augmented sensitivity equations was prohibitively slow for even simple models).

Table A.1: SEIR model parameter and their interpretation on their natural scales.

Parameter	**Interpretation**	**Domain**
α	Rate of infectious contact from outside the population	$[0, \infty)$
β	Per–contact rate of infection within the population	$[0, \infty)$
ω	Rate of transition from $E \to I$	$[0, \infty)$
μ	Rate of transition from $I \to R$	$[0, \infty)$
ρ	Mean case detection probability	$[0, 1]$
ϕ	Negative binomial overdispersion parameter	$[0, \infty)$
N_{eff}	Effective population size	$[0, N]$

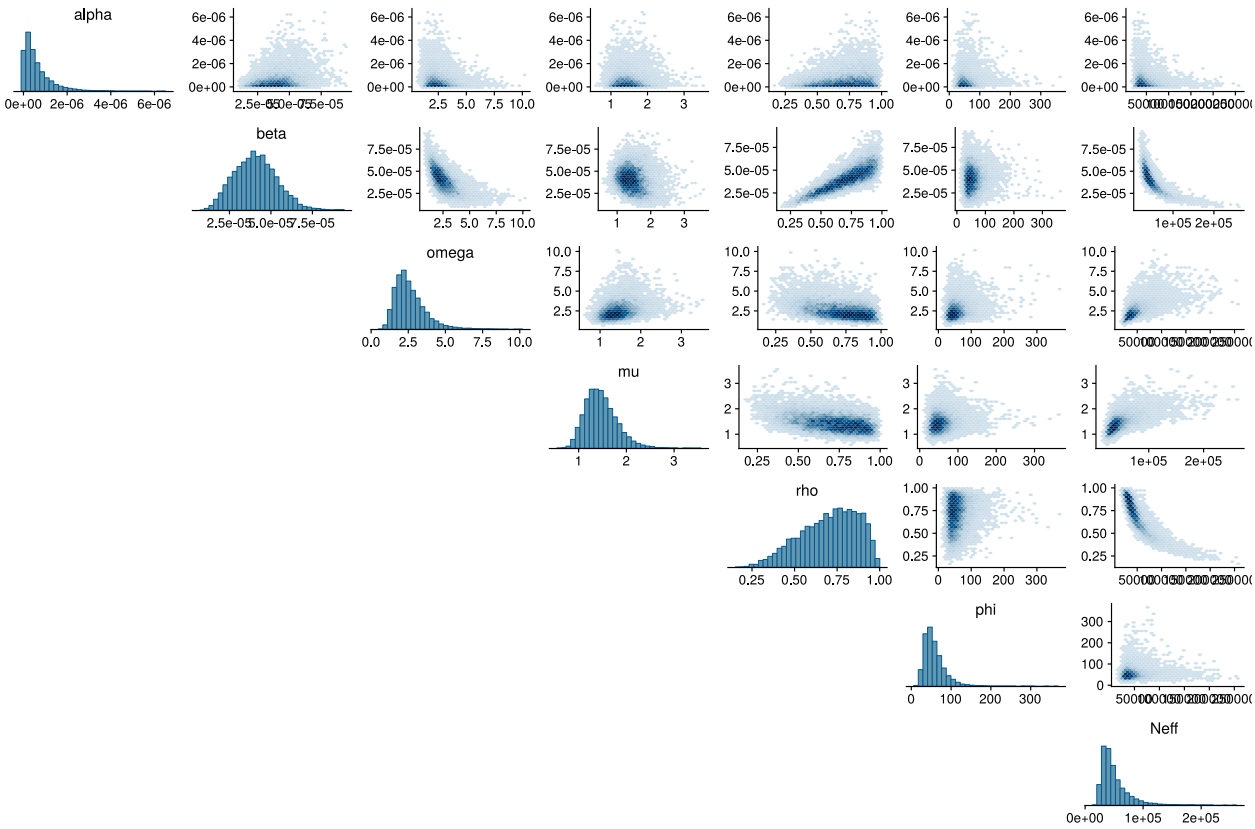

Figure A.2: Marginal histograms and pairwise scatterplots of posterior samples for parameters for the SEIR model fit to the Sierra Leone Ebola dataset using the estimation scale in Table A.3. The parameters on the estimation scales in this figure and their interpretations are provided in Table A.3.

We can mitigate the problems caused by non–linear relationships and strong correlations among parameters by parameterizing the estimation scale in terms of how the parameters jointly affect the model dynamics and then removing the boundary conditions. Table A.3 provides a list of parameters on their estimation scale that are reflective of an initial first pass at how we would expect the parameters to interact. For example, the parameters governing the rates of infectious contact, α and β, combine with the effective population size and the infectious period duration to produce the basic reproductive numbers with respect to initially infected individuals outside and inside the population. Still, we can see that there are some residual non–linear relationships between the log effective population size, the logit

case detection probability, and the effective reproductive number.

Table A.2: SEIR model parameter and their interpretation on a possible set of estimation scales.

