Title

A linear noise approximation for stochastic epidemic models fit to partially observed incidence counts

Authors

Fintzi, Jonathan
Wakefield, Jon
Minin, Vladimir N

Publication Date

2020-01-14

A Linear Noise Approximation for Stochastic Epidemic Models Fit to Partially Observed Incidence Counts

Jonathan Fintzi[1], Jon Wakefield[2,*], Vladimir N. Minin[3,*]
[1]Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Rockville, Maryland, U.S.A.
[2]Departments of Biostatistics and Statistics, University of Washington, Seattle, Washington, U.S.A.
[3]Department of Statistics, University of California, Irvine, California, U.S.A.
[*]corresponding authors: jonno@uw.edu and vminin@uci.edu

Abstract

Stochastic epidemic models (SEMs) fit to incidence data are critical to elucidating outbreak transmission dynamics, shaping response strategies, and preparing for future epidemics. SEMs typically represent counts of individuals in discrete infection states using Markov jump processes (MJP), but are computationally challenging as imperfect surveillance, lack of subject–level information, and temporal coarseness of the data obscure the true epidemic. Analytic integration over the latent epidemic process is generally impossible, and integration via Markov chain Monte Carlo (MCMC) is cumbersome due to the dimensionality and discreteness of the latent state space. Simulation-based computational approaches can address the intractability of the MJP likelihood, but are numerically fragile and prohibitively expensive for complex models. A linear noise approximation (LNA) that replaces the MJP transition density with a Gaussian density has been explored for analyzing prevalence data in large–population settings. Existing LNA frameworks are inappropriate for incidence data and depend on simulation-based methods or an assumption that disease counts are normally distributed. We demonstrate how to reparameterize a SEM to properly analyze incidence data, and fold the LNA into a data augmentation MCMC framework that outperforms deterministic methods, statistically, and simulation-based methods, computationally. Our framework is computationally robust when the model dynamics are complex and can be applied to a broad class of SEMs. We apply our method to national-level surveillance counts from the 2013–2015 West Africa Ebola outbreak, modeling within-country transmission and importation of infections from neighboring countries.

Keywords: Bayesian data augmentation; Ebola outbreak; Elliptical slice sampler; Non-centered parameterization; Surveillance count data.

1 Introduction

Incidence count data reported by public health surveillance systems provide critical information used in preparing for, and responding to, infectious disease outbreaks. These data are used to inform our understanding of outbreak transmission dynamics, to quantify uncertainty about how an epidemic might evolve, and to assess the effectiveness of interventions in interrupting transmission. Incidence counts reflect the number of new cases accumulated in each inter-observation interval, and are distinct from prevalence counts, which record the number of infected individuals at each observation time. Imperfect surveillance and the occurrence of asymptomatic cases result in systematic under-reporting and make it difficult to disentangle whether the data arose from a severe outbreak, observed with low fidelity, or a mild outbreak where most cases were detected. The absence of subject-level data and the temporal coarseness of incidence counts further reduce the amount of available information.

Outbreak data is commonly modeled using mechanistic compartmental models that prescribe physical laws governing the outbreak transmission dynamics as individuals transition between discrete infection states. These models can be specified at varying levels of granularity vis–a–vis the underlying epidemic process, ranging from agent-based models that track the infection histories of individuals (Jewell et al., 2009), to population–level models where the epidemic path tracks the numbers of individuals in each infection state (Lekone and Finkenstädt, 2006, Dukic et al., 2012). The parameters that govern the epidemic dynamics are of particular interest because they are informative about the mechanistic aspects of disease transmission. For example, the basic and effective reproduction numbers, R_0 and R_{eff}, are measures of the expected number of secondary cases per index case. Respectively, these quantities inform us about the likelihood that an outbreak will take off and persist. Moreover, these quantities can provide important insights into how an intervention could be designed to short circuit an outbreak since they are functions of other parameters that govern susceptibility, infectiousness, and rates of contact.

Stochastic epidemic models (SEMs) explicitly model variability in the epidemic process via distributions for waiting times between elementary transition events, such as infections and recoveries. Most commonly, a SEM represents the epidemic as a simple density–dependent Markov jump process (MJP) that evolves on a discrete state space of compartment counts with times of infection state occupancy taken to be exponentially distributed (Allen, 2017). The challenge in fitting a SEM to partially observed incidence is that we must sum over all epidemic paths from which the data could have arisen. This is difficult since the set of possible paths is enormous, even in small populations. Hence, the observed data likelihood of a SEM is intractable when incidence is under–reported.

Computation and inference for SEMs in the presence of under–reporting typically relies on either simulation–based methods or data augmentation (DA), often in

combination with an approximation of the target MJP (O'Neill, 2010). Simulation-based methods generate realizations of the epidemic process that form the basis for inference. This class of methods has been referred as "plug-and-play" since inference requires only the ability to simulate epidemic paths (Bretó et al., 2009). The particle marginal Metropolis-Hastings (PMMH) algorithm of (Andrieu et al., 2010) stands out as a state-of-the-art method for Bayesian inference with a fast and robust implementation for fitting epidemic models as part of the `pomp` `R` package (King et al., 2016a). Despite their flexibility, simulation-based methods can be prohibitively expensive for models with complex dynamics, and may fail entirely in the absence of an adequate model from which to simulate epidemic paths (Dukic et al., 2012).

DA facilitates Bayesian inference by targeting the joint posterior distribution of the latent epidemic process and SEM parameters. Historically, DA algorithms for fitting SEMs have relied on data–agnostic trans-dimensional proposals to sample subject-level disease histories (Gibson and Renshaw, 1998, O'Neill and Roberts, 1999). These proposal distributions are inefficient in the absence of subject–level data since the fraction of missing information is large (Roberts and Stramer, 2001). Modern DA algorithms that sample SEM paths conditionally on the data, rather than simulating unconditioned paths from the model, can be more computationally robust than simulation–based methods, especially in the absence subject–level data or when the SEM dynamics are complex (Pooley et al., 2015, Fintzi et al., 2017). However, the MJP representation of a SEM is itself a barrier to the application of DA machinery in large population settings since repeatedly evaluating the MJP likelihood, which is a product of exponential waiting time densities, at each iteration of a Markov chain Monte Carlo (MCMC) algorithm is prohibitively expensive. Hence, it is often necessary to approximate the MJP with a process whose likelihood is more tractable.

Commonly used MJP approximations broadly divide into two categories, those that approximate the epidemic discretely in time and preserve the discreteness of the MJP state space, and those where the time is treated continuously but where the state space of the approximating process is continuous. Discrete–time approximations include chain–binomial models (Lekone and Finkenstädt, 2006) and conditional Poisson/negative binomial count models (Bjørnstad et al., 2002, Held et al., 2005), both of which are attractive for their relatively low computational cost and have been used to fit models with complex spatio–temporal dynamics (Held et al., 2005, Wakefield et al., 2019). However, care must be taken with the choice of time step, particularly when the periodicity of data acquisition is misaligned with the average time between infection of a primary and secondary case, and when transition rates vary over the course of an outbreak (Glass et al., 2003). Moreover, the approximations underlying discrete time models can limit their applicability in many settings. For example, the commonly used TSIR framework developed in Bjørnstad

et al. (2002) depends on having a reasonable estimate of the number of susceptible individuals in each observation interval.

Among continuous–time approximations, the most common are ordinary and stochastic differential equations (ODEs, SDEs) (Allen, 2017). Deterministic ODE models, that can be viewed as infinite–population functional limits of MJPs, are attractive for their computational tractability, which is particularly important in outbreak settings when time is of the essence. These models also lend themselves to analytic characterizations of the outbreak dynamics, e.g., relating the final outbreak size to the basic reproductive number, which is useful when parameterizing complex models and incorporating prior information, especially in complex settings where certain aspects of a model might be poorly determined. However, deterministic models are known to underestimate uncertainty about the epidemic process (King et al., 2015), and moreover, are unable to address questions that are inherently stochastic, e.g., distributional questions about the size or duration of an outbreak. SDEs can reasonably approximate the stochastic aspects of the MJP in large population settings. However, SDE transition densities are typically unavailable in closed form, hence statistical inference requires simulation-based methods or simplification of the SEM for computation.

In this work, we develop a computational framework for fitting SEMs to partially observed incidence based on a linear noise approximation (LNA) for transition densities of a MJP over inter-observation intervals. The LNA replaces the MJP transition density with a Gaussian density whose moments are obtained by numerically solving systems of ODEs, derived by a Taylor expansion of the diffusion approximation for a density-dependent MJP around its deterministic ODE limit. The LNA has been a fixture in the biochemical reaction network literature on approximations of density–dependent MJPs since at least the 1970s (Kurtz, 1970, 1971). Reviews of the LNA, along with related approximations that derive from various system size expansions of the MJP Kolmogorov forward equation, can be found in (Wallace et al., 2012) and (Schnoerr et al., 2017). The LNA has found applications in the analysis of gene regulatory networks (Komorowski et al., 2009, Finkenstädt et al., 2013) and in outbreak modeling (Ross et al., 2009, Ross, 2012, Fearnhead et al., 2014). However, to our knowledge, the LNA has not been used to analyze incidence data. Rather, its application to outbreak modeling has been limited to either the analysis of prevalence or cumulative incidence data under an assumption of normally distributed disease counts, or has relied on simulation-based methods to allow for non-Gaussian emission distributions, as in Golightly et al. (2015).

Our contributions in this work are threefold. First, we demonstrate how SEMs should be reparameterized, in general, to properly analyze incidence data. In doing so, we aim to correct an unfortunately common error in transmission modeling where incidence counts are wrongly conflated with prevalence data. Second, we demonstrate how the LNA can be folded into a computationally robust Bayesian

data augmentation framework that leverages cutting edge MCMC algorithms and can be used to fit a broad class of SEMs. Critically, we are able to leverage tools that absolve us of the *de facto* requirement that observed disease counts follow a Gaussian distribution for the sake of computational efficiency, the alternative being computationally intensive particle filter methods. Finally, we provide guidance on optimal parameterizations that massively improve MCMC sampling efficiency for LNA paths and model parameters. In particular, we introduce a non–centered parameterization (NCP) for LNA transition densities that enables us to sample LNA paths using the elliptical slice sampling, which is a simple, computationally robust, and efficient MCMC algorithm free of tuning parameters. This enables us to fit models with complex dynamics that would otherwise be impossible to fit without compromising the model, even with cutting edge computational tools such as particle MCMC.

2 Stochastic Epidemic Models for Incidence Data

For clarity, we present the LNA framework in the context of fitting a susceptible–infected–recovered (SIR) model to negative binomial distributed incidence counts. The SIR model describes the transmission dynamics of an outbreak in a closed, homogeneously mixing population of P exchangeable individuals who are either susceptible (S), infected/infectious, (I), or recovered (R). The infection states relate to transmission, not disease. Thus, individuals are recovered when they no longer have infectious contact with others, not when they clear the infection. Note that our framework can easily be generalized beyond the SIR model and used to fit more complex SEMs. We will later consider a more complex multi-country model for the spread of Ebola in West Africa with country-specific susceptible-exposed-infected-recovered (SEIR) dynamics. This model adds an exposed (E), infected, but not yet infectious, compartment for latent infection, and allows for cross-border transmission.

2.1 Measurement Process and Data

Incidence data, $\mathbf{Y} = \{Y_1, \ldots, Y_L\}$, are counts of new infections in time intervals, $\mathcal{I} = \{\mathcal{I}_1, \ldots, \mathcal{I}_L : \mathcal{I}_\ell = (t_{\ell-1}, t_\ell]\}$. The observed incidence may reflect a fraction of the true incidence due to imperfect surveillance or asymptomatic cases. Or, cases may be over–reported if non–specific symptoms lead to over–diagnosis. 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 transitions over $\mathcal{I}_\ell$; so, $\Delta N^c_{SI}(t_\ell)$ is the incidence over $(t_{\ell-1}, t_\ell]$. Heterogeneities in case detection rates could lead to over–dispersion in the reporting distribution. Hence, we model the number of observed

cases as a negative binomial sample of the true incidence with mean case detection rate ρ and over–dispersion parameter ϕ:

$$Y_\ell | \Delta N^c_{SI}(t_\ell), \rho, \phi \sim \text{Neg.Binom.}(\mu_\ell = \rho \Delta N^c_{SI}(t_\ell),\ \sigma^2_\ell = \mu_\ell + \mu^2_\ell/\phi). \tag{1}$$

2.2 Latent Epidemic Process

The SIR model describes the evolution of compartment counts, $\mathbf{X}^c = \{S^c, I^c, R^c\}$, in continuous time on the state space

$$\mathcal{S}^c_X = \{(l, m, n) : l, m, n \in \{0, \ldots, P\},\ l + m + n = P\},$$

where P is the population size. We take the waiting times between state transitions to be exponentially distributed, which implies that $\mathbf{X}^c$ evolves according a MJP. Now, if our data had consisted of prevalence counts, we could approximate the MJP transition densities of $\mathbf{X}^c$ as in Ross et al. (2009) or Fearnhead et al. (2014). However, incidence data reflect the new infections in each inter–observation interval, and so the emission density (1) is expressed in terms of the change in N^c_{SI}, not I, over $(t_{\ell-1}, t_\ell]$. Hence, it is here that we first diverge from the approaches previously taken in the literature. It would be simply incorrect to treat incidence as though it were a change in prevalence. For instance, there could be a positive number of infections but prevalence might not change due to an equal number of recoveries over the inter–observation interval. To correctly specify an emission distribution for incidence data, we must construct the LNA that approximates the transition density of $\mathbf{N}^c$.

2.3 Cumulative Incidence of Infections and Recoveries

The cumulative incidence of infections and recoveries, $\mathbf{N}^c$, is a MJP with state space

$$\mathcal{S}^c_N = \{(j, k) : j, k \in \{0, \ldots, P\}, \mathbf{X}(\mathbf{N}^c(j, k)) \in \mathcal{S}_X\},$$

which is the set of non–decreasing cumulative incidence and recovery counts that do not lead to invalid prevalence paths (e.g., if there more recoveries than infections). Let β denote the per–contact infection rate, and μ the recovery rate. Then the rate at which $\mathbf{N}^c$ transitions from state $\mathbf{n}$ to $\mathbf{n}'$ is

$$\boldsymbol{\lambda}_{\mathbf{n},\mathbf{n}'}(\mathbf{X}^c) = \begin{cases} \lambda_{SI} = \beta SI, & \mathbf{n} = (n_{SI}, n_{IR}),\ \mathbf{n}' = (n_{SI} + 1, n_{IR}), \\ \lambda_{IR} = \mu I, & \mathbf{n} = (n_{SI}, n_{IR}),\ \mathbf{n}' = (n_{SI}, n_{IR} + 1), \\ 0, & \text{for all other } \mathbf{n} \text{ and } \mathbf{n}'. \end{cases} \tag{2}$$

We seek to infer the posterior distribution of $\mathbf{N}^c$ and the parameters, $\boldsymbol{\theta}$, that govern its dynamics along with those of the emission distribution. By the Markov

property and standard hidden Markov model conditional independence assumptions, the complete data likelihood is a product of emission densities and transition densities,

$$\begin{aligned}\pi(\boldsymbol{\theta}, \mathbf{N}^c \mid \mathbf{Y}) \;&\propto\; L(\mathbf{Y} \mid \mathbf{N}^c, \boldsymbol{\theta})\pi(\mathbf{N}^c \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &= \prod_{\ell=1}^{L} \left[\Pr\left(Y_\ell \mid \Delta\mathbf{N}^c_{SI}(t_\ell), \boldsymbol{\theta}\right) \times \pi\left(\mathbf{N}^c(t_\ell) \mid \mathbf{N}^c(t_{\ell-1}), \boldsymbol{\theta}\right)\right] \pi(\boldsymbol{\theta}), \qquad (3)\end{aligned}$$

where $\pi(\boldsymbol{\theta})$ denotes the prior distribution of the model parameters. The challenge in sampling from this posterior is that the transition densities of $\mathbf{N}^c$ are intractable due to the dimensionality of its state space. Moreover, we cannot analytically integrate over the latent epidemic process, except in trivial cases that are of little practical interest. In the following subsections, we use the LNA to approximate transition densities of $\mathbf{N}^c$, turning (3) into a more tractable product of Gaussian transition densities and emission densities. This will facilitate the use of efficient algorithms for sampling from the approximate posterior.

2.4 Diffusion Approximation

We approximate the integer–valued MJPs, $\mathbf{X}^c$ and $\mathbf{N}^c$, with the real–valued diffusion processes, $\mathbf{X}$ and $\mathbf{N}$, that satisfy an SDE, referred to as the chemical Langevin equation (CLE), whose drift and diffusion terms are chosen to match the approximate moments of MJP path increments in infinitesimal time intervals (Wilkinson, 2011, Golightly and Gillespie, 2013). Additional details regarding the diffusion approximation are given in Web Appendix A, and we refer to Gillespie (2000) and Fuchs (2013) for comprehensive discussions.

The state space of $\mathbf{X}$ for the SIR model is

$$\mathcal{S}_X^R = \{(l, m, n) : l, m, n \in [0, P],\; l + m + n = P\},$$

and the state space of $\mathbf{N}$ is

$$\mathcal{S}_N^R = \{(j, k) : j, k \in [0, P],\; \mathbf{X}(\mathcal{V}_{jk}) \in \mathcal{S}_X^R\}.$$

In words, the state space of $\mathbf{X}$ is the real–valued set of compartment volumes that are non–negative and sum to the population size, while the state space of $\mathbf{N}$ is the real–valued set of non–decreasing and non–negative incidence paths, constrained so that they do not lead to invalid prevalence paths (e.g., where there are more recoveries than infections and hence negative number of infected individuals). Let $\boldsymbol{\lambda}(\mathbf{X}(t))$ be the vector of transition rates at time t, e.g., $\boldsymbol{\lambda}(\mathbf{X}(t)) = (\beta S(t)I(t), \mu I(t))$, and $\boldsymbol{\Lambda}(\mathbf{X}(t)) = \text{diag}(\boldsymbol{\lambda}(t))$ For now, we ignore the constraints on $\mathcal{S}_N^R$ and $\mathcal{S}_X^R$, and approximate changes in cumulative incidence of infections and recoveries in an infinitesimal

time step with the CLE,

$$\mathrm{d}\mathbf{N}(t) = \boldsymbol{\lambda}(\mathbf{X}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{X}(t))^{1/2}\mathrm{d}\mathbf{W}_t, \tag{4}$$

where the vector $\mathbf{W}_t$ is distributed a bivariate Brownian motion with independent components.

