# Supplementary Material for "Efficient Data Augmentation for Fitting Stochastic Epidemic Models to Prevalence Data" by

Jonathan Fintzi[1], Xiang Cui[2], Jon Wakefield[1,2], and Vladimir N. Minin[2,3]

[1]Department of Biostatistics, University of Washington, Seattle
[2]Department of Statistics, University of Washington, Seattle
[3]Department of Biology, University of Washington, Seattle

## S1 SIR Model Construction and Lumpability of CTMCs

In this section, we outline why the SIR model of Section **??** is equivalent to the canonical SIR model (Kermack and McKendrick, 1927, Andersson and Britton, 2000) via a property called *lumpability*. The following discussion is not meant to be a comprehensive presentation of the theoretical details behind the connection between the two models. We refer readers seeking a more thorough presentation to Tian and Kannan (2006).

Given a Markov process, $\mathbf{X}$ with state space $\mathcal{S} = \{s_1, \ldots, s_P\}$ and initial probability vector $\pi$, we define a new process, $\overline{\mathbf{X}}$ on state space $\overline{\mathcal{S}} = \{S_1, \ldots, S_{\mathcal{L}}\}$, a partition of $\mathcal{S}$. The jump chain of this new chain is obtained by taking the sequence of subsets of $\overline{\mathcal{S}}$ that contain the corresponding states of the original jump chain. The initial probability distribution of $\overline{\mathbf{X}}(t)$ is

$$\Pr(\overline{\mathbf{X}}(t_0) = S_i) = \Pr_\pi(\mathbf{X}(t_0) \in S_i)$$

and its transition probabilities are given by

$$\Pr(\overline{\mathbf{X}}(t + \Delta t) = S_j | \overline{\mathbf{X}}(t) = \overline{\mathbf{x}}(t'), t' \leqslant t) = \Pr(\overline{\mathbf{X}}(t + \Delta t) \in S_j | \mathbf{X}(t) = \mathbf{x}(t'), t' \leqslant t),$$

where $\overline{\mathbf{x}}(t')$ and $\mathbf{x}(t')$ denote the paths of the original process and the new process. The new process is called the *lumped process*. We say that the original process is *lumpable* with respect to a partition $\overline{\mathcal{S}}$ of $\mathcal{S}$, and that $\overline{\mathbf{X}}(t)$ is the *lumped Markov process* corresponding to $\mathbf{X}(t)$, if for every choice of $\pi$ we have that $\overline{\mathbf{X}}(t)$ is Markov and the transition probabilities do not depend on $\pi$. A necessary and sufficient condition for a CTMC to be lumpable is that its rate matrix, $\mathbf{\Lambda} = (\lambda_{a,b})$, where $\lambda_{a,b}$