Parameter	**Interpretation**	**Domain**
$\log(R_{eff}^{ext}) = \log(\alpha N_{eff}/\mu)$	Log basic reproductive number given an infected outside the population	$(-\infty, \infty)$
$\log(R_{eff} - 1) = \log(\beta N_{eff}/\mu - 1)$	Log basic reproductive number given an infected inside the population and $R_{eff} > 1$.	$(-\infty, \infty)$
$\log(1/\omega)$	Log mean latent period duration	$(-\infty, \infty)$
$\log(1/\mu)$	Log mean infectious period duration	$(-\infty, \infty)$
$\text{logit}(\rho)$	Logit mean case detection probability	$(-\infty, \infty)$
$\log(\phi)$	Log negative binomial overdispersion parameter	$(-\infty, \infty)$
$\log(N_{eff})$	Log effective population size	$(-\infty, \log(N))$

We should give additional consideration to how the previous functions of model parameters interact within the model. The effective population size, on its own, in this model is essentially a nuisance parameter. However, combined with the mean case detection probability, the mean case detection rate, ρN_{eff} should be, more or less, on the same scale as the total number of observed cases modulo the fraction of the effective population to escape infection. We note that estimation of this quantity is stable even when the population size is misspecified, see, e.g., [20, 33]. Moreover, we should suspect, *a priori*, that the basic reproductive number interacts with the mean case detection rate. To understand this, we examine how the basic reproductive number acts on the outbreak size through the final size relation for the deterministic ODE analog to our model [6]:

$$\log \frac{S_0}{S_\infty} = R0 \left(1 - \frac{S_\infty}{N}\right). \tag{A.2}$$

This equation relates the fraction of the population that eventually becomes infected with the basic reproductive number. As $R0$ increases, a larger fraction of the population becomes infected. If the effective population size is large, and if the $R0$ is high, the mean case detection

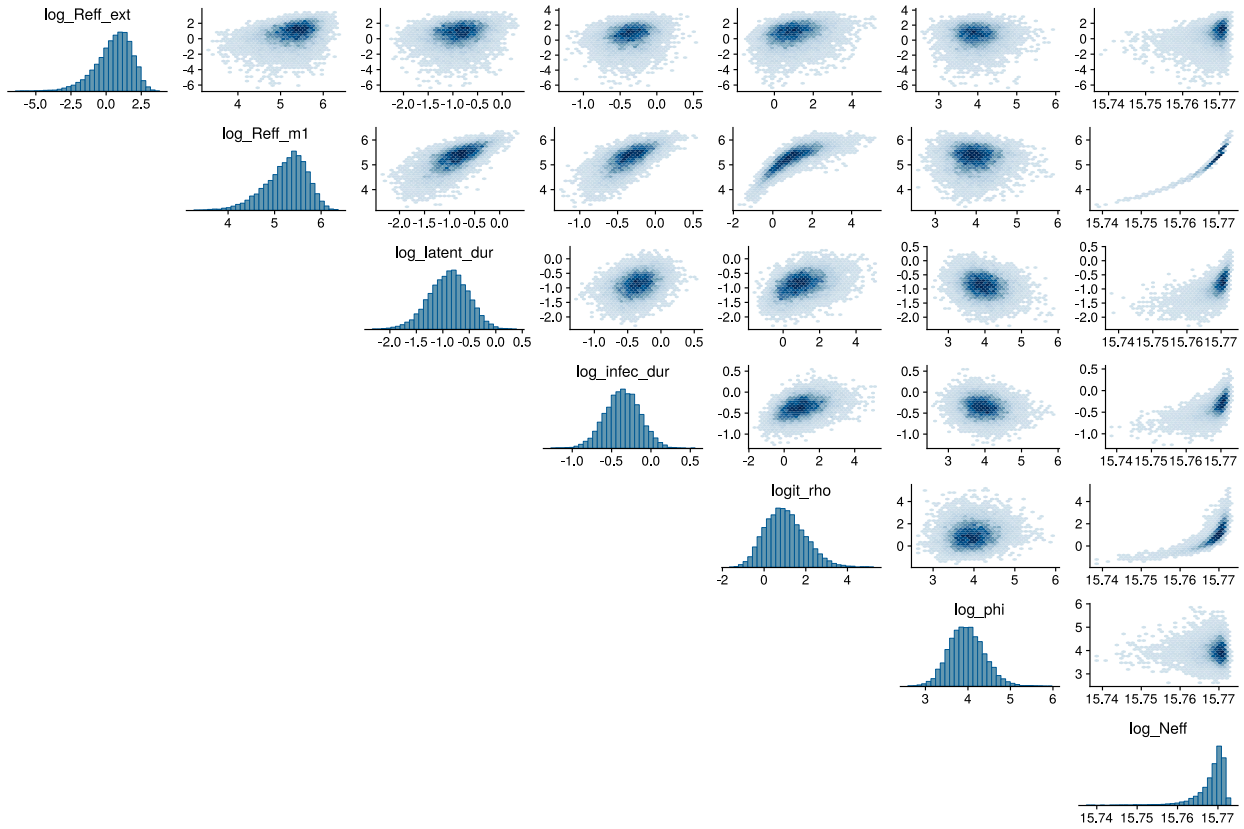

Figure A.3: Marginal histograms and pairwise scatterplots of posterior samples for parameters for the SEIR model fit to the Sierra Leone Ebola dataset using the estimation scale in Table A.3. The parameters on the estimation scales in this figure and their interpretations are provided in Table A.3.