Following Bretó and Ionides (2011) and Ho et al. (2018), we reparameterize $\mathbf{X}(t)$ in terms of $\mathbf{N}(t)$, conditional on the initial conditions $\mathbf{X}(t_0) = \mathbf{x}_0$ and $\mathbf{N}(t_0) = \mathbf{0}$. Let $\mathbf{A}$ denote the matrix whose rows specify changes in the compartment counts per infection or recovery:

$$\mathbf{A} = \begin{array}{c} \\ S\to I \\ I\to R \end{array} \begin{array}{c} \begin{array}{ccc} S & I & R \end{array} \\ \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \end{array}. \tag{5}$$

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

$$\mathbf{X}(t) = \mathbf{x}_0 + \mathbf{A}^T\mathbf{N}(t). \tag{6}$$

For the SIR model,

$$\begin{pmatrix} S(t) \\ I(t) \\ R(t) \end{pmatrix} = \begin{pmatrix} S_0 - N_{SI}(t) \\ I_0 + N_{SI}(t) - N_{IR}(t) \\ R_0 + N_{IR}(t) \end{pmatrix}, \tag{7}$$

so we rewrite (4) as

$$\mathrm{d}\mathbf{N}(t) = \boldsymbol{\lambda}(\mathbf{x}_0 + \mathbf{A}^T\mathbf{N}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{x}_0 + \mathbf{A}^T\mathbf{N}(t))^{1/2}\mathrm{d}\mathbf{W}_t. \tag{8}$$

Changes in compartment volumes have multiplicative effects on transition rates, and hence on increments in incident infections and recoveries. Therefore, we would like perturbations about the drift in (8) to be symmetric on a multiplicative, not an additive scale. This leads us to log transform (8). Let $\widetilde{\mathbf{N}} = \log(\mathbf{N} + \mathbf{1})$, so $\mathbf{N} = \exp(\widetilde{\mathbf{N}}) - \mathbf{1}$. By Itô's lemma (Øksendal, 2003), the SDE for $\widetilde{\mathbf{N}}$ is

$$\mathrm{d}\widetilde{\mathbf{N}}(t) = \underbrace{\mathrm{diag}\left(\exp(-\widetilde{\mathbf{N}}(t)) - 0.5\exp(-2\widetilde{\mathbf{N}}(t))\right) \times \boldsymbol{\lambda}\left(\mathbf{x}_0 + \mathbf{A}^T(\exp(\widetilde{\mathbf{N}}(t)) - \mathbf{1})\right)}_{\boldsymbol{\eta}(\mathbf{x}_0, \widetilde{\mathbf{N}}(t))}\mathrm{d}t +$$
$$\underbrace{\mathrm{diag}\left(\exp(-\widetilde{\mathbf{N}}(t))\right)\boldsymbol{\Lambda}\left(\mathbf{x}_0 + \mathbf{A}^T(\exp(\widetilde{\mathbf{N}}(t)) - \mathbf{1})\right)^{1/2}}_{\boldsymbol{\Phi}(\mathbf{x}_0, \widetilde{\mathbf{N}}(t))^{1/2}}\mathrm{d}\mathbf{W}_t. \tag{9}$$

2.5 Linear Noise Approximation

The LNA for SDEs having the form (9) is derived by decomposing $\widetilde{\mathbf{N}}$ into its deterministic ODE limit and a stochastic residual (Wilkinson, 2011). The SDE (9) is Taylor expanded around its deterministic limit, discarding higher order terms, to obtain a linear SDE for the stochastic residual term. This linear SDE has an explicit solution as a Gaussian random variable. The LNA reasonably approximates the stochastic aspects of a density dependent MJP over short time horizons when conditions (18) and (19) are satisfied (Wallace et al., 2012). Over longer time periods the approximation may deteriorate as stochastic perturbations to the system accumulate. The approximation can be substantially improved by phase–correcting the LNA via a time–transformation (Minas and Rand, 2017), or by using the LNA to approximate the one–step conditional transition densities of the MJP and restarting the approximation at the beginning of each inter–observation interval (Fearnhead et al., 2014). We adopt the latter strategy for its simplicity and ease of interpretation, although both strategies for improving the LNA over long time periods could be applied.

The LNA for (9) approximates the transition density of $\widetilde{\mathbf{N}}$ over the interval $(t_{\ell-1}, t_\ell]$ is

$$\widetilde{\mathbf{N}}(t_\ell) \mid \widetilde{\mathbf{n}}(t_{\ell-1}), \mathbf{x}(t_{\ell-1}), \boldsymbol{\theta} \sim MVN\left(\boldsymbol{\mu}(t_\ell) + \mathbf{m}(\widetilde{\mathbf{n}}(t_{\ell-1}) - \boldsymbol{\mu}(t_{\ell-1})), \boldsymbol{\Sigma}(t_\ell)\right), \tag{10}$$

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

$$\frac{\mathrm{d}\boldsymbol{\mu}(t)}{\mathrm{d}t} = \boldsymbol{\eta}(\boldsymbol{\mu}(t)), \tag{11}$$

$$\frac{\mathrm{d}\mathbf{m}(t)}{\mathrm{d}t} = \mathbf{F}(t)\mathbf{m}(t), \tag{12}$$

$$\frac{\mathrm{d}\boldsymbol{\Sigma}(t)}{\mathrm{d}t} = \mathbf{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\mathbf{F}(t)^T + \boldsymbol{\Phi}(t)^{1/2}\left(\boldsymbol{\Phi}(t)^{1/2}\right)^T, \tag{13}$$

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 $\boldsymbol{\eta}(t)$ and $\boldsymbol{\Phi}(t)$ are given in (9) and $\mathbf{F}(t) = \left(\frac{\partial \boldsymbol{\eta}_i(\boldsymbol{\mu}(t))}{\partial \boldsymbol{\mu}_j(t)}\right)_{i,j \in 1,\ldots,|\widetilde{\mathbf{N}}|}$ is the Jacobian of $\boldsymbol{\eta}(t)$ evaluated along the solution to (11). Note that we never solve (12) since $\mathbf{m}(t_0) = \mathbf{0}$ implies that $\mathbf{m}(t_\ell) = \mathbf{0}\ \forall\ l = 1, \ldots, L$.

The LNA posterior factorizes as a Gaussian state space model,

$$\begin{aligned}
&\pi(\widetilde{\mathbf{N}}, \boldsymbol{\theta} \mid \mathbf{Y}) \propto L(\mathbf{Y} \mid \widetilde{\mathbf{N}}, \boldsymbol{\theta}) \pi(\widetilde{\mathbf{N}} \mid \boldsymbol{\theta}) \mathbb{1}_{\left\{\mathbf{N} \in \mathcal{S}_N^R\right\}} \mathbb{1}_{\left\{\mathbf{X} \in \mathcal{S}_X^R\right\}} \pi(\boldsymbol{\theta}) \\
&= \prod_{\ell=1}^{L} \Pr(Y_\ell \mid \Delta\mathbf{N}(t_\ell), \boldsymbol{\theta}) \pi(\widetilde{\mathbf{N}}(t_\ell) \mid \widetilde{\mathbf{n}}(t_{\ell-1}), \mathbf{x}(t_{\ell-1}), \boldsymbol{\theta}) \times \mathbb{1}_{\left\{\mathbf{N}(t_\ell) \in \mathcal{S}_N^R\right\}} \mathbb{1}_{\left\{\mathbf{X}(t_\ell) \in \mathcal{S}_X^R\right\}} \pi(\boldsymbol{\theta}).
\end{aligned} \tag{14}$$

The emission densities in (14) depend on the incidence, not the log–incidence. Hence, we exponentiate LNA sample paths when computing emission densities. 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: first, to more faithfully approximate the MJP, and second, to avoid numerical instabilities that arise when $\mathbf{N}$ or $\mathbf{X}$ are negative.

2.6 Inference via the Linear Noise Approximation

We will sample LNA paths using the elliptical slice sampling (ElliptSS) algorithm of Murray et al. (2010), which is an efficient and computationally robust MCMC algorithm free of tuning parameters. ElliptSS can be used to sample a latent variable of interest, $\mathbf{Z}$, when the posterior decomposes into a Gaussian prior for $\mathbf{Z}$ and an arbitrary likelihood, $L(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta})$, i.e.,

$$\pi(\boldsymbol{\theta}, \mathbf{Z}|\mathbf{Y}) \propto L(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}) MVN(\mathbf{Z}; \mu_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}}). \tag{15}$$

It is critical to note that ElliptSS cannot be used to sample LNA paths when restarting the LNA ODEs as a strategy to improve the approximation. The mean of $\mathbf{N}(t_\ell)$ depends non–linearly on the value of $\mathbf{N}(t_{\ell-1})$, hence paths of $\widetilde{\mathbf{N}}$ are not jointly Gaussian. In order to facilitate the use of ElliptSS, we introduce a non–centered parameterization (NCP) that maps standard normal random variables, $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$, onto LNA paths (see Algorithm 1 for pseudo–code). Let $\mathbf{Z}(t_\ell) \sim MVN(\mathbf{0}, \mathbf{I})$, and $\widetilde{\mathbf{N}}(t_\ell) \sim MVN\left(\boldsymbol{\mu}(t_\ell), \boldsymbol{\Sigma}(t_\ell)\right)$, where $\boldsymbol{\mu}(t_\ell)$ and $\boldsymbol{\Sigma}(t_\ell)$ solve the LNA ODEs over $(t_{\ell-1}, t_\ell]$. The NCP is $(\boldsymbol{\theta}, \mathbf{Z})$, and maps $\mathbf{Z}$ via

$$\widetilde{\mathbf{N}}(t_\ell) \stackrel{\mathcal{L}}{=} \widetilde{\mathbf{W}}(t_\ell), \;\; \widetilde{\mathbf{W}}(t_\ell) = \boldsymbol{\mu}(t_\ell) + \boldsymbol{\Sigma}(t_\ell)^{1/2} \mathbf{Z}(t_\ell). \tag{16}$$

Our MCMC targets the joint posterior of the parameters and non–centered LNA draws,

$$\pi(\boldsymbol{\theta}, \mathbf{Z} \mid \mathbf{Y}) \propto L(\mathbf{Y} \mid \mathbf{N}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I})) \pi(\mathbf{Z}) \mathbb{1}_{\left\{\mathbf{N}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I}) \in \mathcal{S}_N^R\right\}} \mathbb{1}_{\left\{\mathbf{X}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I}) \in \mathcal{S}_X^R\right\}} \pi(\boldsymbol{\theta}), \tag{17}$$

where $\mathbf{N}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I})$ and $\mathbf{X}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I})$ denote sample paths obtained by centering the LNA draws.

The NCP also helps to alleviate issues of poor MCMC mixing that arise when fitting hierarchical latent variable models via DA MCMC. The problem arises in weak data settings, such as ours (see Web Appendix B and Figure S2), where MCMC samples can become severely autocorrelated when alternately sampling latent variables and model parameters (Papaspiliopoulos et al., 2003, Papaspiliopoulos et al., 2007). In our setting, this can be traced to the use of a centered parameterization (CP) for the LNA in (14). Under the CP, updates to $\boldsymbol{\theta} \mid \widetilde{\mathbf{N}}, \mathbf{Y}$ are made conditionally on a *fixed* LNA path and are accepted if they are concordant with the data

and the current path. This limits the magnitude of perturbations that can be made to the model parameters at each MCMC iteration and results in severe autocorrelation. In contrast, the NCP locates a sample LNA path within the transition densities induced by a set of proposed model parameters, thus allowing us to make more meaningful perturbations to model parameters at each MCMC iteration.

2.7 Parameter updates

In each MCMC iteration, we alternately update $\mathbf{Z} \mid \boldsymbol{\theta}, \mathbf{Y}$ via EliptSS, and $\boldsymbol{\theta} \mid \mathbf{Z}, \mathbf{Y}$ using either a global adaptive random walk Metropolis (GA-RWM) sampler when fitting SEMs with few parameters and simple dynamics, (Algorithm 4 in Andrieu and Thoms (2008)) or an adaptive multivariate normal slice sampler (MVNSS) when the dynamics are more complex (Web Appendix B. The GA-RWM algorithm is faster per–iteration, though we have found the MVNSS algorithm to be somewhat more robust in more complex settings. Identifying an estimation scale for the model hyperparameters is critical to MCMC efficiency and is discussed in detail in Web Appendix B. When the initial state, $\mathbf{X}_0$, is estimated, we assign an informative prior $\mathbf{X}_0 \sim TMVN_{\mathcal{S}_X^R}\left(P\mathbf{p}, P(\text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^T)\right)$, 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 $\mathbf{X}_0$ via ElliptSS (Web Appendix B).

3 Motivating the LNA Using Simulated Data

3.1 Comparison with Common SEM Approximations via Simulation

We benchmarked the LNA against two common approximations of the MJP: its deterministic infinite population ODE limit, and a discrete–time approximation where epidemic paths were simulated within a particle marginal Metropolis–Hastings (PMMH) framework (Andrieu et al., 2010) using a multinomial modification of the τ–leaping algorithm (MMTL) (Bretó and Ionides, 2011). The ODE approximation was chosen for its ubiquity in epidemic modeling, while the MMTL/PMMH approximation was chosen because of a straightforward and general implementation in the popular `pomp` `R` package (King et al., 2016b). Arguably, the MMTL more faithfully approximates the MJP vis–a–vis the LNA since it preserves the discreteness of the state space, while the deterministic ODE is further removed from the MJP.

To better understand how the LNA, ODE, and MMTL/PMMH approximations would perform in finite population settings, we fit simple SIR models to data from 500 outbreaks with SIR dynamics in three populations of different sizes and numbers of initially infected individuals. Each outbreak was simulated from a MJP via

Gillespie's direct algorithm (Gillespie, 1976) conditional on parameters drawn from a set of prior distributions (Table S4). The priors were chosen to reflect outbreak dynamics and detection rates that would be typical of many contact driven outbreaks, while the population sizes were reflective of settings in which the methods might reasonably be applied, i.e., not so big that the outbreaks would essentially evolve deterministically, nor so small that the approximations would be unreasonable. We also assessed how the computational robustness of the methods would hold up when fitting a more complex multi–country SEIR model to a dataset simulated under the model presented Section 4. In all simulations, the models were fit under the same priors from which the parameters were drawn, ensuring that the posterior would be properly calibrated. 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 details and results for the SIR models are provided in Web Appendix C, and further details regarding the multi–country SEIR model are provided in Section 4 and Web Appendix E.

3.2 Results

LNA and MMTL point estimates and credible interval widths were similar for all parameters in each of the three population size regimes, and coverage of credible intervals were close to the nominal 95% levels. Despite the lack of structure model misspecification, the ODE models struggle to reliably recover the model parameters, particularly those governing the sampling process (Figure 1). Coverage of credible intervals for ODE models was low for all model parameters, and only somewhat improved as the population size increased. The distributions of LNA and ODE median absolute deviations relative to MMTL estimates indicate that the ODE estimates were less precise, while credible intervals also tended to be narrower than estimates obtained with the two stochastic approximations. This is in agreement with King et al. (2015), who found that ODE models tend to underestimate uncertainty in epidemic dynamics. Finally, the computational performance of the LNA and MMTL/PMMH approximations was comparable fitting simple SIR models, with the LNA perhaps being slightly faster. This, of course, comes with the usual caveat that comparisons of computational performance depend on the efficiency of each implementation and a variety of algorithmic choices (e.g., timestep of MMTL or number of particles in PMMH).

The parity in computational performance between the LNA and MMTL/PMMH algorithms vanished when we attempted to fit a more complex multi–country SEIR model to simulated data. Despite the absence of any structural model misspecification, the MMTL/PMMH algorithm failed to yield an MCMC run that adequately explored the posterior when given a similar computational budget to the LNA DA MCMC algorithm (Table 1). The computational cost of simulation–based meth-

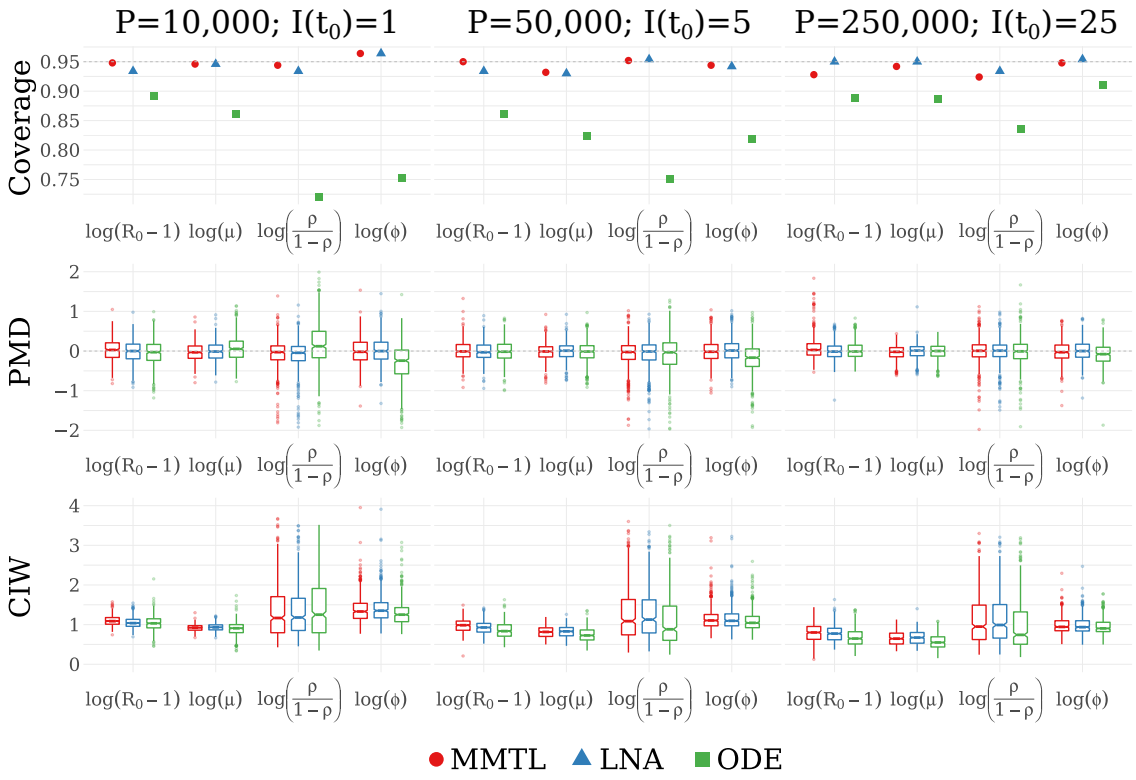

Figure 1: Results for SIR models fit to 500 datasets via multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings, the linear noise approximation (LNA), and deterministic ordinary differential equations (ODE), shown respectively in that order for each parameter. R_0 is the basic 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, posterior median deviations (PMD), and 95% Bayesian credible interval widths (CIW). The simulation was repeated for three regimes of population sizes ($P = 10,000$; $P = 50,000$; $P = 250,000$) and initially infected individuals ($I(t_0) = 1$; $I(t_0) = 5$; $I(t_0) = 25$).

ods, such as PMMH, increases with model complexity and length of the data since more particles are required to obtain an estimate of the likelihood and update the latent path. It might be possible to obtain a working MMTL/PMMH run with even greater expenditure of time and computational resources. However, our inability to obtain a valid posterior sample in an "easy" setting where we knew the true data generating mechanism led us to abandon PMMH as a reliable computational strategy for analyzing real–world data from the West Africa outbreak.

4 Modeling the Spread of Ebola in West Africa

We now turn our attention to modeling the 2013–2015 Ebola outbreak in the West African countries of Guinea, Liberia, and Sierra Leone. Our objective will be to describe the transmission dynamics of the outbreak, and in particular to estimate the basic reproductive numbers for each country. The data, shown in Figure 2, consist of national case counts from the World Health Organization patient database consisting of weekly confirmed and probable Ebola cases (Organization, 2016). Individuals were classified as suspected cases if they presented with Ebola–like symptoms and had contact with a suspected, probable, or confirmed case of a dead or sick animal. Probable cases were defined as suspected cases who were evaluated by a clinician, or who had died but were epidemiologically linked to confirmed cases. A confirmed case was defined as a suspected or probable case testing positive for Ebola virus RNA or IgM Ebola antibodies (Coltart et al., 2017). For illustrative purposes, we chose to simplify the model by lumping together confirmed and probable cases.

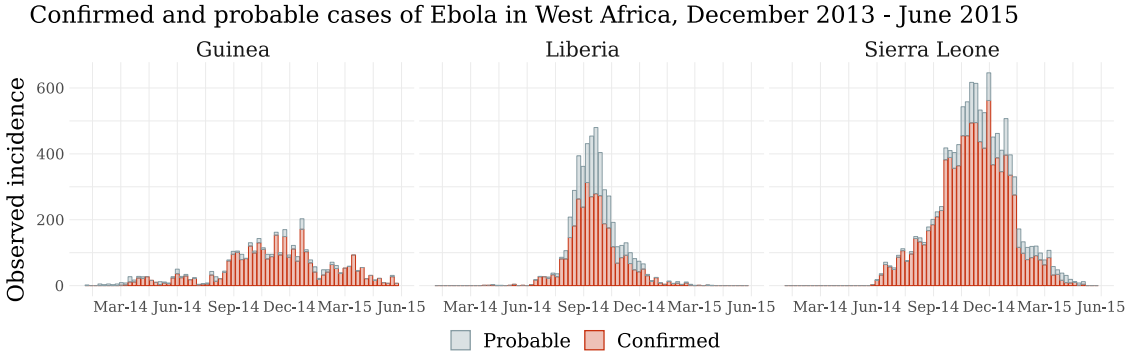

Figure 2: Weekly incidence of confirmed and probable cases of Ebola in Guinea, Liberia, and Sierra Leone. The total incidence was 3,627 in Guinea, 4,994 in Liberia, and 11,317 in Sierra Leone, and the estimated population sizes were 11.8 million, 4.4 million, and 7.1 million, respectively (United Nations, 2017).