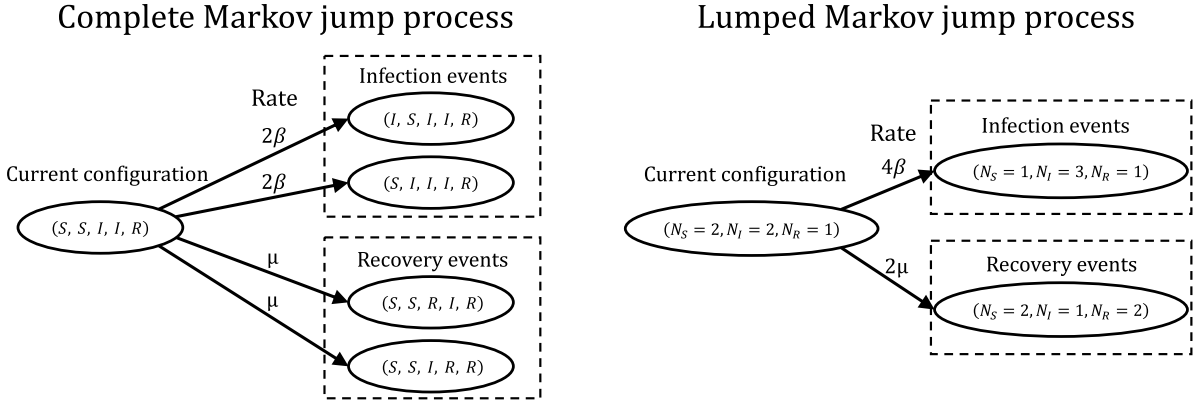


**Figure S1:** Complete and lumped representations of SIR dynamics in a population of five individuals. The per–contact infectivity rate, $\beta$, and the recovery rate, $\mu$, parameterize exponential waiting time distributions between transition events. The complete Markov jump process evolves on the state space of subject state labels, $\mathcal{S} = \{S, I, R\}^N$, with dynamics determined by the subject–level transition rates. Each susceptible may contact two infected individuals, while each infected individual recovers independently. The lumped process evolves on the state space of compartment counts, $\overline{\mathcal{S}} = \{N_S, N_I, N_R : \ N_S + N_I + N_R = N\}$, with dynamics determined by lumped transition rates. The waiting time distributions between transitions are derived by noting that if $\tau_1 \sim \exp(\lambda_1)$ and $\tau_2 \sim \exp(\lambda_2)$, then $\tau_{\min} = \min(\tau_1, \tau_2) \sim \exp(\lambda_1 + \lambda_2)$.

being the rate of transition from $s_a$ to $s_b$, satisfies

$$\sum_{s_b \in S_B} \lambda_{a,b} = \sum_{s_b \in S_B} \lambda_{c,b}$$

for any pair of sets $S_A$ and $S_B$ and for any pair of states $(s_a, s_c)$ in $S_A \in \overline{\mathcal{S}}$.

In Section **??**, we defined the latent process, $\mathbf{X}(\tau) = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$, with state space $\mathcal{S} = \{S, I, R\}^N$. Let $c_u = (x_1, \ldots, x_N)$ denote a configuration of the state labels (e.g. $c_u = (S, I, S, R, I)$), and denote the set of configurations that correspond to a vector of compartment counts by

$$\mathcal{C}_{lmn} = \left\{ c_u : l = \sum_{i=1}^{N} \mathbb{I}\left(x_i = S\right), m = \sum_{i=1}^{N} \mathbb{I}\left(x_i = I\right), n = \sum_{i=1}^{N} \mathbb{I}\left(x_i = R\right), \ l + m + n = N \right\}.$$

The state space of count vectors,

$$\overline{\mathcal{S}} = \left\{ \mathcal{C}_{lmn} : l, m, n \in \{1, \ldots, N\}, l + m + n = N \right\},$$

defines a partition of $\mathcal{S}$ that is obtained by stripping away the subject labels and summing the number of individuals in each disease state.

Given the partition $\overline{\mathcal{S}}$ of $\mathcal{S}$, we may define the CTMC for the canonical SIR model, $\overline{\mathbf{X}} = (N_S, N_I, N_R)$, where $N_S + N_I + N_R = N$ and $N_S, N_I, N_R \in \{0, \ldots, N\}$, on the state space of compartment counts, depicted in Figure S1. This construction is usually presented for computational reasons since dis-

carding the subject labels for infection and recovery events substantially reduces the computational overhead. When the sojourn times are exponentially distributed, the transition rates for the time-homogeneous CTMC are

$$\begin{array}{cc} \underline{\text{Transition}} & \underline{\text{Rate}} \\ (N_S, N_I, N_R) \longrightarrow (N_S - 1, N_I + 1, N_R) & \beta N_S N_I, \\ (N_S, N_I, N_R) \longrightarrow (N_S, N_I - 1, N_R + 1) & \mu N_I. \end{array}$$

The state space $\overline{\mathcal{S}}$ partitions the state space $\mathcal{S}$ into groups of configurations for which the triple of compartment counts are the same. The CTMC $\overline{\mathbf{X}}$ trivially satisfies the condition for lumpability, and thus is the lumped Markov chain of $\mathbf{X}$ with respect to this partition.