probability should be low so that the mean case detection rate is concordant with the scale of the observed counts. This is all to suggest that the combination of parameters that jointly acts on the model is the effective reproductive number, offset by the mean case detection rate. The other reparameterization we suggest is to use the infectious period duration and the ratio of the latent to infectious period durations. The new estimation scale is given in Table A.3. On this estimation scale, the posterior for Sierra Leone is much better behaved, with weaker pairwise correlations and very little in the way of non–linear relationships between the model parameters.

Table A.3: SEIR model parameter and their interpretation on a possible set of estimation scales.

Parameter	**Interpretation**	**Domain**
$\log(R_{eff}^{ext}) = \log(\alpha N_{eff}/\mu)$	Log basic reproductive number given an infected outside the population	$(-\infty, \infty)$
$\log(R_{eff} - 1) + \log(\rho N_{eff})$	Log basic reproductive number given an infected inside the population and $R_{eff} > 1$, offset by the mean case detection rate	$(-\infty, \infty)$
$\log(\omega/\mu)$	Log ratio of mean latent to infectious period durations	$(-\infty, \infty)$
$\log(1/\mu)$	Log mean infectious period duration	$(-\infty, \infty)$
$\text{logit}(\rho)$	Logit mean case detection probability	$(-\infty, \infty)$
$\log(\phi)$	Log negative binomial overdispersion parameter	$(-\infty, \infty)$
$\log(\rho N_{eff})$	Log mean case detection rate	$(-\infty, \log(N))$

A.4 Simulation Details and Additional Results for Section 4.3.1

A.4.1 Simulation Setup and MCMC Details

In this simulation, repeated for each of the three different regimes of population size and initial conditions given in Table 4.3.1, we simulated 500 datasets according to the following procedure:

1. Draw $\log(R0 - 1),\ 1/\mu,\ \text{logit}(\rho),\ \log(\phi)$ from the priors given in Table 4.3.1.

2. Simulate an outbreak, $\mathbf{N}|\boldsymbol{\theta}$, under SIR dynamics from the MJP via Gillespie's direct algorithm [24]. If there were fewer than 15 cases, simulate another outbreak.

3. Simulate the observed incidence, $\mathbf{Y}|\mathbf{N}, \boldsymbol{\theta}$, as a negative binomial sample of the true incidence in each epoch, i.e., $Y_\ell \sim \text{Neg.Binomial}(\rho(\text{N}_{\text{SI}}(\text{t}_\ell) - \text{N}_{\text{SI}}(\text{t}_{\ell-1})), \phi)$. If the outbreak died off before epoch 15, the dataset was truncated at 15 observations (i.e., the dataset consisted of a series of case counts accrued during the outbreak along with