We used the LNA and ODE approximations to fit a multi–country model for

Table 1: (Top) Computational statistics for models fit in the coverage simulation: We report 50% (2.5%, 97.5%) quantiles of total CPU time for five MCMC chains and total log–posterior effective sample size (ESS)/total CPU time for SIR models fit via the ODE, LNA, and MMTL/PMMH approximations. (Bottom) Computational statistics for models fit via the ODE, LNA, and MMTL approximations to a simulated dataset mimicking an Ebola outbreak using the model in Section 4. Computation for the MMTL model was carried out using PMMH via the `pomp` `R` package, while the LNA and ODE models were fit via ElliptSS/MVNSS (Web Appendices B and E). Non–convergence of the MCMC algorithm for the MMTL model was indicated by the fact potential scale reduction factors for four parameters exceeded 1.05, hence we do not report the log–posterior ESS for that model.

Simple SIR models

	ODE	**LNA**	**MMTL**
Total CPU time (minutes)	2.1 (1.3, 3.0)	139.4 (61.1, 280.9)	438.0 (205.4, 788.2)
Total log–posterior ESS/total CPU time	170000 (110000, 250000)	1600 (130, 9100)	980 (460, 2000)

Multi–country SEIR model

	ODE	**LNA**	**MMTL**
Total CPU time (hours)	6	900	914
Total log–posterior ESS/total CPU time	1172.7	2.05	—

the spread of Ebola in Guinea, Liberia, and Sierra Leone under country–specific SEIR transmission dynamics, illustrated in Figure 3. Cross–border transmission was incorporated via virtual migration of infectious individuals and was parameterized by extrinsic reproduction numbers, interpretable as the expected number of secondary cases in a country per index case in another country. Transmission was assumed to commence in Liberia on March 2^{nd}, 2014, and in Sierra Leone on May 4^{th}, 2014, corresponding to three weeks prior to the first cases in those countries. The observed incidence in each country was modeled as a negative binomial sample of the true incidence. The total incidence in each country was small relative to the population size, suggesting that only a fraction of the population was geographically or socially linked to ongoing transmission. Hence, we estimated the effective population size in each country, interpreted as the size of the sub–population within which the outbreak occurred. The implications of estimating the effective population size for identifiability are discussed in Web Appendix D. We assigned priors for effective population sizes by matching outbreak sizes based on deterministic final size relations for SEIR models to outbreak sizes computed by inflating the observed incidence by the case detection rates (Web Appendix E). The priors for other parameters were informed by published estimates of Ebola transmission dynamics (Table S13). We also fit a set of simplified models to data from each country independently to understand the benefit of explicitly modeling cross–border transmission (Web Appendix E). Results from the single–country models did not differ substantively from results obtained with the joint model (Table S19).

The estimated transmission dynamics under the LNA model (Table S19) were consistent with published estimates obtained with stochastic models fit to aggregate incidence data Chretien et al. (2015). The posterior median (95% BCI) basic reproduction numbers, adjusted by the estimated effective population sizes, were 1.2 (1.1, 1.5) in Guinea, 1.9 (1.4, 3.2) in Liberia, and 1.3 (1.2, 1.4) in Sierra Leone. Adjusted basic reproduction numbers estimated under the ODE tended to be higher than estimates obtained under the LNA (top panel, Figure 4). Posterior distributions of latent and infectious period durations in Guinea and Liberia largely recovered the priors, while the estimated durations for Sierra Leone were somewhat shorter, though not unreasonably so. Estimated latent and infectious period durations under the ODE were longer than under the LNA, especially in Guinea and Liberia, where they were much longer than expected. Though we allowed for cross–border transmission, we were not able to resolve its contribution from the data and exactly recovered the priors for extrinsic basic reproduction numbers between each pair of countries. Additional results, including diagnostics, are provided in Web Appendix E.

Compared with the deterministic ODE approximation, the LNA provides a better fit to the data and more appropriately accounts for uncertainty about the outbreak. One clear indication of this is that the LNA posterior predictive distribution,

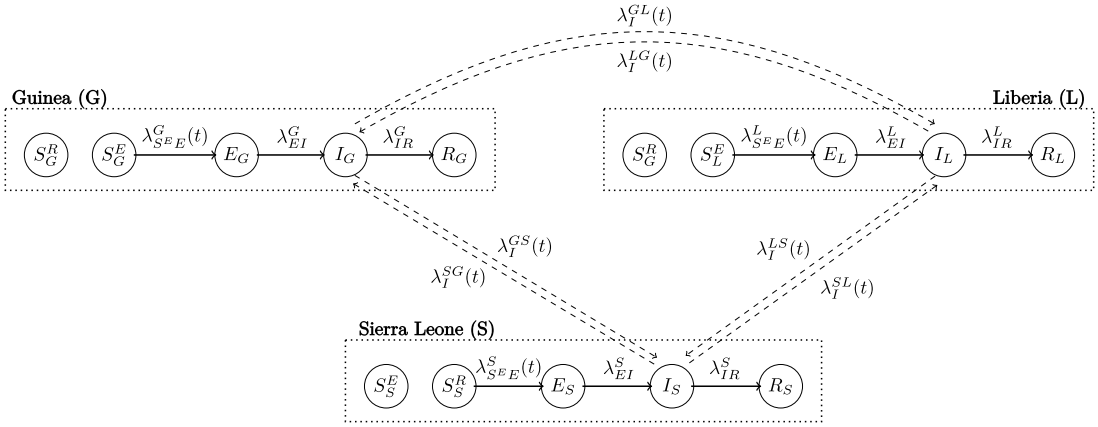

Figure 3: Diagram of state transitions for a joint model for Ebola transmission in Guinea, Liberia, and Sierra Leone. Dotted boxes denote countries, nodes in circles denote the model compartments: susceptible but removed from infectious contact (S^R), susceptible but exposed to infectious contact (S^E), exposed (E), infectious (I), recovered (R). Compartments are subscripted with country indicators. Solid lines with arrows indicate stochastic transitions between model compartments, which occur continuously in time. Dashed lines indicate that infected individuals in one country contribute to the force of infection in another country. Rates at which individuals transition between compartments are denoted by λ and are subscripted by compartments and superscripted by countries, e.g., $\lambda^L_{S^E E}$ is the rate at which susceptible individuals become exposed in Liberia. Transmission in Liberia and Sierra Leone was assumed to commence at 10 and 19 weeks, respectively. Full expressions for the rates are given in Table S12.

which integrates over the joint posterior of the model parameters and latent epidemic process, more closely resembles the observed incidence than does the ODE posterior predictive distribution (bottom panel, Figure 4). ODE posterior predictive intervals (PPIs) tend to be wider than their LNA counterparts, while ODE posterior predictive p-values (PPPs) are more extreme (Figure S11). The poor performance of the ODE is due to its being forced to account for all observations with a single deterministic path, which inevitably leaves many observations poorly explained. On the other hand, the stochastic nature of the LNA, and in particular, the restarting formulation explored in this work, allows the model to capture local variability in the epidemic process. This dynamic also results in inflated estimates of the negative binomial overdispersion parameter under the ODE model since misspecification of the latent epidemic process is absorbed as overdispersion in the emission distribution.

5 Discussion

We have presented a broadly applicable and computationally robust framework for fitting stochastic epidemic models to partially observed incidence data that arise in disease surveillance and outbreak settings. Though the linear noise approximation has been used in outbreak modeling, its application has been restricted to the analysis of cumulative incidence or prevalence data, the latter of which is often unfortunately conflated with incidence. Critically, we can estimate SEM parameters and appropriately account for the stochastic aspects of the MJP without compromising the approximation or resorting to fragile and computationally intensive simulation–based methods.

Our main contributions in this work were to demonstrate how to correctly reparameterize a SEM to admit latent epidemic paths that are compatible with emission distributions for incidence data, and how the LNA approximation can be folded into a computationally robust MCMC machinery while accommodating non–Gaussian emission distributions. We showed in simulations with simple SIR models, which are often used as building blocks in more complex models, that SEMs approximated via the LNA were comparable to those approximated using MMTL, and vastly outperformed ODE approximations that are still in common use. We used our framework to fit a high dimensional model, which was fully stochastic in all aspects of the transmission dynamics, to data from the 2013–2015 outbreak of Ebola in West Africa. We were unable to fit this model using the state–of–the–art PMMH algorithm, even to simulated data in the absence of model misspecification.

For clarity of exposition, we have restricted ourselves to relatively simple models, though we feel it is important to acknowledge that there a variety of ways in which scientific validity of these models could have been improved by introducing additional complexity. For example, it is frequently, if not universally, unreasonable to assume that the transmission and surveillance dynamics are constant over

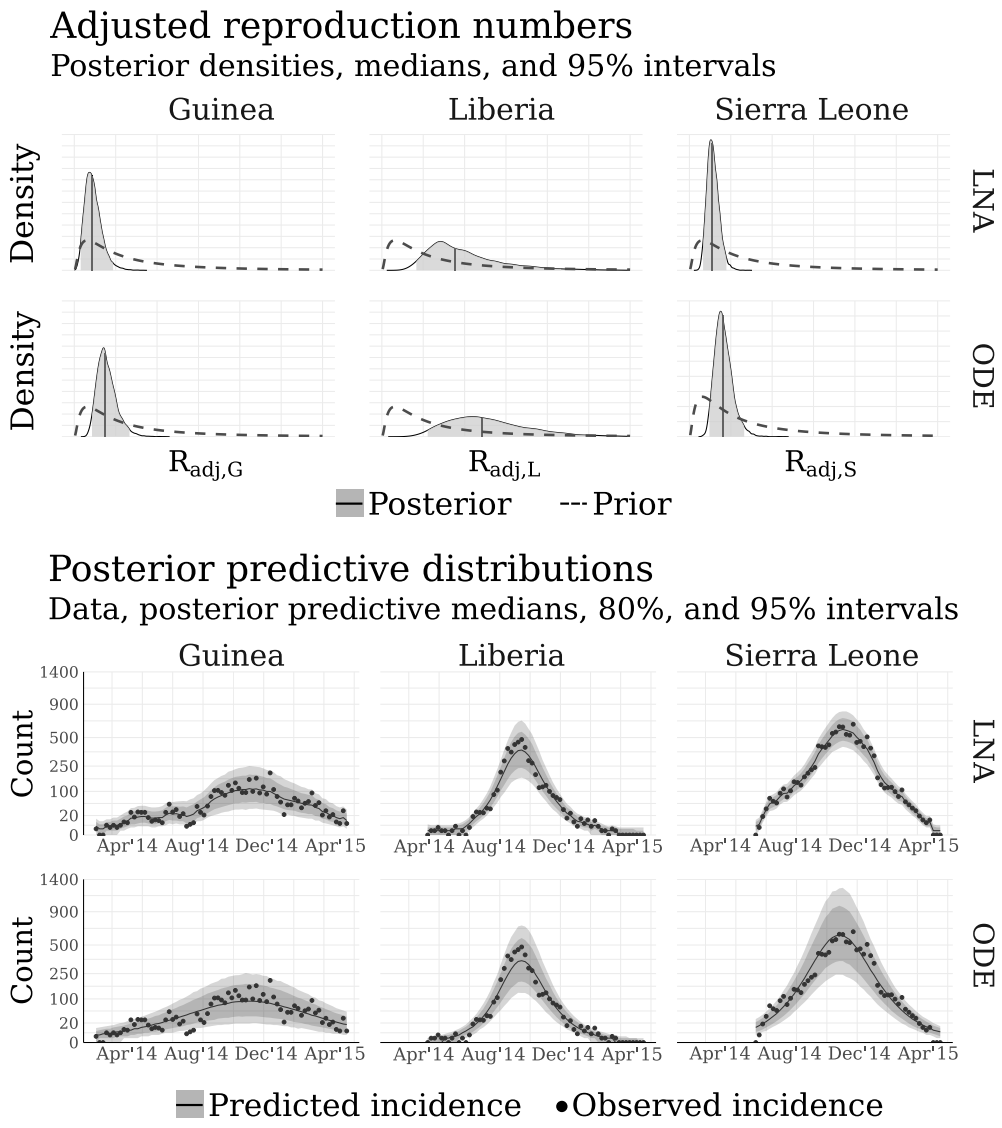

Figure 4: (Top panel) Posterior distributions (shaded densities) of effective population size adjusted basic reproduction numbers under the LNA and ODE models, with prior densities (dashed lines) plotted over the posterior ranges. Solid vertical lines are posterior medians, and shaded regions correspond to 95% credible intervals. The adjusted basic reproduction number is defined with respect to the effective population size. For example, $R_{adj,G} = P_{eff,G}\beta_G/\mu_G$, where $P_{eff,G}$, β_G, and μ_G are the effective population size, within–country per–contact infection rate, and recovery rate for Guinea, respectively. (Bottom panel) Posterior predicted incidence in West Africa under the LNA and ODE models, marginalizing over the joint posterior distributions of the latent incidence and model parameters. Solid lines are pointwise posterior predictive median incidence, shaded bands correspond to 80% and 95% pointwise posterior predictive incidence distributions, and dots are the observed incidence.

time. Fitting models with time–inhomogeneous dynamics presents additional computational difficulties and are an important target for future work. We have also not touched on formal assessments of the predictive performance, issues of model selection, and the effects of model misspecification, though we acknowledge that all of these topics are critical to the application of our methods.

Acknowledgements

J.F., J.W., and V.N.M. were supported by the NIH grant U54 GM111274. J.W. was supported by the NIH grant R01 CA095994. V.N.M. was supported by the NIH grant R01 AI107034. This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov)

Supplementary Materials

The algorithms for fitting LNA and ODE models are implemented in the `stemr R` package, which is available from the following stable GitHub repository along with code for reproducing the results presented in this paper: `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` (Eddelbuettel and François, 2011, Eddelbuettel and Sanderson, 2014). ODE integration functions are dynamically compiled in `C++` with the help of the `odeintr` R package (Keitt, 2017) and ODEs can be integrated using a variety of methods available in the `Odeint C++` library (Ahnert and Mulansky, 2011).

Web Appendices A, B, C, D, and E, referenced throughout this work, are available with this paper at the Biometrics website on Wiley Online Library.

References

Y. Ahmadian, J. W. Pillow, and L. Paninski. Efficient Markov chain Monte Carlo methods for decoding neural spike trains. Neural Computation, 23:46–96, 2011.

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

L. J. S. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. Infectious Disease Modelling, 2:128–142, 2017.

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

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.

O. N. Bjørnstad, B. F. Finkenstädt, and B. T. Grenfell. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecological Monographs, 72:169–184, 2002.

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.

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.

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

G. Chowell and H. Nishiura. Transmission dynamics and control of Ebola virus disease (EVD): a review. BMC medicine, 12:196, 2014.

J. P. Chretien, S. Riley, and D. B. George. Mathematical modeling of the West Africa Ebola epidemic. eLife, 4:e09186, 2015.

C. E. M. Coltart, B. Lindsey, I. Ghinai, A. M. Johnson, and D. L. Heymann. The Ebola outbreak, 2013–2016: old lessons for new epidemics. Philosophical Transactions of the Royal Society B, 372:20160297, 2017.

B. D. Dalziel, M. S. Y. Lau, A. Tiffany, A. McClelland, J. Zelner, J. R. Bliss, and B. T. Grenfell. Unreported cases in the 2014-2016 Ebola epidemic: Spatiotemporal

variation, and implications for estimating transmission. PLoS Neglected Tropical Diseases, 12:e0006161, 2018.

M. C. M. de Jong, O. Diekmann, and H. Heesterbeek. How does transmission of infection depend on population size? Publications of the Newton Institute, 5: 84–94, 1995.

G. Dudas, L.M. Carvalho, T. Bedford, A. J. Tatem, G. Baele, N.R. Faria, D. J. Park, J.T. Ladner, A. Arias, and D. Asogun. Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature, 544:309, 2017.

V. Dukic, H.F. Lopes, and N. G. Polson. Tracking epidemics with Google flu trends data and a state-space SEIR model. Journal of the American Statistical Association, 107:1410–1426, 2012.

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

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

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

B. Finkenstädt, D. J. Woodcock, M. Komorowski, C. V. Harper, J. R. E. Davis, M.R.H. White, and D. A. Rand. Quantifying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: An application to single cell data. The Annals of Applied Statistics, 7:1960–1982, 2013.

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

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.

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

G. J. Gibson and E. Renshaw. Estimating parameters in stochastic compartmental models using Markov chain methods. Mathematical Medicine and Biology, 15: 19–40, 1998.

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.

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

K. Glass, Y. Xia, and B. Grenfell. Interpreting time-series analyses for continuous-time biological models — measles as a case study. Journal of Theoretical Biology, 223:19–25, 2003.

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

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

L. Held, M. Höhle, and M. Hofmann. A statistical framework for the analysis of multivariate infectious disease surveillance counts. Statistical Modelling, 5:187–199, 2005.

L. S. T. Ho, F. W. Crawford, and M. A. Suchard. Direct likelihood-based inference for discretely observed stochastic compartmental models of infectious disease. The Annals of Applied Statistics, 12:1993–2021, 2018.

C. P. Jewell, T. Kypraios, P. Neal, and G. O. Roberts. Bayesian analysis for emerging infectious diseases. Bayesian Analysis, 4:465–496, 2009.

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

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.

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, 2016a.

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, 2016b.

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.

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

T. G. Kurtz. Limit theorems for sequences of jump Markov processes. Journal of Applied Probability, 8:344–356, 1971.

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

F. Liang, C. Liu, and R. Carroll. Advanced Markov chain Monte Carlo methods: learning from past samples, volume 714. John Wiley & Sons, Hoboken, 2011.

J. C. Miller. A note on the derivation of epidemic final sizes. Bulletin of Mathematical Biology, 74:2125–2141, 2012.

G. Minas and D. A. Rand. Long-time analytic approximation of large stochastic oscillators: Simulation, analysis and inference. PLoS Computational Biology, 13: e1005676, 2017.

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

United Nations. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2017 Revision. `https://esa.un.org/unpd/wpp/DataQuery`, 2017. Last accessed: February 28, 2018.

R. M. Neal. Slice sampling. Annals of Statistics, pages 705–741, 2003.

B. Øksendal. Stochastic Differential Equations. Springer, New York, 2003.

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

P. D. O'Neill and G. O. Roberts. Bayesian inference for partially observed stochastic epidemics. Journal of the Royal Statistical Society: Series A (Statistics in Society), 162:121–129, 1999.

World Health Organization. World Health Organization. Ebola data and statistics. `http://apps.who.int/gho/data/node.ebola-sitrep.quick-downloads?lang=en`, May 11, 2016. Last accessed: February 28, 2018.

O Papaspiliopoulos, GO Roberts, and M Sköld. Non-centered parameterisations for hierarchical models and data augmentation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and D. West, editors, Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting, volume 307, pages 307–326. Oxford University Press, USA, 2003.

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

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

C. M. Pooley, S.C. Bishop, and G. Marion. Using model-based proposals for fast parameter inference on discrete state space, continuous-time Markov processes. Journal of The Royal Society Interface, 12:20150225, 2015.

G. O. Roberts and O. Stramer. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. Biometrika, 88:603–621, 2001.

J. V. Ross. On parameter estimation in population models III: Time-inhomogeneous processes and observation error. Theoretical Population Biology, 82:1–17, 2012.

J. V. Ross, D. E. Pagendam, and P. K. Pollett. On parameter estimation in population models II: multi-dimensional processes and transient dynamics. Theoretical Population Biology, 75:123–132, 2009.