# S2 Computing the matrix exponential

The transition probability matrix (TPM), $\mathbf{P}(t)$, for a time–homogeneous CTMC over an interval of length $t$, solves the matrix differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t) = \mathbf{\Lambda}\mathbf{P}(t), \quad \text{s.t. } \mathbf{P}(0) = \mathbf{I},$$

where $\mathbf{\Lambda}$ is the transition rate matrix for the CTMC and $\mathbf{I}$ is an identity matrix of the same size as $\mathbf{\Lambda}$ (Wilkinson, 2011). Therefore, $\mathbf{P}$ is computed using the matrix exponential solution of the above differential equation, $\mathbf{P} = \exp(t\mathbf{\Lambda})$. This is the most intensive step in our algorithm. However, we may lessen the computational burden to a large extent by leveraging the fact that we are computing the matrix exponential for the same rate matrix for possibly many values of $t$. Therefore, computing the matrix exponential using the eigen decomposition of $\mathbf{\Lambda}$ and caching the resulting eigenvalues and eigenvectors will be relatively efficient (Moler and Van Loan, 2003). We outline this computation in the following two cases: when the eigenvalues of $\mathbf{\Lambda}$ are all real (e.g. as with the SIR and SEIR models), and when $\mathbf{\Lambda}$ has complex eigenvalues (e.g. as is possibly the case with the SIRS model).

## S2.1 Case 1: $\mathbf{\Lambda}$ has real eigenvalues

Suppose that $\mathbf{\Lambda}_{n\times n} = \mathbf{U}\mathbf{V}\mathbf{U}^{-1}$, where $\mathbf{V}$ is a diagonal matrix of eigenvalues, $v_1, \ldots, v_n$, and $\mathbf{U}$ is the matrix whose columns are the corresponding right eigenvectors. Then,

$$\mathrm{e}^{t\mathbf{V}} = \mathrm{diag}(\mathrm{e}^{v_1 t}, \ldots, \mathrm{e}^{v_n t}).$$

That $\mathbf{U}$ is nonsingular yields

$$\mathrm{e}^{t\mathbf{\Lambda}} = \mathbf{U}\mathrm{e}^{t\mathbf{V}}\mathbf{U}^{-1}.$$

## S2.2 Case 2: $\mathbf{\Lambda}$ has complex eigenvalues

In the event that $\mathbf{\Lambda}$ has complex eigenvalues, we may obtain a real-valued TPM by transforming $\mathbf{\Lambda}$ into its real canonical form (Hirsch et al., 2013). Suppose that $\mathbf{\Lambda}$ has $r$ real eigenvalues, $v_1, \ldots, v_r$, with corresponding real eigenvectors, $\mathbf{u}_1, \ldots, \mathbf{u}_r$, and $n - r$ pairs of complex conjugate eigenvalues. Let $(\mathbf{u}_j|\mathbf{w}_j)$ denote the real and imaginary parts of the eigenvector corresponding to the $j^{th}$ eigenvalue, $\alpha_j + i\beta_j$, for $j = r+1, \ldots, n$, and define the matrix $\mathbf{T} = (\mathbf{u}_1|\ldots|\mathbf{u}_r|\mathbf{u}_{r+1}|\mathbf{w}_{r+1}|\ldots|\mathbf{u}_n|\mathbf{w}_n)$. The real canonical form for a rate matrix with complex eigenvalues can now be written as $\mathbf{V} = \mathbf{T}^{-1}\mathbf{\Lambda}\mathbf{T}$, where $\mathbf{V} = \mathrm{diag}(v_1, \ldots, v_r, \mathbf{B}_{r+1}, \ldots, \mathbf{B}_n)$, and each $\mathbf{B}_j,\ j = r+1, \ldots, n$ is given by

$$\mathbf{B}_j = \begin{pmatrix} \alpha_j & \beta_j \\ -\beta_j & \alpha_j \end{pmatrix},$$

which implies that

$$\mathrm{e}^{t\mathbf{B}_j} = \mathrm{e}^{\alpha_j t} \begin{pmatrix} \cos(\beta_j t) & \sin(\beta_j t) \\ -\sin(\beta_j t) & \cos(\alpha_j) \end{pmatrix},$$

and hence $\mathrm{e}^{t\mathbf{V}} = \mathrm{diag}(\mathrm{e}^{v_1 t}, \ldots, \mathrm{e}^{v_r t}, \mathrm{e}^{t\mathbf{B}_{r+1}}, \ldots, \mathrm{e}^{t\mathbf{B}_n})$. Therefore, we can compute the matrix exponential of $t\mathbf{\Lambda}$ as $\mathrm{e}^{t\mathbf{\Lambda}} = \mathbf{T}\mathrm{e}^{t\mathbf{V}}\mathbf{T}^{-1}$.