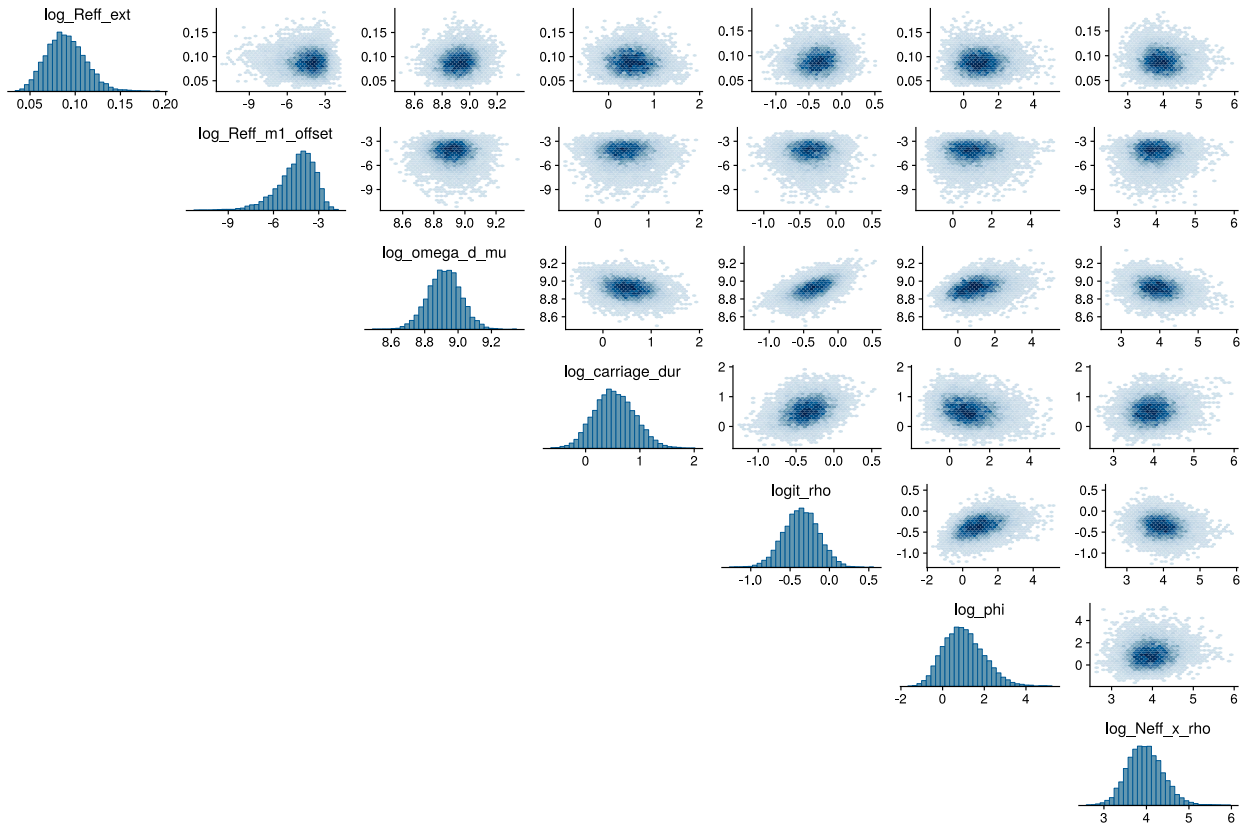

Figure A.4: Marginal histograms and pairwise scatterplots of posterior samples for parameters for the SEIR model fit to the Sierra Leone Ebola dataset using the estimation scale in Table A.3. The parameters on the estimation scales in this figure and their interpretations are provided in Table A.3.

a series of trailing zeros accrued after the outbreak died off). If the outbreak lasted longer than 50 epochs, the dataset was truncated at 50 observations

We proceed to fit SIR models using the LNA, ODE, and MMTL approximations. Priors for model parameters were assigned as in Table 4.3.1. Five MCMC chains per model were initialized at random values near the true parameters and run for 35,000 iterations per chain. The first 10,000 iterations used to warm up each chain and adaptively estimate the empirical covariance matrix to be used in the multivariate Gaussian random walk Metropolis–Hastings proposals for parameters. The empirical covariance matrix was initialized as 0.01 times an

identity matrix. After the warm–up period, the empirical covariance matrix was frozen and the final 25,000 iterations from each chain were combined to form the final MCMC sample. Convergence was assessed using potential scale reduction factors (PSRFs) [12], computed via the `coda` R package [43]. PSRFs were less than 1.05 in cases.

For models fit via the LNA and ODE approximations, the covariance matrix was adapted as in algorithm 4 of [5]. The gain factor sequence was $\gamma_n = 0.25(1+0.05n)^{0.50001}$, and a small nugget variance, equal to 0.00001 times an identity matrix, was added during the adaptation phase. The target acceptance rate used in the adaptation was 0.234. The models were implemented using the stemr R package [19].

Inference via the MMTL approximation within PMMH were fit using the `pomp` R package [32]. We used 500 particles in the PMMH algorithm. This choice was made to mitigate issues of particle degeneracy that occurred with fewer particles for some datasets. The time step for MMTL was set to 1/7, which, for example, corresponds to τ–leaping over one day increments given weekly incidence data. The MCMC was initialized in the same way as LNA and ODE models, but the empirical covariance matrix was adapted according to a different cooling schedule. The gain factor sequence provided by the package is $\gamma_n = n^{\alpha}$, where the cooling term, α, was set to 0.999. For some of the datasets, the PMMH algorithm degenerated during the adaptive phase of the MCMC. If this was the case, the MCMC was restarted at a different set of random initial conditions. The posterior sample consisted of the combined samples from all five MCMC chains after discarding the initial samples from the adaptation phase.

A.4.2 Additional Results

A.5 Supplementary Simulations with Fixed Parameters

A.5.1 Simulation Setup

The simulations presented in this section supplement the results of Section 4.3.1 in assessing the statistical and computation performance of the LNA approximation vis–a–vis the ODE and MMTL approximations. In contrast to the previous coverage simulation, here we will fix the model parameters to one of four regimes, presented in Table A.5.2, that are characterized by either fast or moderate outbreak dynamics, and high or low detection probability. In each setting, we simulated 500 outbreaks from a MJP with SIR dynamics in a population of 50,000 individuals, five of whom were initially infected and the rest of whom were susceptible. The observed incidence in each epoch was a negative binomial sample of the true incidence. Outbreaks for which the number of observed cases was less than 25 were re–simulated. MCMC chains were tuned and SIR models were fit via the LNA, ODE, and MMTL approximations as described in Section A.4.1. The initial comparment volumes were fixed at the true values. The models were fit using diffuse priors, also presented in Table A.5.2. We caution that the priors used in this exercise are perhaps unreasonably diffuse, particularly in the context of epidemic modeling where prior information is often available, and that we expect credible intervals will be overly wide as a result (we would still expect nominal coverage to be incorrect, even under tighter priors, since the datasets were simulated under fixed parameter regimes).