S. V. Scarpino, A. Iamarino, C. Wells, D. Yamin, M. Ndeffo-Mbah, N. S. Wenzel, S. J. Fox, T. Nyenswah, F.L. Altice, A. P. Galvani, et al. Epidemiological and viral genomic sequence analysis of the 2014 Ebola outbreak reveals clustered transmission. Clinical Infectious Diseases, 60:1079–1082, 2014.

D. Schnoerr, G. Sanguinetti, and R. Grima. Approximation and inference methods for stochastic biochemical kinetics — a tutorial review. Journal of Physics A: Mathematical and Theoretical, 50:093001, 2017.

M. B. Thompson. Slice Sampling with Multivariate Steps. PhD dissertation, 2011.

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.

G. E. Velásquez, O. Aibana, E. J. Ling, I. Diakite, E. Q. Mooring, and M. B. Murray. Time from infection to disease and infectiousness for Ebola virus disease, a systematic review. Clinical Infectious Diseases, 61:1135–1140, 2015.

J. Wakefield, T. Q. Dong, and V. N. Minin. Spatio-temporal analysis of surveillance data. In L. Held, N. Hens, P. O'Neill, and J Wallinga, editors, Handbook of Infectious Disease Data Analysis, pages 455–476. CRC Press, Boca Raton, 2019.

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.

D. J. Wilkinson. Stochastic Modelling for Systems Biology. CRC Press, Boca Raton, 2011.

Web Appendix A: Diffusion Approximations for Markov Jump Processes

There are a variety of methods for arriving at the diffusion approximation for a MJP (Fuchs, 2013). We outline an intuitive, though somewhat informal, construction of a stochastic differential equation (SDE), referred to as the chemical Langevin equation (CLE), where the drift and diffusion terms are chosen to match the approximate moments of MJP path increments in infinitesimal time intervals (Wilkinson, 2011, Golightly and Gillespie, 2013), and refer to Gillespie (2000) and Fuchs (2013) for more detailed presentations.

Denote the compartment counts at time t by $\mathbf{X}^c(t) = \mathbf{x}_t^c$. We want to approximate the numbers of infections and recoveries in a small time interval, $(t, t + \mathrm{d}t]$, i.e., $\mathbf{N}^c(t + \mathrm{d}t) - \mathbf{N}^c(t)$. Suppose that we can choose $\mathrm{d}t$ so 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:

$$\boldsymbol{\lambda}(\mathbf{X}^c(t')) \approx \boldsymbol{\lambda}(\mathbf{x}^c(t)),\ \forall t' \in (t, t + \mathrm{d}t]. \tag{18}$$

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

$$\boldsymbol{\lambda}(\mathbf{x}^c(t)) \gg \mathbf{1}. \tag{19}$$

Condition (18), holds when $\mathrm{d}t$ is infinitesimally small, and implies that the numbers of infections and recoveries in $(t, t + \mathrm{d}t]$ are essentially independent since their rates of occurrence are approximately constant within the interval (Gillespie, 2000). This condition also implies that the numbers of infections and recoveries are 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 $N_{IR}^c(t + \mathrm{d}t) \sim \text{Poisson}(\mu I(t) \mathrm{d}t)$ (Wilkinson, 2011). Condition (19) is likely to be satisfied in large populations (Wallace et al., 2012) and implies that the Poisson innovations are well–approximated by Gaussian random variables.

When (18) and (19) hold, we can approximate the integer–valued processes, $\mathbf{X}^c$ and $\mathbf{N}^c$, with the real–valued processes, $\mathbf{X}$ and $\mathbf{N}$. The state space of $\mathbf{X}$ for the SIR model is

$$\mathcal{S}_X^R = \{(l, m, n) : l, m, n \in [0, P],\ l + m + n = P\},$$

and the state space of $\mathbf{N}$ is

$$\mathcal{S}_N^R = \{(j, k) : j, k \in [0, P],\ \mathbf{X}(\mathcal{V}_{jk}) \in \mathcal{S}_X^R\}.$$

In words, the state space of $\mathbf{X}$ is 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., where there are more recoveries than infections and hence negative number of infected individuals). For now, we ignore the constraints on $\mathcal{S}_N^R$ and $\mathcal{S}_X^R$, and approximate changes in cumulative incidence of infections and recoveries in an infinitesimal time step as

$$\mathbf{N}(t + \mathrm{d}t) - \mathbf{N}(t) \approx \boldsymbol{\lambda}(\mathbf{X}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{X}(t))^{1/2}\mathrm{d}t^{1/2}\mathbf{Z}, \tag{20}$$

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

$$\mathrm{d}\mathbf{N}(t) = \boldsymbol{\lambda}(\mathbf{X}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{X}(t))^{1/2}\mathrm{d}\mathbf{W}_t, \tag{21}$$

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

Web Appendix B: Algorithms and Additional MCMC Details

Algorithms for Sampling LNA Paths

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 (11) and (13) 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
    $\mathbf{N}(t_\ell) \leftarrow \mathbf{N}(t_{\ell-1}) + \exp(\widetilde{\mathbf{N}}(t_\ell)) - \mathbf{1}$
    $\mathbf{X}(t_\ell) \leftarrow \mathbf{X}(t_{\ell-1}) + \mathbf{A}^T(\mathbf{N}(t_\ell) - \mathbf{N}(t_{\ell-1}))$
    $\widetilde{\mathbf{N}}(t_\ell) \leftarrow \mathbf{0},\ \boldsymbol{\mu}(t_\ell) \leftarrow \mathbf{0},\ \boldsymbol{\Sigma}(t_\ell) \leftarrow \mathbf{0}$ ▷ restart initial conditions
return ▷ incidence and/or prevalence sample paths
  $\mathbf{N} = \{\mathbf{N}(t_0), \mathbf{N}(t_1), \ldots, \mathbf{N}(t_L)\}$
  $\mathbf{X} = \{\mathbf{X}(t_0),\ \mathbf{X}(t_1), \ldots, \mathbf{X}(t_\ell)\}$

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:

$$\begin{aligned} \psi &\sim \text{Unif}(0, \omega) \\ L_\psi \leftarrow -\psi;\ R_\psi &\leftarrow L_\psi + \psi \\ \phi &\sim \text{Unif}(L_\psi, R_\psi) \end{aligned}$$

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

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 ElliptSS bracket width at $\omega < 2\pi$. This is motivated by the observation that when the model dynamics are complex, the ElliptSS 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 ElliptSS algorithm we must solve the LNA ODEs in order to compute the observed data likelihood. Thus, if we can reduce the number of ElliptSS steps, we will be able to shorten the run time of our MCMC.

We 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 ElliptSS 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. In order to facilitate tuning of the initial bracket width, the elliptical slice sampling Algorithm 2 was modified slightly from the one in Murray et al. (2010) with respect to the initial angle proposal so that the distribution of angles for accepted proposals would be symmetric around zero.

Figure S1 presents histograms of the number of contractions per ElliptSS update and the accepted angles before and after contracting the initial ElliptSS bracket width for the Ebola model of Section 4. In this instance, we were able to substantially reduce the number of contractions, and hence likelihood evaluations, per ElliptSS update while leaving the distribution of accepted angles essentially unchanged. We like to call this a "free lunch".

Centered vs. Non–centered Parameterization

A central computation challenge for fitting hierarchical latent variable models via DA MCMC is that samples can become autocorrelated when alternately updating latent variables and model parameters (Papaspiliopoulos et al., 2003, Papaspiliopoulos et al., 2007). DA MCMC that alternately updates LNA paths and model parameters is no exception (Figure 1A).

This phenomenon of poorly mixing MCMC chains can be traced to the use of a centered parameterization (CP) for the LNA in (14). Under the CP, updates to $\boldsymbol{\theta}|\widetilde{\mathbf{N}}, \mathbf{Y}$ are made conditionally on a *fixed* LNA path. Therefore, parameter proposals are accepted if they are concordant with the data *and* the current path. Small perturbations to model parameters can significantly shift the LNA transition densities and render the current path unlikely under the proposal. This limits the magnitude of perturbations to the model parameters at each MCMC iteration and

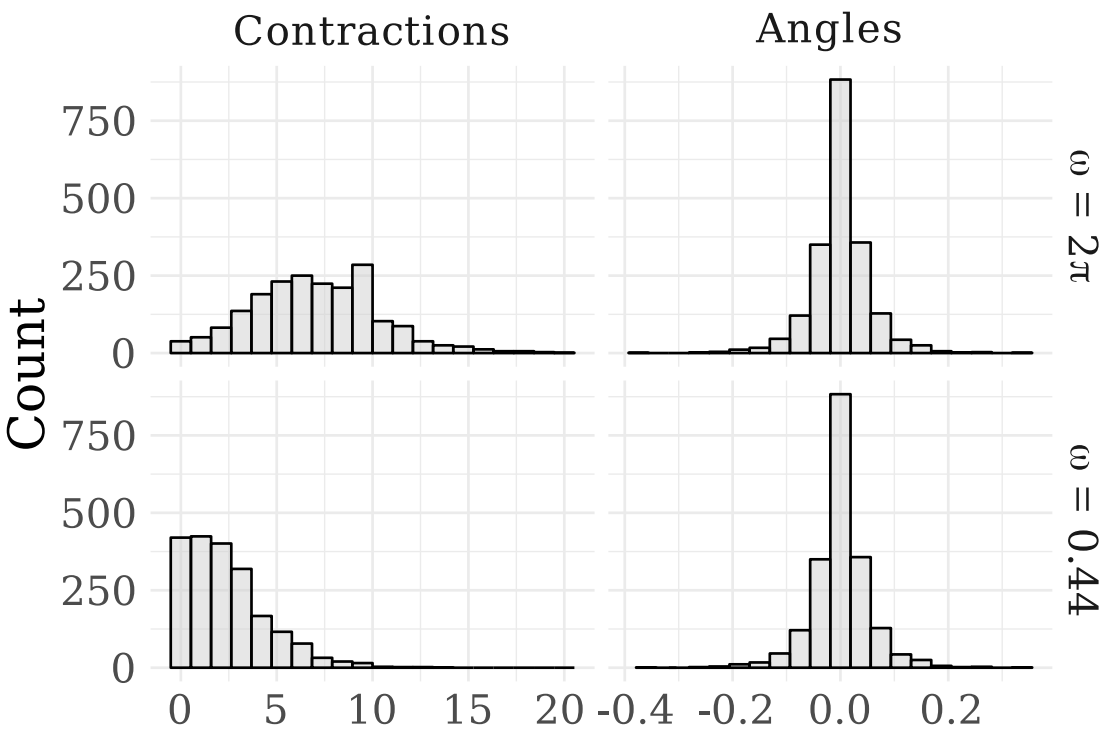

Figure S1: Distributions of the numbers of contractions per ElliptSS update and the accepted angles for an MCMC chain for the Ebola model of Section 4 fit to a simulated dataset. An initial bracket width of 2π was used for the first 5,000 iterations (top row), after which the initial bracket width was 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).

results in severely autocorrelated posterior samples.

The use of a NCP for latent LNA paths massively improves MCMC mixing. Figure 1A shows traceplots of model parameters for one of the MCMC chains for an SIR model fit to Poisson distributed incidence data using the CP. Each MCMC chain was run for 2.5 million iterations, following a tuning run of equal length, but only yielded an effective sample sizes in the low double digits for the basic reproduction number and infectious period duration. In contrast, the NCP yielded effective sample sizes per–chain of between 500–700 for each of the model parameters in only 50,000 iterations following a tuning run of equal length. Figure 1B shows traceplots for one of the NCP MCMC chains. Note that since the posterior distribution of the restarting LNA under the CP does not decompose into the form required to use EllipSS, applying EllipSS to the CP of the restarting LNA amounts to essentially further approximating the random variates in that representation with Gaussian random variables.

Multivariate Normal Slice Sampler

Univariate slice samplers can suffer from poor mixing in moderate– to high–dimensional settings much in the same way as Gibbs samplers. One option for reducing autocorrelation in MCMC samples is to update blocks of parameters. Methods for slice sampling in multiple dimensions are explored in Neal (2003), Thompson (2011), Tibbits et al. (2014). These include slice sampling in hyperrectangles, the use of adaptive Gaussian crumbs that guide slice proposals, and slice sampling along eigenvectors of the estimated posterior covariance matrix.

We present a simple method for sampling a parameter vector, $\boldsymbol{\theta} \in \mathbb{R}^d$, where we perform univariate slice sampling updates along rays drawn from a non–isotropic angular central Gaussian distribution, which is tuned to match the covariance structure of the posterior. This helps to account for linear correlations among model parameters. The method, which we refer to as the multivariate normal slice sampler (MVNSS), is similar to the algorithm in Ahmadian et al. (2011). Our approach differs in that we typically adapt the proposal covariance matrix using a Robbins–Monro recursion. This adaptation is helped by slight modifications to the algorithm during the adaptation phase of the MCMC (Andrieu and Thoms, 2008, Liang et al., 2011). The computational cost of MVNSS does not increase dramatically with the dimensionality of the parameter space, though we have found that multiple MVNSS updates per MCMC iteration can, in some cases, improve performance. The algorithm is amenable to tuning of the initial bracket width as in Tibbits et al. (2014), which helps to reduce the number of likelihood evaluations per iteration.

Suppressing the dependence on the data for notational clarity, slice sampling $\boldsymbol{\theta} \in \mathbb{R}^d$ from its posterior $\pi(\boldsymbol{\theta}|\mathbf{Y})$ is largely the same as sampling $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \propto f(\boldsymbol{\theta})$. Let $\boldsymbol{\Sigma} = \text{Cov}(\boldsymbol{\theta}) = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is the lower triangular matrix of the Cholesky

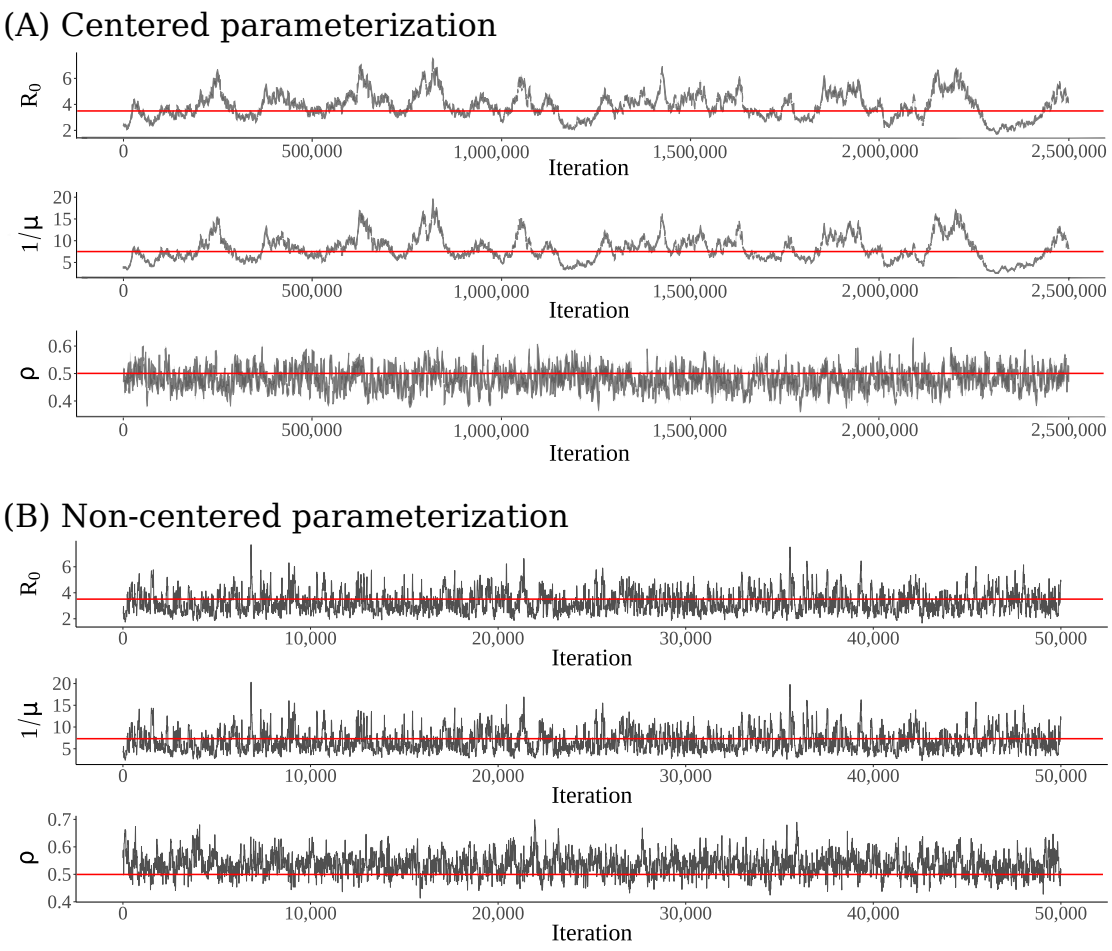

Figure S2: (A) Posterior traceplots from a single MCMC chain for an SIR model fit to Poisson distributed incidence data targeting the centered posterior (14). (B) Posterior traceplots from a single MCMC chain targeting the non–centered posterior (17). MCMC alternated between updating the latent path via elliptical slice sampling, and updating parameters via a multivariate random walk Metropolis algorithm. $R_0 = \beta P/\mu$ 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.

decomposition of $\mathbf{L}$ (any other matrix square root would do). In practice, $\boldsymbol{\Sigma}$ is approximated by $\widehat{\boldsymbol{\Sigma}}_n$, which is estimated over an initial MCMC run. The strategy in MVNSS is to propose $\boldsymbol{\theta}^{prop} = \boldsymbol{\theta}^{cur} + c\boldsymbol{\xi}$, where $\boldsymbol{\xi} = h(\mathbf{z}),\ \mathbf{z} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}),\ h(\mathbf{z}) = \mathbf{z}/||\mathbf{z}||$, and to sample c in a univariate slice sampling update. Normalizing $\mathbf{z}$ allows us to more easily tune the initial bracket width. During an adaptation phase, we construct proposals as $\boldsymbol{\theta}^{prop} = \boldsymbol{\theta}^{cur} + ch(w\boldsymbol{\xi}_1 + (1-w)\boldsymbol{\xi}_2)$, where $\boldsymbol{\xi}_1 = h(\mathbf{z}_1),\ \mathbf{z}_1 \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}),\ \boldsymbol{\xi}_2 = h(\mathbf{z}_2),\ \mathbf{z}_2 \sim MVN(\mathbf{0}, \mathbf{I}_d),\ h(\mathbf{z}) = \mathbf{z}/||\mathbf{z}||$, and $w \in [0, 1]$. The weight given to $\mathbf{z}_2$ is typically quite small, but helps to avoid degeneracy of the empirical covariance matrix during adaptation. We give the non–adaptive version of the algorithm below.

Algorithm 3 Multivariate normal slice sampling with stepping out.

procedure MVNSS($\boldsymbol{\theta}^{cur},\ \mathbf{L},\ S' = (0, \omega)$)
  $u \sim \text{Unif}(0, f(\boldsymbol{\theta}^{cur}))$ ▷ Set threshold
  $\mathbf{z} \sim MVN(\mathbf{0}, \mathbf{I}_d),\ \boldsymbol{\xi} \leftarrow h(\mathbf{Lz})$ ▷ Propose direction
  $p \sim \text{Unif}(0, 1);\ L \leftarrow -\omega p,\ U \leftarrow L + \omega$ ▷ Position S' around 0
  $S' \leftarrow$ StepOut(u, S') ▷ Step out bracket
    while $u < f(\boldsymbol{\theta} - L\boldsymbol{\xi})$ **do** $L \leftarrow L - \omega$
    while $u < f(\boldsymbol{\theta} + U\boldsymbol{\xi})$ **do** $U \leftarrow U + \omega$
  $c \sim \text{Unif}(L, U);\ \boldsymbol{\theta}^{prop} \leftarrow \boldsymbol{\theta}^{cur} + c\boldsymbol{\xi}$ ▷ Propose new value
  if $f(\boldsymbol{\theta}^{prop}) > u$ **then** $\boldsymbol{\theta}^{new} \leftarrow \boldsymbol{\theta}^{prop}$ ▷ Accept proposal
    return $\boldsymbol{\theta}^{new}$
  else ▷ Shrink bracket
    if $c < 0$ **then** $L \leftarrow c$ **else** $U \leftarrow c$
    GoTo 8

Adapting the proposal covariance in MVNSS

Adaptive MCMC algorithms aim to improve computational efficiency by using MCMC samples to learn optimal values of tuning parameters on the fly. MCMC proposal kernels can be adapted in a number of different ways, but must be adapted with care to preserve the stationarity of the target distribution. In order for an adaptive MCMC algorithm to preserve the stationary distribution, it must satisfy two conditions, vanishing adaptation, and bounded convergence (Andrieu and Thoms, 2008).