# S3 Forward-Backward Algorithm for Sampling the Disease State at Observation Times

The stochastic forward-backward algorithm (Scott, 2002) enables us to efficiently sample from $\pi\left(\mathbf{X} \mid \mathbf{Y}, \mathbf{X}_{(-j)}, \boldsymbol{\theta}\right)$ by recursively accumulating, in a "forward" pass, information about the probability of various paths through $\mathcal{S}$, conditional on the data, and then recursively sampling a trajectory in a "backwards" pass. Let $\mathbf{Y}_{t_1}^{t_\ell} = (Y_1, \ldots, Y_\ell)$ denote the observations made at times $t_1, \ldots, t_\ell$, and similarly, let $\mathbf{X}_{j,t_{L-\ell+1}}^{t_L} = (\mathbf{X}_j(t_{L-\ell+1}), \ldots, \mathbf{X}_j(t_L))$ denote the state of $\mathbf{X}_j$ at times $t_{L-\ell+1}, \ldots, t_L$. In the forward recursion, we construct a sequence of matrices $\boldsymbol{Q}_j^{(t_2)}, \ldots, \boldsymbol{Q}_j^{(t_L)}$, where $\boldsymbol{Q}_j^{(t_\ell)} = \left(q_{j,r,s}^{(t_\ell)}\right)$, and $q_{j,r,s}^{(t_\ell)} = \Pr\left(\mathbf{X}_j(t_\ell) = s, \mathbf{X}_j(t_{\ell-1}) = r \mid \mathbf{Y}_{t_1}^{t_\ell}, \mathbf{X}_{(-j)}, \boldsymbol{\theta}\right)$. Let $\mathbf{P}_{r,s}^{(j)}(t_{\ell-1}, t_\ell) = \Pr\left(\mathbf{X}_j(t_\ell) = s \mid \mathbf{X}(t_{\ell-1}) = r, \boldsymbol{\theta}; \mathbf{X}_{(-j)}\right)$. If there are changes in the numbers of infected individuals in interval $\mathcal{I}_\ell$, we construct the transition probability matrix for that interval as in (**??**). Then,

$$q_{j,r,s}^{(t_\ell)} \propto \pi_j^{(t_\ell)}\left(r \mid \mathbf{X}_{(-j)}, \boldsymbol{\theta}\right) \times \mathbf{P}_{r,s}^{(j)}\left(t_{\ell-1}, t_\ell\right) \times f\left(Y_{t_\ell} \mid \mathbf{X}_j(t_\ell), \mathbf{X}_{(-j)}(t_\ell), \rho, \mathbf{p}_{t_1}\right), \tag{1}$$

where $\pi_j^{(t_\ell)}\left(r \mid \mathbf{X}_{(-j)}, \boldsymbol{\theta}, \rho\right) = \sum_r q_{j,r,s}^{(t_j)}$ and with proportionality reconciled via $\sum_r \sum_s q_{j,r,s}^{(t_j)} = 1$.

In the backwards pass, we sample the sequence of states at times $t_1, \ldots, t_L$ from the distribution $\pi\left(\mathbf{X} \mid \mathbf{Y}, \mathbf{X}_{(-j)}, \boldsymbol{\theta}, \rho, \mathbf{p}_{t_1}\right)$. To do this, we first note that

$$\begin{aligned}
\pi\left(\mathbf{X} \mid \mathbf{Y}, \mathbf{X}_{(-j)}, \boldsymbol{\theta}, \rho, \mathbf{p}_{t_1}\right) &= \pi\left(\mathbf{X}_j(t_L) \mid \mathbf{Y}_{t_1}^{t_L}, \mathbf{X}_{(-j)}, \boldsymbol{\theta}, \rho, \mathbf{p}_{t_1}\right) \prod_{\ell=1}^{L-1} \pi\left(\mathbf{X}_j(t_{L-\ell}) \mid \mathbf{X}_{j,t_{L-\ell+1}}^{t_L}, \mathbf{X}_{(-j)}, \mathbf{Y}_{t_1}^{t_L}, \boldsymbol{\theta}, \rho, \mathbf{p}_{t_1}\right) \\
&= \pi\left(\mathbf{X}_j(t_L) \mid \mathbf{Y}_{t_1}^{t_L}, \mathbf{X}_{(-j)}, \boldsymbol{\theta}, \rho, \mathbf{p}_{t_1}\right) \prod_{\ell=1}^{L-1} \pi\left(\mathbf{X}_j(t_{L-\ell}) \mid \mathbf{X}_{j,t_{L-\ell+1}}, \mathbf{X}_{(-j)}, \mathbf{Y}_{t_1}^{t_{L-\ell+1}}, \boldsymbol{\theta}, \rho, \mathbf{p}_{t_1}\right),
\end{aligned}$$