A.5.2 Results

Coverage for credible intervals of ODE models tended to fall below nominal levels in spite of the bias towards wide intervals due to the diffusivity of the priors. This was particularly the case in parameter regimes 1 and 3, where the basic reproductive number was lower (and hence the simulated outbreak trajectories further from their thermodynamic limits). In these parameter regimes, coverage was particularly poor due as estimates of the outbreak dynamics tended to be farther from their true values and credible intervals were too tight

and did not properly account for uncertainty about the parameter estimates, particularly those governing the measurement process. Coverage levels for models fit via the LNA and MMTL approximations exceeded their nominal levels as expected.

ODE models remained the most computationally performant. However, in this exercise, the LNA substantially outperformed the MMTL approximation within PMMH in terms of ESS and ESS per CPU time. We believe this is largely attributable to the diffusivity of the priors, which not only fail to regularize the posterior, but likely pull it towards unreasonable regions of the parameter space. As a general comment, we would strongly caution practitioners against adopting such priors more broadly. While it may seem appealing to adopt such diffuse priors in pursuit of being "agnostic" to the underlying outbreak dynamics, one of the very good reasons for working within the Bayesian paradigm in this context is that we have quite a bit of prior information regarding the outbreak dynamics and reasonable ranges for the case detection probability. For example, we often have historic examples of outbreaks in similar settings that we can look to in specifying priors about the basic reproductive number.

A.5.3 Results

A.6 LNA Implementation Details and LNA Model Vignettes

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	log(R0)	0.93	0 (-0.51, 0.55)	1.04 (0.83, 1.36)	1340 (274, 4029)	0.63 (0.13, 3.4)
LNA	log(μ)	0.95	-0.01 (-0.5, 0.48)	0.93 (0.74, 1.13)	1024 (199, 3988)	0.48 (0.1, 2.91)
LNA	logit(ρ)	0.93	-0.05 (-1.44, 0.56)	1.18 (0.54, 2.86)	1225 (316, 3941)	0.62 (0.17, 4.05)
LNA	log(ϕ)	0.96	0 (-0.66, 0.71)	1.35 (0.92, 2.22)	2021 (687, 4535)	1 (0.31, 3.53)
MMTL	log(R0)	0.95	0.03 (-0.48, 0.55)	1.09 (0.88, 1.39)	7483 (5453, 9197)	—
MMTL	log(μ)	0.95	-0.03 (-0.51, 0.45)	0.92 (0.75, 1.09)	7481 (5442, 9197)	—
MMTL	logit(ρ)	0.94	-0.03 (-1, 0.61)	1.17 (0.51, 2.98)	6725 (4328, 8506)	—
MMTL	log(ϕ)	0.96	-0.02 (-0.69, 0.71)	1.33 (0.91, 2.21)	7486 (5405, 9215)	—
ODE	log(R0)	0.89	-0.03 (-0.76, 0.56)	1.03 (0.65, 1.38)	6438 (4956, 7650)	185 (105, 340)
ODE	log(μ)	0.86	0.05 (-0.51, 0.79)	0.91 (0.53, 1.2)	6420 (4908, 7663)	184 (103, 344)
ODE	logit(ρ)	0.72	0.12 (-1.03, 1.37)	1.25 (0.48, 2.94)	6237 (4346, 7556)	203 (111, 401)
ODE	log(ϕ)	0.75	-0.25 (-1.64, 0.54)	1.25 (0.88, 2.11)	6529 (5459, 7778)	189 (116, 340)

Table A.4: Detailed small population (N = 10,000) regime results for the coverage simulation presented in Section 4.3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over–dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	log(R0)	0.93	-0.03 (-0.5, 0.49)	0.93 (0.66, 1.28)	2158 (397, 4981)	0.67 (0.16, 3)
LNA	log(μ)	0.93	0.01 (-0.43, 0.41)	0.83 (0.58, 1.07)	1919 (307, 5276)	0.63 (0.12, 3.05)
LNA	logit(ρ)	0.95	-0.01 (-1, 0.5)	1.13 (0.47, 2.81)	1704 (439, 4636)	0.61 (0.18, 3.49)
LNA	log(ϕ)	0.94	0.01 (-0.57, 0.7)	1.1 (0.79, 1.86)	2916 (1179, 5442)	1.02 (0.38, 3.1)
MMTL	log(R0)	0.95	-0.01 (-0.46, 0.51)	0.98 (0.67, 1.32)	7284 (4713, 9107)	—
MMTL	log(μ)	0.93	-0.01 (-0.47, 0.37)	0.82 (0.58, 1.05)	7167 (4488, 8912)	—
MMTL	logit(ρ)	0.95	-0.03 (-0.98, 0.56)	1.09 (0.43, 2.96)	6452 (3999, 8318)	—
MMTL	log(ϕ)	0.94	-0.02 (-0.59, 0.65)	1.1 (0.79, 1.84)	7096 (4640, 8999)	—
ODE	log(R0)	0.86	-0.02 (-0.56, 0.59)	0.84 (0.52, 1.27)	6592 (5213, 7771)	187 (88, 354)
ODE	log(μ)	0.82	-0.01 (-0.66, 0.47)	0.73 (0.44, 1.07)	6558 (5129, 7664)	190 (87, 359)
ODE	logit(ρ)	0.75	-0.03 (-1.13, 0.85)	0.88 (0.37, 2.75)	6421 (5017, 7643)	209 (107, 410)
ODE	log(ϕ)	0.82	-0.17 (-1.36, 0.57)	1.04 (0.78, 1.7)	6637 (5452, 7755)	193 (102, 365)