The main computational tool used in the adaptive variations of the MCMC algorithms in this dissertation is the Robbins–Monro recursion, which allows us to continuously adapt the tuning parameters of an MCMC kernel. The Robbins–Monro recursion is a stochastic approximation algorithm that searches for a solution to an equation, $f(\theta) = \alpha$, that has a unique root at $\theta^\star$. The function $f(\theta)$ is not

directly observed. Instead, we use a noisy sequence of estimates, $h(\widehat{\theta}_n)$, satisfying $\mathrm{E}(h(\widehat{\theta}_n)) = f(\theta)$ to recursively approximate $\theta^\star$. The recursion takes the form,

$$\theta_{n+1} = \theta_n + \gamma_{n+1}(h(\widehat{\theta}_n) - \theta^\star).$$

Hence, the recursion increments θ by an amount proportional to the difference between $h(\widehat{\theta}_n)$ and its target. The gain factor sequence, $\{\gamma_{n+1}\}$, is a deterministic non–increasing, positive sequence such that

$$(i)\ \lim_{n\to\infty} \gamma_n = 0, \qquad (ii)\ \sum_{n=1}^{\infty} \gamma_n = \infty, \qquad (iii)\ \sum_{n=1}^{\infty} \gamma_n^{1+\lambda} < \infty,\ \lambda > 0.$$

Note that since $\gamma_n \to 0$ and $\mathrm{E}(h(\widehat{\theta}_n)) = f(\theta)$, it follows that $|\theta_{n+1} - \theta_n| \to 0$ as $n \to \infty$, i.e., the recursion is constructed to satisfy diminishing adaptation. Condition (ii) ensures that the gain sequence does not decay so fast that there are values of θ in its state space, Θ, that cannot be reached. Condition (iii) ensures bounded convergence of the sequence $\{\theta_n\}$. Gain factor sequences of the form

$$\gamma_n = C(1 + pn)^{-\alpha},\ \alpha \in (0.5, 1],\ p > 0 \tag{22}$$

will satisfy these conditions (Andrieu and Thoms, 2008, Liang et al., 2011). We adapt the proposal covariance over the course of an initial MCMC tuning run, which is followed by a final run with a fixed MCMC kernel. The samples accumulated during the adaptation phase are discarded.

Let μ_n and Σ_n denote the empirical mean and covariance of the posterior samples from the first n MCMC iterations, and $\{\gamma_n\}$ be a sequence of gain factors. In each adaptive MCMC iteration, we sample $\boldsymbol{\theta}^{new}|\boldsymbol{\theta}^{cur}$ via MVNSS, and update the empirical mean and covariance via the following recursions:

$$\begin{aligned}
\boldsymbol{\mu}_n &= \boldsymbol{\mu}_{n-1} + \gamma_n(\boldsymbol{\theta}_n^{new} - \boldsymbol{\mu}_{n-1}), \\
\boldsymbol{\Sigma}_n &= \boldsymbol{\Sigma}_{n-1} + \gamma_n \left((\boldsymbol{\theta}_n^{cur} - \boldsymbol{\mu}_{n-1})(\boldsymbol{\theta}_n^{new} - \boldsymbol{\mu}_{n-1})^T - \boldsymbol{\Sigma}_{n-1}\right).
\end{aligned}$$

Initializing the LNA Draws

In simple models, biologically plausible parameter values will generally lead to valid LNA paths, and we can initialize the LNA draws by simply drawing $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$. 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 resampling step after line 6 in Algorithm 1, in which $\mathbf{Z}(t_\ell)$ is redrawn in place until we have obtained a valid LNA path that respects the constraints on the latent state space. However, such a procedure does not sample from the correct distribution since $\mathbf{Z}$ is not distributed as a truncated multivariate Gaussian. To correct

for this, we "warm–up" the LNA path with an initial run of ElliptSS iterations in which the likelihood only consists of the indicators for whether the path is valid. Note that ElliptSS, or any other valid MCMC algorithm for updating $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$, will never lead to an invalid LNA path being accepted if the current LNA draws and model parameters correspond to a valid path. Similarly, any valid MCMC algorithm for updating model parameters conditional on LNA draws will also preserve the validity of LNA paths.

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:

$$\mathbf{X}_0 \sim TMVN_{\mathcal{S}_X^R}(P\mathbf{p}, \alpha P(\mathbf{P} - \mathbf{p}\mathbf{p}^T)), \tag{23}$$

where $\mathbf{p}$ is a vector of subject–level initial state probabilities, $\mathbf{P} = \text{diag}(\mathbf{p})$, P is the population size, α is an over–dispersion parameter, and the subscript $\mathcal{S}_X^R$ specifies the state space of $\mathbf{X}$ (so that the compartment volumes add up to P and each compartment volume is non–negative and less than the total population size at time t_0). Thus, the initial distribution is the truncated normal approximation of either a multinomial distribution with size P and probability vector $\mathbf{p}$ if $\alpha = 1$, or of a dirichlet–multinomial distribution with parameters $\boldsymbol{\alpha} \implies \mathbf{p} = \boldsymbol{\alpha}/\boldsymbol{\alpha}^T\mathbf{1}$, and over–dispersion $\alpha = (P + \boldsymbol{\alpha}^T\mathbf{1})/(1 + \boldsymbol{\alpha}^T\mathbf{1})$. 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 clarity.

Let $\mathbf{m} = P\mathbf{p}$, $\mathbf{V} = \alpha P(\mathbf{P} - \mathbf{p}\mathbf{p}^T)$, 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 4 Sampling LNA draws and initial volumes via elliptical slice sampling.

procedure doElliptSS2($\mathbf{Z}^X_{cur}, \mathbf{Z}^{X_0}_{cur}, \boldsymbol{\theta}, \mathbf{Y}, \mathcal{I}, \omega = 2\pi$)

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:

$$\psi \sim \text{Unif}(0, \omega)$$
$$L_\psi \leftarrow -\psi;\ R_\psi \leftarrow L_\psi + \psi$$
$$\phi \sim \text{Unif}(L_\psi, R_\psi)$$

Make the initial proposal:

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

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

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 straightforward 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 an SEIR model fit to incidence data from Sierra Leone. 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 effective population size is roughly the size of the initially susceptible population. The model parameters on their natural scales are provided in Table S1. Each of the model parameters is obviously interpretable on its natural scale, but upon examining the pairwise scatterplots of the posterior (Figure S3) 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).

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 S2 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 adjusted reproductive number.

A heuristic argument for an estimation scale that further simplifies the posterior

geometry proceeds by analogy with the analogous deterministic ODE model. In particular, we will consider how the functions of model parameters in Table S2 act on the model dynamics through the final size relation, and how they are informed by the data. The final size relation for the ODE model (Miller, 2012) relates the fraction of the population that eventually becomes infected π, with the basic reproductive number:

$$\pi = 1 - \mathrm{e}^{-R0\pi}. \tag{24}$$

As $R0$ increases, a larger fraction of the population becomes infected. As the effective population size, P_{eff}, increases, so too does the overall scale of the outbreak, and the number of cases we would expect to detect. The expected scale of the observed outbreak is related to the quantity, $\pi \times \rho \times P_{eff}$, which should be concordant with the total number of observed cases. We can interpret the product, $\rho \times P_{eff}$, as a measure of scale the observed incidence data. An implication of this is that the combination of parameters we would expect to jointly act on the model, *a posteriori*, is the adjusted reproduction number offset by the case detection rate, $R0 \times \rho \times P_{eff}$, which will enter our estimation scale as $\log(R0 \times \rho \times P_{eff})$. The other modification of the estimation scale in Table S2 replaces the log latent period with the log ratio of infectious to latent period durations. The new estimation scale is given in Table S3. On this estimation scale, the posterior is much better behaved, with weaker pairwise correlations and little in the way of non–linear relationships between the model parameters.

Table S1: SEIR model parameter and their interpretation on their natural scales.

Param.	**Interpretation**	**Domain**
α	Rate of exogenous infection	$[0, \infty)$
β	Per–contact endogenous infection rate	$[0, \infty)$
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	Neg. binom. overdispersion parameter	$[0, \infty)$
P_{eff}	Effective population size	$[0, N]$

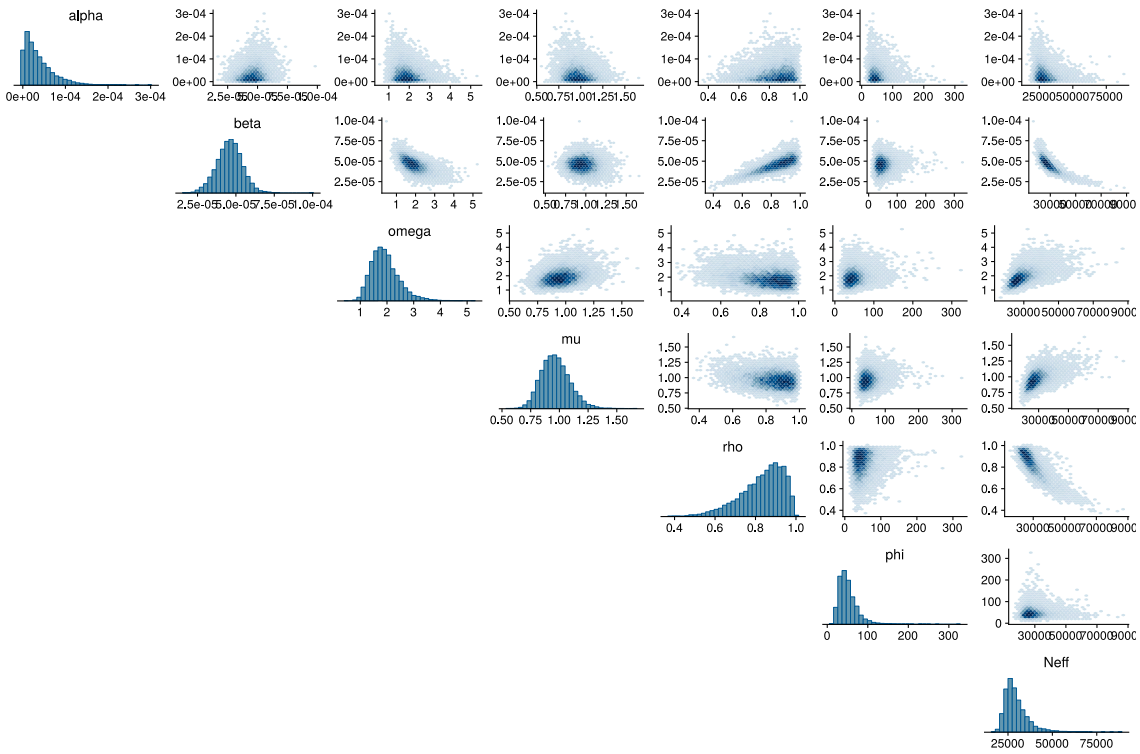

Figure S3: Marginal histograms and pairwise scatterplots of posterior samples for parameters for the SEIR model fit to the Sierra Leone Ebola dataset using the estimation scale in Table S3. The parameters on the estimation scales in this figure and their interpretations are provided in Table S1.

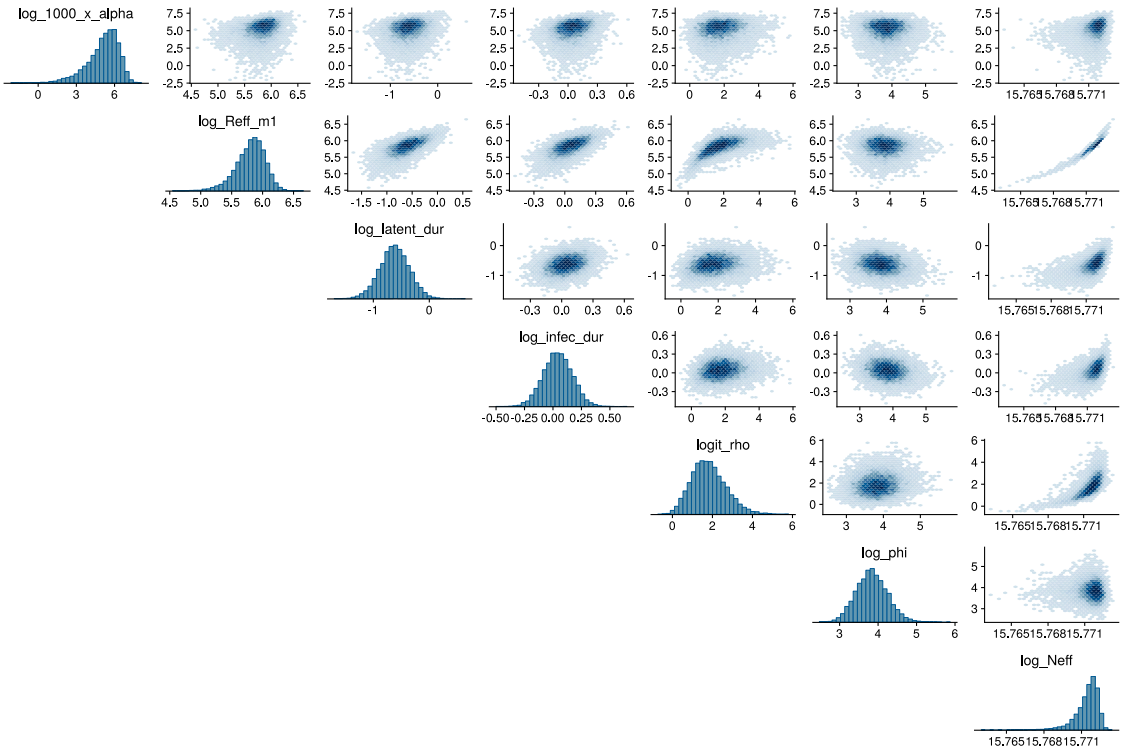

Figure S4: Marginal histograms and pairwise scatterplots of posterior samples for parameters for the SEIR model fit to the Sierra Leone Ebola dataset using the estimation scale in Table S3. The parameters on the estimation scales in this figure and their interpretations are provided in Table S2.

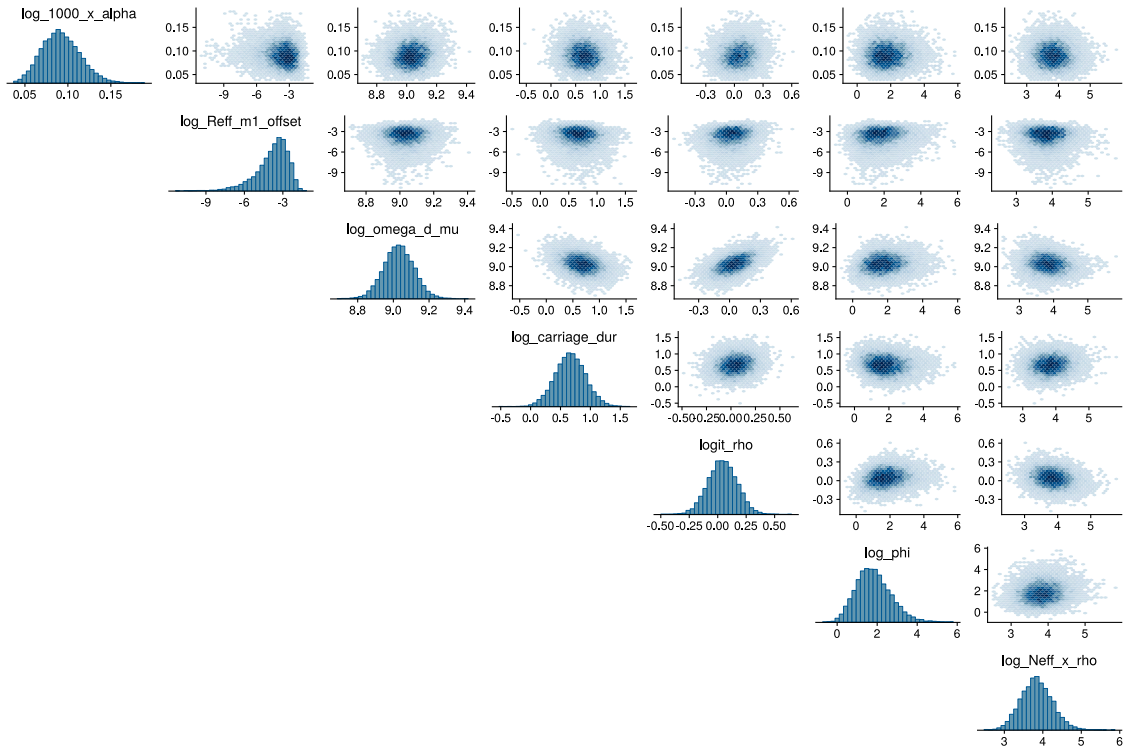

Figure S5: Marginal histograms and pairwise scatterplots of posterior samples for parameters for the SEIR model fit to the Sierra Leone Ebola dataset using the estimation scale in Table S3. The parameters on the estimation scales in this figure and their interpretations are provided in Table S3.

Table S2: SEIR model parameter and their interpretation on a possible set of estimation scales.

Parameter	Interpretation	Doma
$\log(1000\alpha)$	Log effective number of additional infecteds per 1000 foreign infecteds	$(-\infty, \infty$
$\log(R_{adj} - 1) = \log(\beta P_{eff}/\mu - 1)$	Log basic reproductive number for native index case and $R_{adj} > 1$.	$(-\infty, \infty$
$\log(1/\omega)$	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	$(-\infty, \infty$
$\log(P_{eff})$	Log effective population size	$(-\infty, \log$

Table S3: SEIR model parameters and their interpretation on a possible set of estimation scales.

Parameter	Interpretation
$\log(1000\alpha)$	Log effective number of additional infecteds per 1000 infecteds outside the population
$\log(R_{adj} - 1) + \log(\rho P_{eff})$	Log adjusted reproductive number for native index case with offset, and $R_{adj} > 1$
$\log(\omega/\mu)$	Log ratio of mean latent to infectious period durations
$\log(1/\mu)$	Log mean infectious period duration
$\text{logit}(\rho)$	Logit mean case detection probability
$\log(\phi)$	Log negative binomial overdispersion parameter
$\log(\rho P_{eff})$	Log mean case detection rate

Web Appendix C: Section 3 — Simulation Setup and MCMC Details

Simulation Setup

In this simulation, repeated for each of the three different regimes of population size and initial conditions given in Table S4, we simulated 500 datasets as follows:

1. Draw $\log(R0 - 1),\ 1/\mu,\ \text{logit}(\rho),\ \log(\phi)$ from the priors given in Table S4.

2. Simulate an outbreak, $\mathbf{N}|\boldsymbol{\theta}$, under SIR dynamics from the MJP via Gillespie's direct algorithm (Gillespie, 1976). 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 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 S4. 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.

For models fit via the LNA and ODE approximations, the covariance matrix was adapted as in algorithm 4 of Andrieu and Thoms (2008). 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 (Fintzi, 2018).

Inference via the MMTL approximation within PMMH were fit using the `pomp` R package (King et al., 2016b). 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 of the form $\gamma_n = n^{\alpha}$, where the cooling term, α, was set to 0.999. For some of the datasets, the PMMH algorithm became degenerate during the adaptive phase of the MCMC. When this was the case, the MCMC was restarted at a different set of random initial conditions. The posterior samples from all five MCMC chains were combined after discarding the initial samples from the adaptation phase.

Additional Coverage Simulation Results

Table S4: Population sizes, initial conditions, and priors under which datasets were simulated. Five hundred datasets were simulated for each of the population size regimes. Each 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)**
$R_0 - 1$	Basic reproduction # - 1	LogNormal(0, 0.5)	$\implies R_0 = 2.00$ (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. over–dispersion	Exponential(0.1)	6.93 (0.25, 36.89)

Table S5: Median (2.5%, 97.5%) quantiles of average run time per MCMC chain, in minutes, for models fit in the coverage simulation presented in Section 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).

Population size	**ODE**	**LNA**	**MMTL**
10,000	2 (2.1, 2.2)	1.2 (1.3, 1.4)	2.9 (2.9, 3.8)
50,000	108.6 (161.5, 165.2)	55.6 (67.6, 63.4)	187.1 (274.5, 360.8)
250,000	429.2 (443.8, 439.8)	212.7 (197.9, 202.3)	759.1 (761, 831.7)

Table S6: Detailed small population (N = 10,000) regime results for the coverage simulation presented in Section 3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings (PMMH), 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 total effective sample size divided by the total CPU time (ESS/CPU time).

Method	**Parameter**	**Coverage**	**PMD**	**95% CIW**	**ESS**	**ESS/CPU time**
ODE	$\log(R_0)$	0.93	-0.03 (-0.76, 0.56)	1.03 (0.65, 1.38)	6438 (4956, 7650)	197069.12 (121822.81, 328146.02)
ODE	$\log(\mu)$	0.95	0.05 (-0.51, 0.79)	0.91 (0.53, 1.2)	6420 (4908, 7663)	196036.49 (121435.19, 332097.76)
ODE	$\text{logit}(\rho)$	0.93	0.12 (-1.03, 1.37)	1.25 (0.48, 2.94)	6237 (4346, 7556)	190241.73 (113179.5, 316262.46)
ODE	$\log(\phi)$	0.96	-0.25 (-1.64, 0.54)	1.25 (0.88, 2.11)	6529 (5459, 7778)	203807.37 (127287.69, 316314.92)
LNA	$\log(R_0)$	0.95	0 (-0.51, 0.55)	1.04 (0.83, 1.36)	1340 (274, 4029)	746.74 (95.83, 3720.42)
LNA	$\log(\mu)$	0.95	-0.01 (-0.5, 0.48)	0.93 (0.74, 1.13)	1024 (199, 3988)	575.02 (73.92, 3659.91)
LNA	$\text{logit}(\rho)$	0.94	-0.05 (-1.44, 0.56)	1.18 (0.54, 2.86)	1225 (316, 3941)	678.89 (120.47, 3927.53)
LNA	$\log(\phi)$	0.96	0 (-0.66, 0.71)	1.35 (0.92, 2.22)	2021 (687, 4535)	1098.84 (236.9, 4519.87)
MMTL	$\log(R_0)$	0.89	0.03 (-0.48, 0.55)	1.09 (0.88, 1.39)	7483 (5453, 9197)	1056.93 (535.23, 2203.76)
MMTL	$\log(\mu)$	0.86	-0.03 (-0.51, 0.45)	0.92 (0.75, 1.09)	7481 (5442, 9197)	1048.48 (525.15, 2264.44)
MMTL	$\text{logit}(\rho)$	0.72	-0.03 (-1, 0.61)	1.17 (0.51, 2.98)	6725 (4328, 8506)	970.67 (432.17, 1924)
MMTL	$\log(\phi)$	0.75	-0.02 (-0.69, 0.71)	1.33 (0.91, 2.21)	7486 (5405, 9215)	1037.88 (528.07, 2038.94)

Table S7: Detailed medium population (N = 50,000) regime results for the coverage simulation presented in Section 3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings (PMMH), 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 total effective sample size divided by the total CPU time (ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	ESS/CPU time
ODE	$\log(R_0)$	0.93	-0.02 (-0.56, 0.59)	0.84 (0.52, 1.27)	6592 (5213, 7771)	190145.76 (121327.57, 303671.72)
ODE	$\log(\mu)$	0.93	-0.01 (-0.66, 0.47)	0.73 (0.44, 1.07)	6558 (5129, 7664)	190261.35 (119467.86, 301261.95)
ODE	$\text{logit}(\rho)$	0.95	-0.03 (-1.13, 0.85)	0.88 (0.37, 2.75)	6421 (5017, 7643)	184747.84 (111231.49, 299349.25)
ODE	$\log(\phi)$	0.94	-0.17 (-1.36, 0.57)	1.04 (0.78, 1.7)	6637 (5452, 7755)	191414.32 (127413.28, 290226.21)
LNA	$\log(R_0)$	0.95	-0.03 (-0.5, 0.49)	0.93 (0.66, 1.28)	2158 (397, 4981)	792.2 (103.73, 4065.6)
LNA	$\log(\mu)$	0.93	0.01 (-0.43, 0.41)	0.83 (0.58, 1.07)	1919 (307, 5276)	668.29 (81.33, 4149.88)
LNA	$\text{logit}(\rho)$	0.95	-0.01 (-1, 0.5)	1.13 (0.47, 2.81)	1704 (439, 4636)	624.23 (109.45, 3443.11)
LNA	$\log(\phi)$	0.94	0.01 (-0.57, 0.7)	1.1 (0.79, 1.86)	2916 (1179, 5442)	1078.91 (343.56, 4154.25)
MMTL	$\log(R_0)$	0.86	-0.01 (-0.46, 0.51)	0.98 (0.67, 1.32)	7284 (4713, 9107)	981.83 (513.39, 2404.97)
MMTL	$\log(\mu)$	0.82	-0.01 (-0.47, 0.37)	0.82 (0.58, 1.05)	7167 (4488, 8912)	961.27 (505.32, 2396.12)
MMTL	$\text{logit}(\rho)$	0.75	-0.03 (-0.98, 0.56)	1.09 (0.43, 2.96)	6452 (3999, 8318)	874.11 (402.85, 2138.24)
MMTL	$\log(\phi)$	0.82	-0.02 (-0.59, 0.65)	1.1 (0.79, 1.84)	7096 (4640, 8999)	980.97 (513.13, 2092.91)

Table S8: Detailed large population (N = 250,000) regime results for the coverage simulation presented in Section 3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ–leaping (MMTL) within particle marginal Metropolis–Hastings (PMMH), 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 total effective sample size divided by the total CPU time (ESS/CPU time).

Method	**Parameter**	**Coverage**	**PMD**	**95% CIW**	**ESS**	**ESS/CPU time**
ODE	$\log(R_0)$	0.95	-0.01 (-0.37, 0.55)	0.65 (0.36, 1.22)	6828 (5566, 8025)	184765.65 (108850.36, 292911.44)
ODE	$\log(\mu)$	0.95	0 (-0.49, 0.34)	0.55 (0.3, 1)	6815 (5520, 7877)	185289.87 (109182.94, 290079.18)
ODE	$\text{logit}(\rho)$	0.93	-0.01 (-0.78, 0.68)	0.74 (0.27, 2.62)	6575 (5219, 7838)	180225.91 (104056.77, 286249.25)
ODE	$\log(\phi)$	0.95	-0.08 (-0.59, 0.46)	0.9 (0.64, 1.44)	6721 (5571, 7683)	181453.2 (109830.37, 281222.09)
LNA	$\log(R_0)$	0.93	-0.01 (-0.38, 0.51)	0.77 (0.47, 1.24)	3248 (638, 6127)	1194.75 (109.38, 4862.77)
LNA	$\log(\mu)$	0.94	0.01 (-0.44, 0.33)	0.67 (0.4, 1.02)	3134 (499, 6038)	1164.4 (91.22, 4937.49)
LNA	$\text{logit}(\rho)$	0.92	0.01 (-0.8, 0.57)	0.99 (0.35, 2.66)	2514 (572, 5905)	950.41 (103.59, 4628.48)
LNA	$\log(\phi)$	0.95	0 (-0.44, 0.54)	0.94 (0.67, 1.5)	3987 (2201, 6184)	1450.67 (421.38, 4654.55)
MMTL	$\log(R_0)$	0.89	0.03 (-0.38, 1.04)	0.8 (0.31, 1.27)	7030 (3789, 8915)	941.38 (387.8, 2279.51)
MMTL	$\log(\mu)$	0.89	-0.02 (-0.46, 0.32)	0.65 (0.38, 0.98)	6858 (3602, 8871)	913.59 (384.87, 2278.62)
MMTL	$\text{logit}(\rho)$	0.84	0.01 (-0.7, 0.65)	0.95 (0.34, 2.64)	6166 (3282, 7906)	849.92 (329.3, 2033.99)
MMTL	$\log(\phi)$	0.91	-0.03 (-0.51, 0.52)	0.94 (0.66, 1.5)	6266 (3735, 8960)	853.25 (448.79, 2032.21)

Web Appendix D: Identifiability when Estimating the Effective Population Size

When the scale of an outbreak is small relative to the population size, it may be unreasonable to assume that the entire population mixes homogeneously and participates in propagating the epidemic. An alternative is to split the population into two sub–populations, one that is effectively removed from infectious contact, and another at–risk sub–population within which infectious contacts arise via homogeneous mixing of infectious and susceptible individuals. For example, in the case of the SIR model, we split susceptibles into two compartments, S^R and S^E, where S^R is a susceptible sub–population that is effectively removed, and S^E is a susceptible population that may become exposed. In this model, the *effective population size* is $P_{eff} = S^E + I + R$, and is interpreted as the size of the population at risk of infection. The transmission model is otherwise constructed in the same way as the canonical SIR model, except with S^E replacing S.

The effective population size and the mean case detection probability are weakly identifiable parameters in that we require prior information about their scales to disentangle their effects. To see why this is so, note that ρ and P_{eff} enter into the complete data likelihood, (16), through the priors and through emission distributions that have means of the form, $\rho\Delta N_{SI}(t_\ell)$. The incidence here should be understood as an increment in compartment concentrations scaled by the effective population size, since the LNA is a density dependent process (Komorowski et al., 2009, Wilkinson, 2011, Fearnhead et al., 2014). Thus, the emission densities have means of the form, $\rho\left(N'_{S^SI}(t_\ell) - N'_{S^SI}(t_{\ell-1})\right)P_{eff}$, where $\mathbf{N}' = \mathbf{N}/P_{eff}$ is the equivalent representation of the LNA in terms of compartment concentrations.

Absent prior information, we will have difficulty identifying both the case detection probability and effective population size. Despite this, there are a few reasons that ρ and P_{eff} might be weakly identifiable, rather than completely unidentifiable. First, certain stochastic aspects of the outbreak, such as the probability of a major outbreak and persistence of transmission, depend on the population size. Furthermore, the scale of the observed incidence and the observed outbreak duration are informative about the minimal effective population size. Therefore, the fact that we observe part of the outbreak is itself informative. Finally, a rough estimate the true population size is typically available, providing an upper bound on the effective population size. By extension, the latent epidemic process is also weakly identifiable in models where the effective population size and case detection probability are estimated.

In contrast, the effective detectable population at risk, $\rho \times P_{eff}$, along with parameters governing the outbreak dynamics, are directly informed by the data and remain identifiable. It might seem paradoxical that we can infer the dynamics of an outbreak when we are unable to estimate the latent process. To understand why

this is so, note that a SEM can be rewritten in terms of concentrations by dividing the compartment counts by the population size, yielding the so–called "true mass–action model". The dynamics of this equivalent model, expressed, for example, by the basic reproductive number $R0$ and recovery rate for an SIR model, are known to be independent of the population size (de Jong et al., 1995). Combinations of model parameters yield latent paths, $\mathbf{N}'$, expressed in terms of increments in concentrations, and are weighted in the posterior proportionally (modulo the prior) to the likelihood of scaled paths, $\rho P_{eff}\mathbf{N}'$. The temporal nature of the data is important here because the curvature of scaled paths should roughly match that of the data. Thus, we are leveraging curvature in the data to make inferences, not just the pointwise emission probabilities.

We check that $\rho \times P_{eff}$ and the parameters governing the outbreak dynamics are identifiable via a simple simulation. We drew 500 sets of parameters from the priors given in Table S9, and simulated an outbreak and a dataset for each set of parameters. The models were fit via the LNA under the same priors from which the parameters were drawn using the MCMC procedure used to fit models for the main coverage simulation (described in the first section of Web Appendix C). The results are summarized in Table S10. The nominal coverage rates for all model parameters is approximately correct. However, the relative widths of the posterior credible intervals for the case detection rate are substantially narrower than the relative widths of the credible intervals for ρ and P_{eff} vis–a–vis their prior intervals. This suggests that the data are informative about $\rho \times P_{eff}$. The priors for ρ and P_{eff} are not completely flat, and the scale of the observed counts is itself informative about P_{eff}. Therefore, we still expect, and observe, some contraction in the posteriors for ρ and P_{eff}, individually.

Table S9: Parameters and priors used in simulating 500 SIR outbreaks where the effective population size was a parameter in the model. The true population size was 100,000. Each outbreak was simulated from a MJP with SIR dynamics. The observed incidence was a negative binomial sample of the true incidence.

Parameter	Interpretation	Prior	Median (95% Interval)
$R0 - 1$	Basic reproduction # - 1	LogNormal(0, 0.5)	$\implies R0 = 2.00$ (1.38, 3.66)
$1/\mu$	Mean infectious period	LogNormal(0.7, 0.35)	2.01 (1.01, 4.00)
$\rho/(1-\rho)$	Odds of case detection	LogNormal(0, 1.4)	$\implies \rho = 0.5$ (0.06, 0.94)
ϕ	Neg.Binom. overdispersion	Exponential(0.1)	6.93 (0.25, 36.89)
P_{eff}	Effective population size	Unif(5000, 50000)	27500 (6125, 48875)

Table S10: Results for the models fit to the outbreaks simulated under SIR dynamics with random effective population sizes. Reported are the nominal coverage rates of 95% credible intervals, the median (95% CI) of the posterior median deviations (PMD), credible interval widths (CIW), and credible interval widths relative to the 95% prior interval widths (Rel.CIW). The relative widths of the credible intervals for $\rho \times P_{eff}$ are computed with respect to the induced prior resulting from the marginal priors for ρ and P_{eff}.

Parameter	Coverage	PMD	CIW	Rel.CIW
R_0	0.94	-0.02 (-0.9, 0.58)	1.1 (0.47, 2.43)	0.48 (0.21, 1.06)
μ	0.95	0 (-0.36, 0.21)	0.49 (0.29, 0.85)	0.66 (0.4, 1.16)
ρ	0.95	-0.01 (-0.36, 0.24)	0.55 (0.14, 0.74)	0.62 (0.16, 0.84)
P_{eff}	0.96	800 (-18900, 16100)	32200 (14600, 41100)	0.75 (0.34, 0.96)
$\rho \times P_{eff}$	0.93	35 (-9000, 4200)	6750 (640, 26750)	0.18 (0.02, 0.72)
ϕ	0.95	0.04 (-13.49, 8.95)	9.1 (0.31, 46.85)	0.25 (0.01, 1.28)

Web Appendix E: Modeling the Spread of Ebola in West Africa

We modeled the spread of Ebola in Guinea, Liberia, and Sierra Leone under SEIR transmission dynamics within each country, incorporating cross–border transmission via virtual migration of infectious individuals. The model was first fit to a dataset simulated from the true model under known parameters, and then to incidence data from the 2013–2015 West Africa outbreak, each time using both the LNA and ODE approximation computation. The priors and model fitting procedures were largely the same for each dataset and for each approximation. Hence, the following sections apply to all Ebola models.

Model Parameters and Rates of State Transition

Priors for Model Parameters

Priors for Effective Population Sizes and Initial Compartment Counts

At the time when transmission was assumed to commence in each country, susceptible individuals in the population were considered to either be geographic or social proximate to the transmission process, in which case they could possibly become exposed, or were detached from the transmission process. We separate the S compartment into a compartment for individuals who are susceptible and connected to

Table S11: Parameters and their interpretations. Subscripts, A, B, indicate countries. All parameters governing rates of state transition for Liberia and Sierra Leone are zero until three weeks prior to the first detected case, when transmission was assumed to commence in each country.

Parameter	Interpretation	State Transition
$\beta_A(t)$	Per–contact rate of transmission within country A.	$S_A^E \to E_A$
$\alpha_{AB}(t)$	Per–contact rate of transmission from country A to B.	$S_A^E \to E_A$
$\omega_A(t)$	Rate at which latent individuals become infectious.	$E_A \to I_A$
$\mu_A(t)$	Rate at which infectious individuals recover.	$I_A \to R_A$
$P_{eff,A}$	Effective population size.	—
ρ_A	Mean case detection rate.	—
ϕ_A	Negative binomial overdispersion.	—

Table S12: Rates of state transition. Subscripts for rates indicate model compartments and superscripts indicate countries, while subscripts for compartments and parameters indicate countries. All rates of state transition for Liberia and Sierra Leone are zero until three weeks prior to the first detected case, when transmission was assumed to commence in each country.

Rate	State Transition
$\lambda^A_{S^E E}(t) = \beta_A(t)\left(I_A + \alpha_{BA}(t)I_B + \alpha_{CA}(t)I_C\right) S_A^E$	$S_A^E \to E_A$
$\lambda^A_{EI}(t) = \omega_A(t)E_A$	$E_A \to I_A$
$\lambda^A_{IR}(t) = \mu_A(t)I_A$	$I_A \to R_A$

Table S13: Parameters for the West Africa Ebola outbreak model, prior distributions, 95% prior intervals, and references that informed the choice of priors. Subscripts, G, L, S, indicate specific countries, or generic countries A, B if a prior is common to several parameters. Adjusted reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta P_{eff}/\mu$. The mean hyper–parameters of priors for effective population sizes given in the table were used in modeling the West Africa outbreak. The mean hyper–parameters used in the simulated data example were 9.8, 10.5, and 10.6 for Guinea, Liberia, and Sierra Leone, respectively. The negative binomial overdispersion rate hyper–parameter in the simulated data example was also set to 1, reflecting that we expected less overdispersion in the setting where the model was not possibly misspecified.

Parameter	**Interpretation**	**Prior**	**Med. (90% Prob. Interval)**	**References**
$R_{adj,A} - 1$	Adjusted reproduction #-1	LogNormal(log(0.5), 1.08^2)	$\implies R_{adj,A} = 1.50(1.08, 3.95)$	Chowell (2014); Chretien (2015); Coltart (2017)
$\frac{\alpha_{AB} P_{eff,B}}{\mu_A}$	Adjusted reproduction #, $A \to B$	Exponential(rate=40)	$R_{adj,AB} = 0.017$ $(0.001, 0.075)$	Low cross–border transmission; Dudas (2017)
ω_A/μ_A	Relative latent vs. infectious period	LogNormal(0.0, 0.3^2)	$(1/\mu_A)/(1/\omega_A) = 1.00$ (0.56, 1.80)	Chowell (2014); Velasquez (2015)
$1/\mu_A$	Infectious period duration	LogNormal(0, 0.3^2)	$7/\mu_A = 7$ (3.9, 11.5)	Chowell (2014); Velasquez (2015)
$P_{eff,G}$	Effective population size	LogNormal(9.6, 0.62^2)	$P_{eff,G} = 14765$ (4380, 40938)	Scale of counts, see discussion below
$P_{eff,L}$	Effective population size	LogNormal(9.9, 0.62^2)	$P_{eff,L} = 19930$ (5912, 55260)	Scale of counts, see discussion below
$P_{eff,S}$	Effective population size	LogNormal(10.7, 0.62^2)	$P_{eff,S} = 44356$ (13158, 122984)	Scale of counts, see discussion below
ρ_A	Mean case detection prob.	LogitNormal(0.85, 0.75^2)	$\implies \rho_A = 0.7, (0.35, 0.89)$	Very high and very low ρ unlikely; Scarpino (2014); Dalziel (2018)
$1/\sqrt{\phi_A}$	Neg.Binomial overdispersion	Exponential(0.69)	$\implies = 0.69(0.05, 3)$	—

Table S14: Estimation scales and interpretations for Ebola model parameters. Subscripts, A, B indicate countries.