Table A.5: Detailed medium population (N = 50,000) regime results for the coverage simulation presented in Section 4.3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over–dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	log(R0)	0.95	-0.01 (-0.38, 0.51)	0.77 (0.47, 1.24)	3248 (638, 6127)	1.13 (0.22, 3.99)
LNA	log(μ)	0.95	0.01 (-0.44, 0.33)	0.67 (0.4, 1.02)	3134 (499, 6038)	1.12 (0.18, 3.95)
LNA	logit(ρ)	0.93	0.01 (-0.8, 0.57)	0.99 (0.35, 2.66)	2514 (572, 5905)	1.03 (0.24, 4.71)
LNA	log(ϕ)	0.95	0 (-0.44, 0.54)	0.94 (0.67, 1.5)	3987 (2201, 6184)	1.63 (0.71, 3.95)
MMTL	log(R0)	0.93	0.03 (-0.38, 1.04)	0.8 (0.31, 1.27)	7030 (3789, 8915)	—
MMTL	log(μ)	0.94	-0.02 (-0.46, 0.32)	0.65 (0.38, 0.98)	6858 (3602, 8871)	—
MMTL	logit(ρ)	0.92	0.01 (-0.7, 0.65)	0.95 (0.34, 2.64)	6166 (3282, 7906)	—
MMTL	log(ϕ)	0.95	-0.03 (-0.51, 0.52)	0.94 (0.66, 1.5)	6266 (3735, 8960)	—
ODE	log(R0)	0.89	-0.01 (-0.37, 0.55)	0.65 (0.36, 1.22)	6828 (5566, 8025)	193 (104, 417)
ODE	log(μ)	0.89	0 (-0.49, 0.34)	0.55 (0.3, 1)	6815 (5520, 7877)	198 (105, 428)
ODE	logit(ρ)	0.84	-0.01 (-0.78, 0.68)	0.74 (0.27, 2.62)	6575 (5219, 7838)	210 (118, 486)
ODE	log(ϕ)	0.91	-0.08 (-0.59, 0.46)	0.9 (0.64, 1.44)	6721 (5571, 7683)	209 (117, 413)

Table A.6: Detailed large population (N = 250,000) regime results for the coverage simulation presented in Section 4.3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over–dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Population size	ODE	LNA	MMTL
10,000	0.39 (0.21, 0.62)	21.73 (10.83, 37.74)	85.23 (42.31, 152.48)
50,000	0.42 (0.23, 0.62)	32.27 (13.4, 55.8)	88.36 (38.63, 153.54)
250,000	0.45 (0.25, 0.78)	33.08 (12.56, 70.86)	87.4 (39.8, 166.87)

Table A.7: Median (2.5%, 97.5%) quantiles of run times, in minutes, for MCMC chains in the coverage simulation presented in Section 4.3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE).

Table A.8: Parameter regimes under which datasets were simulated and priors used to fit SIR models. Five hundred datasets were simulated for each of the parameter regimes from a MJP with SIR dynamics. $R0 = \beta N/\mu$ is the basic reproductive number and μ is the recovery rate. The observed incidence was a negative binomial sample of the true incidence in each inter–observation interval with case detection probability ρ and overdispersion parameter ϕ.

	Regime 1	**Regime 2**	**Regime 3**	**Regime 4**
	Low R0/Low ρ	High R0/Low ρ	Low R0/High ρ	High R0/High ρ
R0	1.75	3.25	1.75	3.25
$\boldsymbol{\rho}$	0.25	0.25	0.75	0.75
μ	1	0.4	1	0.4
ϕ	5	5	5	5

Parameter	**Interpretation**	**Prior**	**Median (95% Interval)**
$R0 - 1$	Basic reproduction # - 1	LogCauchy(0.4, 1)	$\implies R0 = 2.50$ (1.00, 4.9×10^5)
$1/\mu$	Mean infectious period	LogCauchy(-0.7, 1)	1.43 (4.3×10^{-6}, 4.7×10^5)
ρ	Mean case detection prob.	Unif(0, 1)	0.5 (0.025, 0.975)
ϕ	Neg.Binom. overdispersion	LogCauchy(1.5,1)	4.48 (1.4×10^{-5}, 1.5×10^6)