Parameter	Interpretation	Support
$\log(\beta_A P_{eff,A}/\mu_A - 1) + \log(P_{eff,A}\rho_A)$	Log adjusted effective endogenous reproduction # - 1, offset by the magnitude of the observed outbreak	$(-\infty, \infty)$
$\log(\alpha_{AB} P_{eff,B}/\mu_A)$	Log adjusted effective exogenous reproduction #, $A \to B$	$(-\infty, \infty)$
$\log(\omega_A/\mu_A)$	Log relative duration, latent vs. infectious period	$(-\infty, \infty)$
$\log(1/\mu_A)$	Log infectious period duration	$(-\infty, \infty)$
$\log(P_{eff,A}\rho_A)$	Log scale of detected outbreak	$(-\infty, \infty)$
$\text{logit}(\rho_A)$	Log odds of case detection	$(-\infty, \infty)$
$\log(1/\sqrt{(}\phi_A))$	Log overdispersion	$(-\infty, \infty)$

exposure, S^E, and individuals who are susceptible but effectively removed, S^R. The fraction of the initially susceptible population that could possibly become exposed corresponds the effective population size. We identify the scale of the prior for S^E by matching the total number of cases that were detected in each country to the number of cases that we would expect to detect given the effective population size under deterministic outbreak with SEIR dynamics.

The final size relation for the SEIR model (Miller, 2012) relates the fraction of the population that becomes infected, π, to the basic reproduction number via:

$$(1 - \pi) = \exp^{-\pi R_0}.$$

Tables S15 and S16 give the effective fraction of the population that could possibly become exposed in each country in order for the expected number of detected cases under deterministic SEIR dynamics to match the observed number of cases under a range of values for the adjusted reproduction number, R_{adj}, and mean case detection rate, ρ.

In our Ebola models, we used a multivariate normal approximation to a multinomial distribution for the initial distribution of individuals. The hyperparameter for country A, was set to

$$\boldsymbol{\alpha}_A = (S_A = N_A - 30, E_A = 15, I_A = 10, R_A = 5),$$

and we compute the effective number of susceptibles as $S_A^E = S - S_A^R$. The effective population size is $P_{eff,A} = S_A^E + E_A + I_A + R_A$.

Ebola Model MCMC Details for Models Fit via the LNA and ODE

The model fitting procedure and priors were the same for both models, and were also the same for models fit via the LNA and ODE approximations. We ran five chains for each model, initialized at random parameter values, for 150,000 iterations per chain, the first 50,000 of which consisted of an adaptive tuning phase. The model parameters, not including the initial compartment volumes, were jointly updated via MVNSS (Web Appendix C). The empirical covariance for the MVNSS algorithm was adapted over the first 100,000 iterations using the gain factor sequence, $\gamma_n = 0.5(1 + 0.01n)^{-0.9}$. The contribution of isotropic Gaussian noise to the proposal was initialized at 0.001 and reduced throughout the adaptation phase according to the sequence $\iota_n = 0.001(1 + 0.01n)^{-0.99}$. The covariance matrix was blocked by country, treating parameters belonging to different countries as independent in the MCMC proposal kernel. Migration parameters were blocked with parameters corresponding to the destination country. The initial compartment volumes were jointly updated in a separate EllipSS update than the rest of the latent

Table S15: Rough estimates of effective population sizes under different reproduction numbers and detection rates that are needed to match the observed case counts from the West Africa Ebola outbreak to expected counts of detected cases under deterministic SEIR dynamics.

	Detection rate (ρ)		
$\mathbf{R_0}$	0.4	0.6	0.8
	Guinea		
1.25	14400	9560	7200
1.5	21700	14500	10900
1.75	31500	21000	15800
	Liberia		
1.25	19800	13200	9940
1.5	29900	19900	15000
1.75	43500	29000	21700
	Sierra Leone		
1.25	45000	30000	22500
1.5	67800	45200	33900
1.75	98500	65700	49200

Table S16: Rough estimates of effective population sizes under different reproduction numbers and detection rates that are needed to match the observed case counts from a simulated Ebola outbreak to expected counts of detected cases under deterministic SEIR dynamics.

	Detection rate (ρ)		
$\mathbf{R_0}$	0.4	0.6	0.8
	Guinea		
1.25	17600	11700	8850
1.5	26500	17700	13200
1.75	38500	25700	19200
	Liberia		
1.25	37000	24700	18500
1.5	55800	37200	27900
1.75	81000	54000	40500
	Sierra Leone		
1.25	40500	27000	20200
1.5	60900	40600	30500
1.75	88500	59000	44200

Table S17: Priors for the initial compartment volumes at the times when transmission was assumed to commence in Guinea, Liberia, and Sierra Leone. The initial compartment volumes for each country are assigned independent truncated multivariate normal priors (Web Appendix B). If the population size for country A is N_A, and the initial state probability is denoted $\mathbf{p}_{0,A} = \mathbf{X}_{0,A}/N_A = (S_{0,A}, E_{0,A}, I_{0,A}, R_{0,A})/N_A$, the prior is a truncated multivariate normal approximation of a multinomial distribution with mean $N_A\mathbf{p}_{0,A}$ and covariance $N_A(\mathbf{P}_{0,A} - \mathbf{p}_{0,A}\mathbf{p}_{0,A}^T)$.

Country	**Prior mean initial volumes** $(\mathbf{X}_0)$
Guinea	$(11.8 \times 10^6 - 30,\ 15,\ 10,\ 5)$
Liberia	$(4.4 \times 10^6 - 30,\ 15,\ 10,\ 5)$
Sierra Leone	$(7.1 \times 10^6 - 30,\ 15,\ 10,\ 5)$

LNA paths. The MCMC alternated between one ElliptSS and one MVNSS updates per MCMC iteration. The ElliptSS bracket width was reset after the first 5,000 MCMC iterations to $\omega = 2\sqrt{2\log(10)}\sigma_{ElliptSS}$, where $\sigma_{ElliptSS}$ was the standard deviation of the accepted angles over the initial iterations. The MCMC estimation scales for each country were parameterized as in Table S3 with the one addition being the log ratio of adjusted reproductive numbers, subtract one, for Guinea, $\log\left((R^{(2)}_{eff,G} - 1)/(R^{(1)}_{eff,G} - 1)\right)$. Convergence was assessed visually by inspection of traceplots of posterior samples, and via potential scale reduction factors (PSRFs) (Brooks and Gelman, 1998) computed via the `coda` R package (Plummer et al., 2006). The LNA models each took roughly three days to run while run times for ODE models took between 1–2 hours.

Ebola Model MCMC Details for Computation with Particle Marginal Metropolis–Hastings

Computation using the MMTL approximation within PMMH was done using the `pomp` R package King et al. (2016b). We used 1,000 particles in the PMMH algorithm. 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 and parameterized in the same way as LNA and ODE models. Parameter updates were made using a multivariate Metropolis algorithm with Gaussian proposals. The proposal covariance matrix was adapted over an initial run of 50,000 iterations, with adaptation starting after the first 100 iterations were completed. The gain factor sequence implemented in the package is of the form $\gamma_n = n^\alpha$, where the cooling term, α, was set to 0.99975. MCMC was run for 300,000 iterations following the adaptive MCMC phase. The posterior samples from all five MCMC chains were combined after discarding the initial samples from the adaptation phase.

We were unable to obtain a convergent and well mixing MCMC sampler, despite substantial expenditure of time and effort, even for the dataset simulated under the assumed model. Hence, PMMH was abandoned as a computational strategy in the analysis of data from the West Africa Ebola outbreak. The multivariate PSRF, computed using the `coda` package in `R` (Plummer et al., 2006), was 1.89 and four of the 24 parameters had univariate PSRF values above 1.05, indicating that the posterior had not been adequately explored.

Single–Country Models Fit to Ebola Data From Guinea, Liberia, and Sierra Leone

A typical first step in learning about the overall outbreak dynamics is to separately model the incidence data from each country. This is less challenging than fitting a multi–country model that incorporates cross–border transmission since each model will have fewer parameters and a smaller latent state space. Furthermore iterating through simplified models is helpful in identifying parameterizations that simplify the posterior geometry (see Web Appendix B).

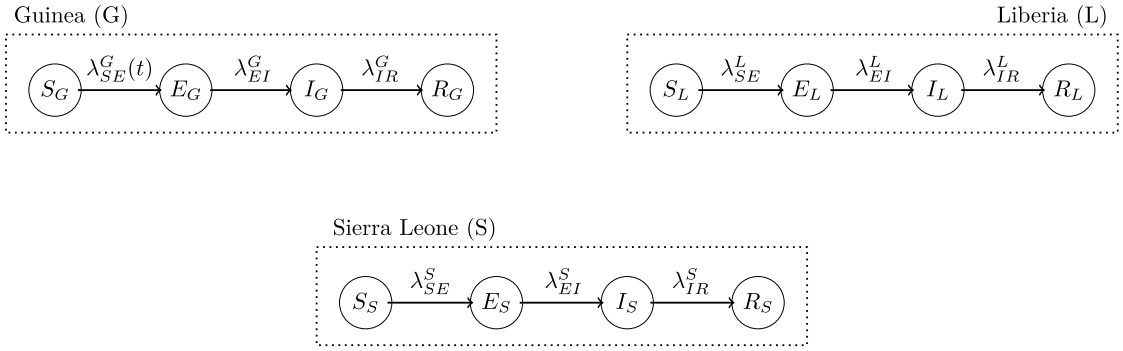

Figure S6: Diagram of state transitions for SEIR models fit to Ebola incidence data from Guinea, Liberia, and Sierra Leone. Dotted boxes denote countries, nodes in circles denote the model compartments: susceptible (S), exposed (E), infectious (I), recovered (R). Compartments are subscripted with country indicators. The number of susceptible individuals is equal to the effective population size, estimated as a parameter in the model, minus the numbers of exposed, infected, and recovered individuals. Solid lines with arrows indicate stochastic transitions between model compartments, which occur continuously in time. Rates at which individuals transition between compartments are denoted by λ and are subscripted by compartments and superscripted by countries, e.g., λ^L_{SE} is the rate at which susceptible individuals become exposed in Liberia. Transmission in Liberia and Sierra Leone was assumed to commence at 10 and 19 weeks, respectively. Full expressions for the rates are given in Table S18.

We fit separate SEIR models, diagrammed in Figure S6, to the incidence data from each country using both the LNA and ODE approximations. Transmission was modeled in Liberia beginning March 2, 2014, and in Sierra Leone from May 4, 2014, corresponding to three weeks prior to the first confirmed or probable cases in those countries. The force of infection in each country included a constant term for infectious contact from outside the population, but did not explicitly link exogenous transmission to the prevalence in other countries. The transition rates for the single–country models are given in Table S18. To account for the small scale of each outbreak relative to the population size in the country, we estimated the effective population size as a parameter in the model. The number of susceptible individuals was then equal to the effective population size, minus the numbers of exposed, infected, and recovered individuals. To complete the model specification, the observed incidence was modeled as a negative binomial sample of the true incidence in each inter–observation interval, as in (1).

Table S18: Rates of state transition for single–country models for Ebola transmission. Subscripts for rates indicate model compartments. All rates of state transition for Liberia and Sierra Leone are zero until three weeks prior to the first detected case, when transmission was assumed to commence in each country.

Rate	State Transition
$\lambda_{S^E E}(t) = (\alpha(t) + \beta(t)I)\, S^E$	$S^E \to E$
$\lambda_{EI}(t) = \omega(t)E$	$E \to I$
$\lambda_{IR}(t) = \mu(t)I$	$I \to R$

The priors for single–country model parameters were the same as those used in fitting the multi–country model, given in Tables S13 and S17. The exception was the prior for the baseline rate of transmission from outside the population, where for each country we took $1000\alpha \sim$Exponential(rate=40), reflecting our assumption that there were probably fewer than a dozen (two dozen under the more diffuse prior regime) transmission events per 1,000 infected individuals outside the country. This was based in part on results published in Dudas et al. (2017), who estimated between one–half to two dozen reintroduction events, depending on the country, between April 2014 and May 2015. We also used the same parameterizations for the estimation scales on which our MCMC samplers explored the posteriors (Table S14), again with the exception of the baseline rate of exogenous transmission. For each single–country model, we parameterized the estimation scale for this parameter as $\log(1000\alpha)$. All other MCMC details — number of iterations, adaptation tuning parameters, etc. — were the same as those used in fitting the multi–country Ebola model. Posterior medians and credible intervals for all parameters are reported in Table S19.

Additional Results and Diagnostics for the Model Fit via the Linear Noise Approximation — Simulated Data Example

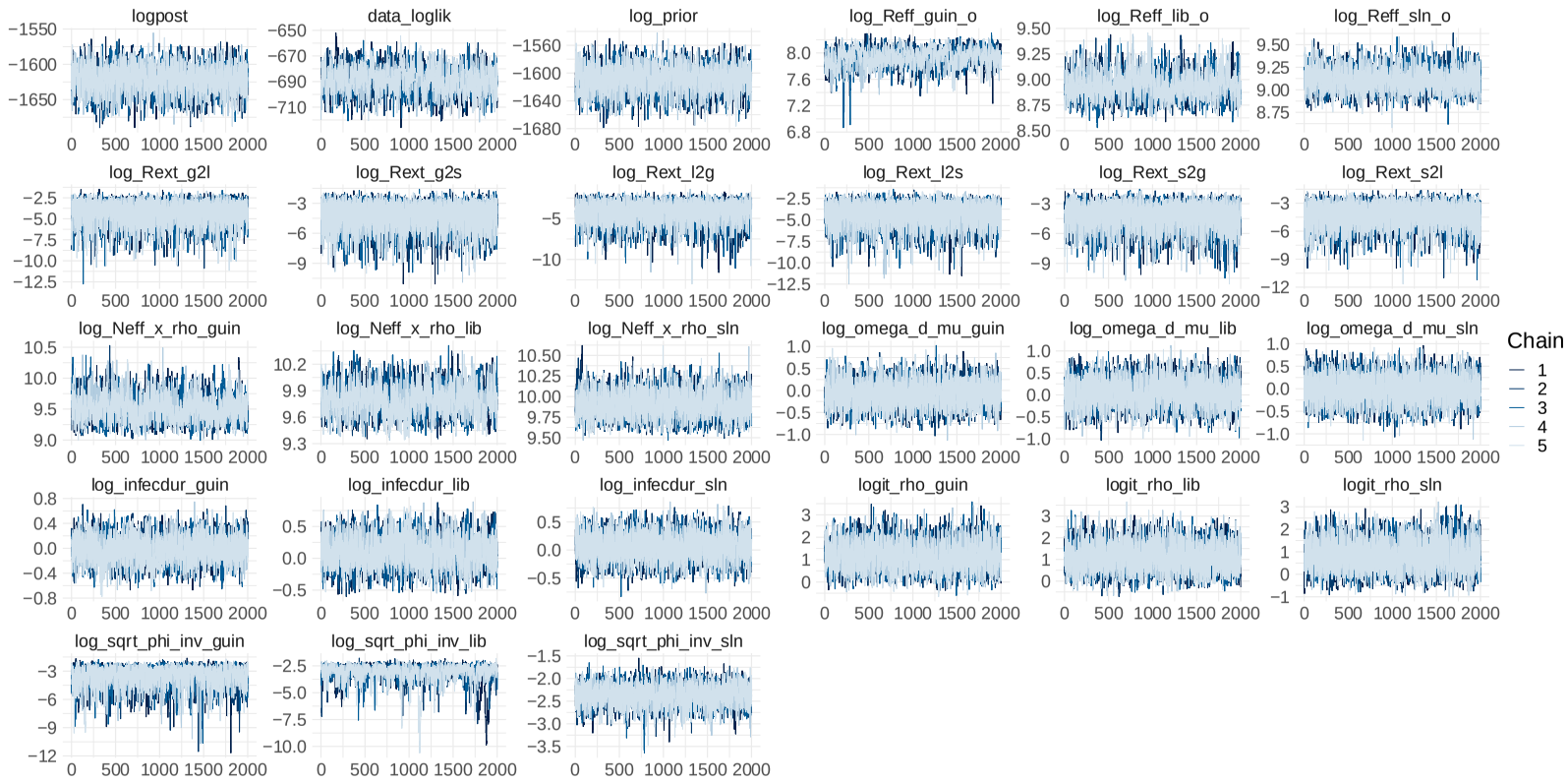

Figure S7: Traceplots for an Ebola LNA model to data from a simulated outbreak in West Africa. Traceplots were thinned to show every 25th MCMC iteration.

Additional Results and Diagnostics for the Model Fit via the Ordinary Differential Equation Approximation — Simulated Data Example

Additional Results for LNA and ODE Models Fit to Data from the West Africa Outbreak

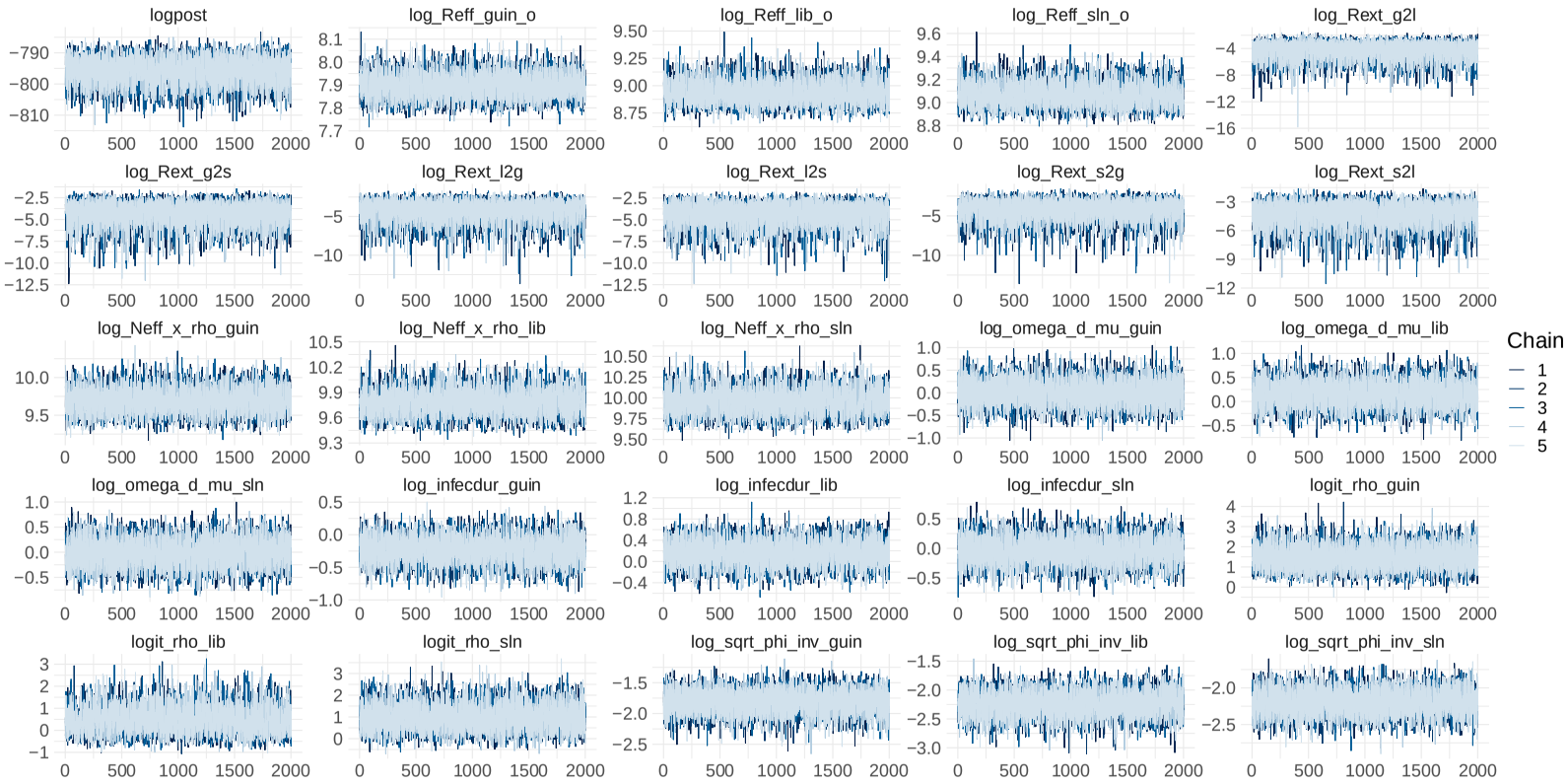

Figure S8: Traceplots for an Ebola ODE model to data from a simulated outbreak in West Africa. Traceplots were thinned to show every 25th MCMC iteration.