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	log(R0)	0.99	0.23 (-0.33, 0.85)	1.9 (1.2, 3.49)	1200 (251, 2508)	19.1 (2.7, 76.4)
LNA	log(μ)	0.98	-0.21 (-0.81, 0.28)	1.73 (1.04, 3.42)	1086 (228, 2295)	27.9 (3.7, 100.5)
LNA	logit(ρ)	0.98	-0.06 (-0.46, 0.4)	1.07 (0.73, 1.67)	1233 (360, 2180)	8.22 (2.03, 21.4)
LNA	log(ϕ)	0.98	0.01 (-0.51, 0.74)	1.32 (1.17, 1.71)	2950 (1145, 4334)	10.8 (2.1, 30.0)
MMTL	log(R0)	0.99	0.22 (-0.86, 0.87)	4.4 (1.92, 53.26)	232 (110, 445)	—
MMTL	log(μ)	1.00	-0.22 (-0.79, 0.52)	2.02 (1.26, 3.63)	125 (58, 283)	—
MMTL	logit(ρ)	0.99	-0.06 (-0.47, 0.48)	1.21 (0.83, 1.72)	492 (258, 874)	—
MMTL	log(ϕ)	0.97	-0.01 (-0.54, 0.71)	1.32 (1.16, 1.72)	942 (541, 1791)	—
ODE	log(R0)	0.84	0.3 (-0.77, 2.43)	1.78 (1.12, 5.93)	3364 (252, 6513)	3894 (169, 13949)
ODE	log(μ)	0.78	-0.24 (-2.57, 0.7)	1.56 (0.93, 5.81)	3307 (243, 6489)	6454 (277.52, 15693)
ODE	logit(ρ)	0.78	-0.08 (-0.67, 0.79)	0.77 (0.51, 1.52)	5225 (2398, 6933)	2500 (869, 5719)
ODE	log(ϕ)	0.94	-0.1 (-0.73, 0.57)	1.26 (1.14, 1.54)	5702 (2768, 6991)	1451 (532, 3117)

Table A.9: Detailed results for the fixed parameter simulation in which outbreaks and datasets were simulated under parameter regime 1, characterized by slow outbreak dynamics (R0 = 1.75) and low mean case detection probability (ρ = 0.25). Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over–dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	log(R0)	0.99	-0.35 (-0.87, -0.05)	2.01 (1.42, 3.03)	1551 (399, 3065)	64.2 (10.3, 265.1)
LNA	log(μ)	0.99	0.34 (0.09, 0.8)	1.9 (1.34, 2.92)	1431 (368, 2865)	23.8 (5.0, 101.7)
LNA	logit(ρ)	0.94	0.12 (-0.23, 0.55)	1.02 (0.73, 1.71)	1354 (509, 2357)	13.1 (3.2, 40.6)
LNA	log(ϕ)	0.96	0.02 (-0.53, 0.81)	1.38 (1.25, 1.76)	3286 (1756, 4442)	19.5 (6.3, 50.5)
MMTL	log(R0)	0.99	-0.41 (-1.11, -0.08)	13.94 (2.95, 55.97)	132 (31, 314)	—
MMTL	log(μ)	0.99	0.4 (0.12, 0.99)	2.74 (2.13, 3.69)	213 (90, 1944)	—
MMTL	logit(ρ)	0.90	0.18 (-0.18, 0.64)	1.78 (1.1, 2.73)	375 (171, 737)	—
MMTL	log(ϕ)	0.97	-0.03 (-0.59, 0.78)	1.42 (1.25, 2)	586 (315, 993)	—
ODE	log(R0)	0.99	-0.22 (-0.84, 0.65)	1.85 (1.31, 3.58)	3979 (1875, 5747)	9412 (2368, 34873)
ODE	log(μ)	0.98	0.21 (-0.78, 0.8)	1.72 (1.21, 3.53)	3976 (1851, 5752)	3824 (9812, 12367)
ODE	logit(ρ)	0.85	0.05 (-0.34, 0.52)	0.71 (0.5, 1.08)	5024 (2846, 6390)	2753 (1205, 6858)
ODE	log(ϕ)	0.95	-0.03 (-0.64, 0.67)	1.33 (1.23, 1.6)	5634 (4377, 6661)	2027 (995, 4479)