Table S19: Posterior medians (95% Bayesian credible intervals) of parameters for Ebola models fit to incidence data from the 2013–2015 outbreak in Guinea, Liberia, and Sierra Leone. Country–specific parameters are denoted by subscripts G, L, and S. Priors are given in Table S13.

		Posterior median (95% BCI)			
		Multi-country models		*Country-specific models*	
Parameter	**Interpretation**	LNA	ODE	LNA	ODE
$R^G_{adj} = \frac{\beta_G P_{eff,G}}{\mu_G}$	Adjusted basic reproduction number	1.2 (1.1, 1.5)	1.4 (1.2, 1.7)	1.3 (1.1, 1.5)	1.4 (1.2, 1.7)
$7/\omega_G$	Latent period, days	6.4 (3, 14.5)	12.2 (6.4, 22.2)	7.7 (3.6, 16.1)	12.8 (6.8, 23.4)
$7/\mu_G$	Infectious period, days	6.4 (3.8, 11.6)	9.9 (6.2, 16.1)	7.3 (4.3, 12.7)	10.2 (6.4, 16.3)
$1000\alpha_G$	Baseline rate of exogenous transmission × 1,000	———	———	0 (0, 0.1)	0 (0, 0.1)
$R^{LG}_{ext} = \frac{\alpha_{LG} P_{eff,G}}{\mu_L}$	Adjusted extrinsic reproduction number	0 (0, 0.1)	0 (0, 0.1)	———	———
$R^{SG}_{ext} = \frac{\alpha_{SG} P_{eff,G}}{\mu_S}$	Adjusted extrinsic reproduction number	0 (0, 0.1)	0 (0, 0.1)	———	———
$P_{eff,G} \times \rho_G$	Scale of detected outbreak	10700 (6500, 21300)	7700 (5400, 11400)	9600 (6100, 17400)	7500 (5400, 11100)
$1/\sqrt{\phi_G}$	Negative binomial overdispersion	0.4 (0.3, 0.6)	0.6 (0.5, 0.7)	0.5 (0.4, 0.6)	0.6 (0.5, 0.7)
$R^L_{adj} = \frac{\beta_L P_{eff,L}}{\mu_L}$	Adjusted basic reproduction number	1.9 (1.4, 3.2)	2.2 (1.6, 3.5)	2 (1.5, 3.1)	2.1 (1.6, 3.3)
$7/\omega_L$	Latent period, days	7.8 (3.4, 18.6)	11.2 (4.7, 20.1)	9.2 (4.4, 18.4)	10.6 (4.9, 19.2)
$7/\mu_L$	Infectious period, days	7.5 (4.2, 13.7)	9.2 (5.3, 15.3)	8.5 (5.1, 14.4)	9.1 (5.3, 15.2)
$1000\alpha_L$	Baseline rate of exogenous transmission × 1,000	———	———	0 (0, 0)	0 (0, 0)
$R^{GL}_{ext} = \frac{\alpha_{GL} P_{eff,L}}{\mu_G}$	Adjusted extrinsic reproduction number	0 (0, 0.1)	0 (0, 0.1)	———	———
$R^{SL}_{ext} = \frac{\alpha_{SL} P_{eff,L}}{\mu_S}$	Adjusted extrinsic reproduction number	0 (0, 0.1)	0 (0, 0.1)	———	———
$P_{eff,L} \times \rho_L$	Scale of detected outbreak	6300 (4800, 9300)	5600 (4600, 7900)	5900 (4600, 8200)	5700 (4400, 7800)
$1/\sqrt{\phi_L}$	Negative binomial overdispersion	0.3 (0.2, 0.5)	0.4 (0.3, 0.6)	0.4 (0.3, 0.6)	0.4 (0.3, 0.6)
$R^S_{adj} = \frac{\beta_S P_{eff,S}}{\mu_S}$	Adjusted basic reproduction number	1.3 (1.2, 1.4)	1.4 (1.2, 1.7)	1.3 (1.2, 1.5)	1.5 (1.3, 1.9)
$7/\omega_S$	Latent period, days	4.2 (2.4, 7.1)	5.8 (3.2, 9.8)	5 (2.7, 8.7)	7.4 (4.1, 12.8)
$7/\mu_S$	Infectious period, days	4.7 (3.1, 7)	5.9 (3.8, 8.9)	5.3 (3.5, 8.2)	7.1 (4.5, 10.9)
$1000\alpha_S$	Baseline rate of exogenous transmission × 1,000	———	———	0 (0, 0)	0 (0, 0.1)
$R^{GS}_{ext} = \frac{\alpha_{GS} P_{eff,S}}{\mu_G}$	Adjusted extrinsic reproduction number	0 (0, 0.1)	0 (0, 0.1)	———	———
$R^{LS}_{ext} = \frac{\alpha_{LS} P_{eff,S}}{\mu_L}$	Adjusted extrinsic reproduction number	0 (0, 0.1)	0 (0, 0.1)	———	———
$P_{eff,S} \times \rho_S$	Scale of detected outbreak	27800 (20600, 40500)	20300 (17500, 32900)	25200 (18500, 37200)	20300 (15300, 27900)
$1/\sqrt{\phi_S}$	Negative binomial overdispersion	0.1 (0.1, 0.2)	0.4 (0.3, 0.5)	0.3 (0.2, 0.4)	0.5 (0.4, 0.6)

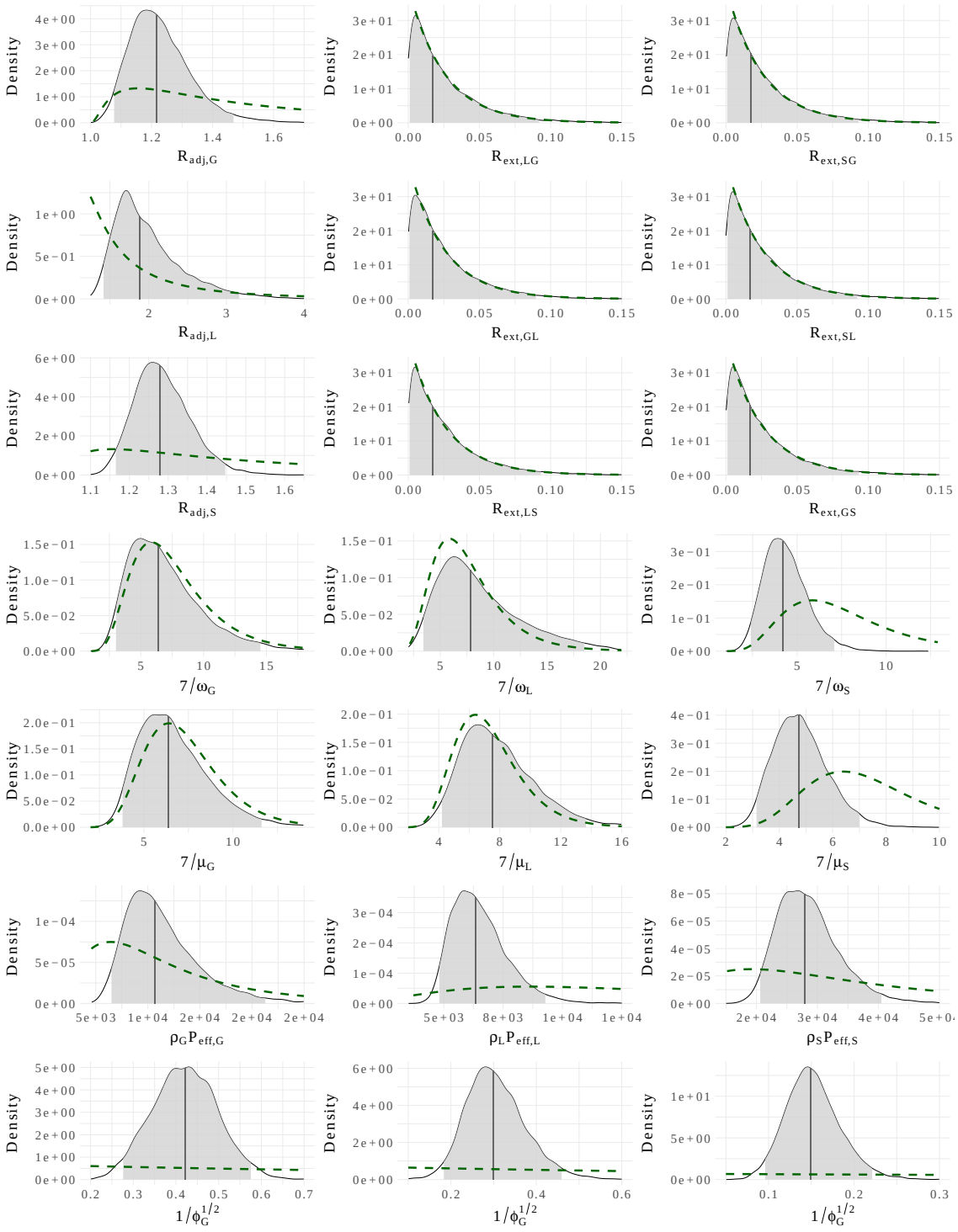

Figure S9: Posterior distributions of LNA model parameters for the model fit to data from the West Africa Ebola outbreak. We show posterior medians (solid gray lines), 95% Bayesian credible intervals (light gray areas under the posterior densities), prior densities (induced priors for the reporting rate and latent period durations) over the posterior ranges (dashed green curves). Priors and interpretations of parameters are specified in Table S13.

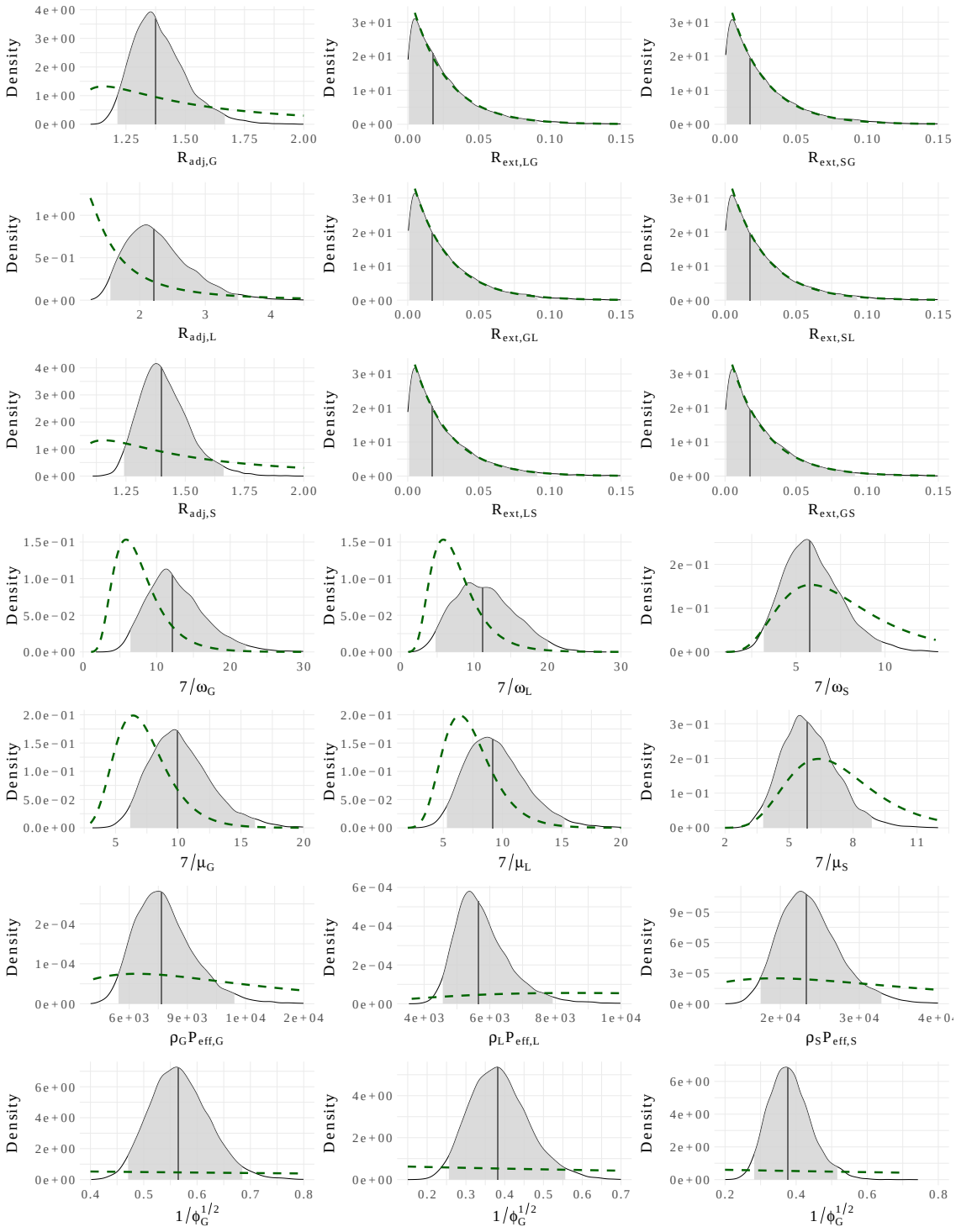

Figure S10: Posterior distributions of ODE model parameters for the model fit to data from the West Africa Ebola outbreak under tight priors. We show posterior medians (solid gray lines), 95% Bayesian credible intervals (light gray areas under the posterior densities), prior densities (induced priors for the reporting rate and latent period durations) over the posterior ranges (dashed green curves). Priors and interpretations of parameters are specified in Table S13.

Table S20: Posterior estimates of initial numbers of exposed and infected individuals for the Ebola models fit to data from the West Africa outbreak. The effective number of susceptibles is equal to the effective population size minus the numbers of exposed, infected, and recovered individuals, but is not reported along with the number of recovered individuals since the effective population size is only weakly identified (Web Appendix D). Country–specific parameters are denoted by subscripts G, L, and S. Priors are given in Table S17.

Parameter	LNA	ODE
$E_{0,G}$	8.9 (2.1, 17.6)	12.4 (4.5, 20.2)
$E_{0,L}$	9 (2, 16.9)	0.6 (0, 3.3)
$E_{0,S}$	6.4 (0.8, 14.1)	14.6 (7.1, 21.5)
$I_{0,G}$	6.9 (1.4, 13.4)	7.9 (2, 14.4)
$I_{0,L}$	6 (1, 12)	0.3 (0, 1.6)
$I_{0,S}$	7.3 (1.7, 13.2)	11.1 (5.5, 16.7)

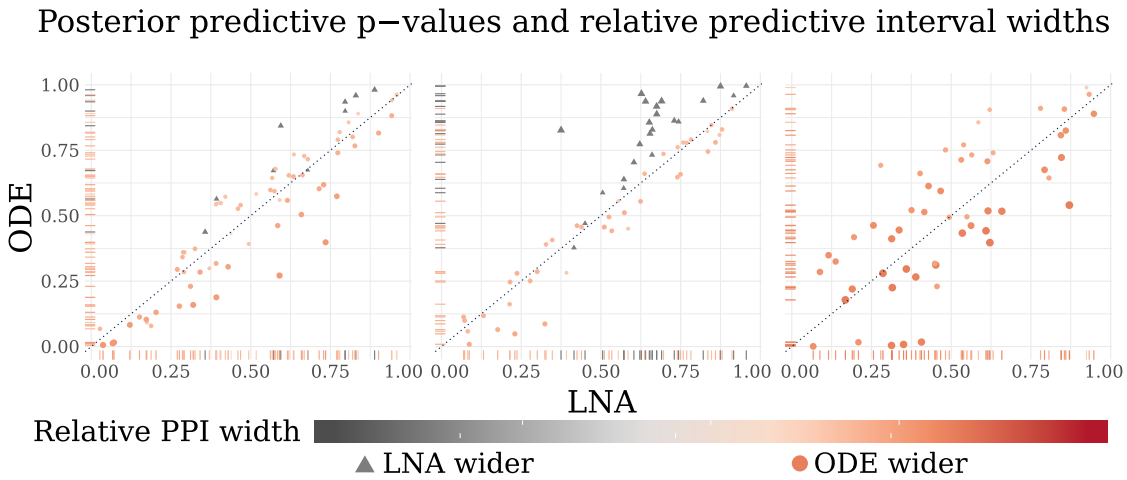

Figure S11: (Bottom panel) Comparison of posterior predictive p-values (PPPs) and relative posterior predictive interval (PPI) widths for LNA and ODE models fit to data from the West Africa Ebola outbreak. Each point corresponds to the observed incidence in a given week. The X–Y coordinates give the PPPs under each model. The size and color of each point corresponds to the relative PPI width, computed as $(\widehat{\sigma}^{ODE}_{post,\ell} - \widehat{\sigma}^{LNA}_{post,\ell})/\widehat{\sigma}^{LNA}_{post,\ell}$, and the sign of the relative width is further emphasized by the shape of the point. Dots indicate that PPIs for the ODE model are wider, while triangles corresponds to observations for which the PPI produced by the LNA model was wider.

Table S21: Effective sample sizes and potential scale reduction factors for LNA and ODE Ebola multi–country model parameters.

	LNA		**ODE**	
Parameter	ESS	PSRF	ESS	PSRF
$\log(\beta_G P_{eff,G}/\mu_G - 1) + \log(P_{eff,G}\rho_S)$	1013	1.01	6449	1.00
$\log(\beta_L P_{eff,L}/\mu_L - 1) + \log(P_{eff,L}\rho_L)$	559	1.02	6505	1.00
$\log(\beta_S P_{eff,S}/\mu_A - 1) + \log(P_{eff,S}\rho_S)$	756	1.01	5604	1.00
$\log(\alpha_{GL} P_{eff,L}/\mu_G)$	2866	1.00	8626	1.00
$\log(\alpha_{GS} P_{eff,S}/\mu_G)$	1406	1.01	5012	1.00
$\log(\alpha_{LG} P_{eff,G}/\mu_L)$	3335	1.00	7964	1.00
$\log(\alpha_{LS} P_{eff,S}/\mu_L)$	863	1.01	4172	1.00
$\log(\alpha_{SG} P_{eff,G}/\mu_S)$	683	1.01	6304	1.00
$\log(\alpha_{SL} P_{eff,L}/\mu_S)$	1396	1.01	3475	1.00
$\log(P_{eff,G}\rho_G)$	6280	1.00	6581	1.00
$\log(P_{eff,L}\rho_L)$	6166	1.00	9172	1.00
$\log(P_{eff,S}\rho_S)$	7544	1.00	9312	1.00
$\log(\omega_G/\mu_G)$	6002	1.00	8624	1.00
$\log(\omega_L/\mu_L)$	7212	1.00	9266	1.00
$\log(\omega_S/\mu_S)$	7087	1.00	5961	1.00
$\log(7/\mu_G)$	884	1.01	3818	1.00
$\log(7/\mu_L)$	544	1.01	5112	1.00
$\log(7/\mu_S)$	840	1.01	2736	1.00
$\log(\rho_G/(1-\rho_G))$	1431	1.01	1696	1.00
$\log(\rho_L/(1-\rho_L))$	783	1.01	2416	1.00
$\log(\rho_S/(1-\rho_S))$	3503	1.00	4041	1.00
$1/\sqrt{\phi_G}$	607	1.01	9627	1.00
$1/\sqrt{\phi_L}$	794	1.03	7436	1.00
$1/\sqrt{\phi_S}$	980	1.01	9384	1.00