Table A.10: Detailed results for the fixed parameter simulation in which outbreaks and datasets were simulated under parameter regime 2, characterized by fast outbreak dynamics (R0 = 3.25) and low mean case detection probability ($\rho = 0.25$). Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over–dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	log(R0)	0.96	0.26 (-0.22, 0.93)	1.68 (1.01, 3.21)	1354 (198, 3054)	16.0 (1.6, 70.2)
LNA	log(μ)	0.96	-0.23 (-0.85, 0.16)	1.49 (0.84, 3.23)	1210 (191, 2931)	25.1 (2.8, 83.6)
LNA	logit(ρ)	0.99	-0.2 (-0.96, 0.84)	3.1 (1.61, 4.6)	935 (393, 1893)	6.9 (2.4, 18.7)
LNA	log(ϕ)	0.97	0.02 (-0.52, 0.67)	1.25 (1.12, 1.52)	2851 (1425, 4283)	13.4 (4.3, 31.6)
MMTL	log(R0)	0.99	0.24 (-0.48, 0.96)	2.49 (1.44, 36.95)	313 (139, 634)	—
MMTL	log(μ)	0.99	-0.24 (-0.87, 0.24)	1.75 (1, 3.44)	141 (57, 359)	—
MMTL	logit(ρ)	1.00	-0.2 (-0.93, 0.86)	3.48 (2.04, 4.99)	439 (244, 711)	—
MMTL	log(ϕ)	0.97	-0.02 (-0.53, 0.64)	1.25 (1.12, 1.56)	723 (372, 1376)	—
ODE	log(R0)	0.79	0.34 (-0.41, 2.6)	1.52 (0.95, 5.73)	3688 (285, 6453)	2951 (127, 10756.39)
ODE	log(μ)	0.77	-0.28 (-2.73, 0.36)	1.31 (0.8, 5.69)	3650 (263, 6457)	5349 (177, 14237)
ODE	logit(ρ)	0.80	-0.19 (-1.41, 1.57)	2.27 (0.9, 4.66)	3418 (1896, 5721)	1767 (704, 4279)
ODE	log(ϕ)	0.91	-0.15 (-0.75, 0.48)	1.16 (1.08, 1.34)	5787 (2679, 7035)	1754 (640, 4124)

Table A.11: Detailed results for the fixed parameter simulation in which outbreaks and datasets were simulated under parameter regime 3, characterized by slow outbreak dynamics (R0 = 1.75) and high mean case detection probability (ρ = 0.75). Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over–dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	log(R0)	1.00	-0.27 (-0.68, 0.04)	1.7 (1.24, 2.64)	2217 (504, 4142)	68.6 (10.0, 342.8)
LNA	log(μ)	1.00	0.25 (-0.01, 0.62)	1.62 (1.18, 2.63)	2014 (493, 3894)	26.4 (4.7, 85.4)
LNA	logit(ρ)	0.98	0.25 (-0.58, 1.35)	3.49 (2.06, 4.63)	1290 (563, 2572)	14.8 (4.9, 59.9)
LNA	log(ϕ)	0.96	0.03 (-0.52, 0.76)	1.31 (1.2, 1.62)	3506 (1910, 4689)	27.0 (11.2, 66.1)
MMTL	log(R0)	1.00	-0.35 (-0.93, -0.04)	13.37 (1.83, 46.16)	181 (48, 409)	—
MMTL	log(μ)	1.00	0.33 (0.06, 0.83)	2.65 (1.61, 3.72)	272 (128, 1994)	—
MMTL	logit(ρ)	0.96	0.4 (-0.49, 1.57)	4.49 (3.15, 7.21)	309 (133, 545)	—
MMTL	log(ϕ)	0.97	-0.06 (-0.63, 0.64)	1.62 (1.26, 2.61)	457 (213, 848)	—
ODE	log(R0)	0.99	-0.18 (-0.78, 0.84)	1.67 (1.14, 3.73)	4463 (1987, 6336)	7879 (1915, 29567)
ODE	log(μ)	0.99	0.18 (-0.99, 0.73)	1.56 (1.07, 3.67)	4424 (1952, 6312)	3254 (867, 8316)
ODE	logit(ρ)	0.89	0.13 (-0.86, 1.71)	2.59 (1.1, 4.57)	3378 (1971, 5315)	2271 (916, 7206)
ODE	log(ϕ)	0.94	-0.07 (-0.66, 0.62)	1.26 (1.17, 1.48)	5697 (4518, 6929)	2490 (1211, 6558)

Table A.12: Detailed results for the fixed parameter simulation in which outbreaks and datasets were simulated under parameter regime 4, characterized by fast outbreak dynamics (R0 = 3.25) and high mean case detection probability ($\rho = 0.75$). Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over–dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Parameter Regime	ODE	LNA	MMTL
Regime 1	0.3 (0.21, 0.34)	19.93 (14.96, 28.21)	62.85 (54.16, 83.94)
Regime 2	0.28 (0.17, 0.36)	14.41 (11.33, 20.43)	48.89 (35.47, 67.2)
Regime 3	0.31 (0.2, 0.4)	19.47 (15.57, 28.28)	62.09 (55.75, 83.89)
Regime 4	0.28 (0.17, 0.36)	14.66 (11.43, 20.58)	49.04 (35.72, 67.53)

Table A.13: Median (2.5%, 97.5%) quantiles of run times, in minutes, for MCMC chains in fixed parameter coverage simulations. Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). True parameter values are presented in Table A.5.2. Parameter regimes 1 and 3 had slower outbreak dynamics (R0 $= 1.75$, vs. R0 $= 3.25$). Parameter regimes 3 and 4 had higher case detection rates ($\rho = 0.75$, vs $\rho = 0.25$).