where the second equality follows from the conditional independence of the HMM. We proceed by first drawing $\mathbf{X}_j(t_L)$ from $\pi_j^{(t_L)}\left(\cdot \mid \mathbf{X}_{(-j)}, \boldsymbol{\theta}, \rho\right)$, and then drawing $\mathbf{X}_j(t_\ell)$, $\ell = L-1, \ldots, 1$, each in turn from the categorical distribution with masses proportional to column $\mathbf{x}_j(t_{\ell+1})$ of $\boldsymbol{Q}_j^{(t_{\ell+1})}$.

# S4 Simulating endpoint Conditioned Time–Homogeneous CTMC Paths

In this section, we briefly summarize the modified rejection sampling and uniformization algorithms for simulating a path from an endpoint-conditioned time-homogeneous CTMC. The following discussion is not meant to be comprehensive, and we refer the reader to the excellent paper by Hobolth and Stone (2009) for a more thorough discussion. We also refer the reader to the `ECctmc R` package for a fast implementation of these algorithms which we relied upon in implementing our data augmentation algorithm (Fintzi, 2016).

Our goal is to simulate a path for a time–homogeneous CTMC, $\mathbf{X}$, in the interval $[0, T]$, conditional on $\mathbf{X}(0) = a$ and $\mathbf{X}(T) = b$. Let $\mathbf{\Lambda}$ be the rate matrix for the process. Let $\Lambda_a$ denote the $a, a$ diagonal element of $\mathbf{\Lambda}$, and similarly let $\Lambda_{a,b}$ denote the rate given by the $a, b$ element. We also denote by $\mathbf{P}(T)$ the transition probability matrix for the CTMC over $[0, T]$, and $P_{ab}(T)$ the probability of beginning in state $a$ and ending in state $b$.

## S4.1 Modified rejection sampling

The modified rejection algorithm proposes paths by explicitly sampling the first transition time when it is known that at least one transition occurred (i.e. when $a \neq b$). The remainder of the path is proposed by forward sampling, for instance, via Gillespie's direct algorithm. The proposed path is then accepted if $\mathbf{X}(T) = b$. When it is not known whether a transition occurred (i.e. when $a = b$), a path is proposed via ordinary forward simulation and accepted if $\mathbf{X}(T) = b$.

We sample the first transition time via the inverse–CDF method, sampling $u \sim \text{Unif}(0, 1)$ and applying the inverse-CDF function

$$F^{-1}(u) = \frac{-log\left[1 - u \times \left(1 - \mathrm{e}^{-T\Lambda_a}\right)\right]}{\Lambda_a}. \tag{2}$$

We found that the modified rejection algorithm worked well in fitting the SIR and SIRS models. In the examples we studied in which these models were fit, subject–paths over intervals where the endpoints required multiple jumps ($S \to R$, or $I \to S$) were almost never considered. Therefore, usually only a single transition time was required to be sampled in a given interval, and accomplishing this using the inverse–CDF method was quite fast.

## S4.2 Uniformization

The uniformization algorithm samples the path for a time–homogeneous CTMC conditional on the state at the interval endpoints by coupling the original process to a Markov chain determined by an auxilliary Poisson point process. State transitions, including virtual transition where the state

does not change, occur at points of this auxilliary process, and the sequence of state labels is drawn from the corresponding Markov chain.

We construct the transition rate matrix of the auxilliary Markov chain, $\mathbf{Y}$, as $R = I + \frac{1}{\mu}\mathbf{\Lambda}$, where $\mu = \max_a \mathbf{\Lambda}_a$. Then number of state transitions, $N$, conditional on $\mathbf{X}(0) = a,\ \mathbf{X}(T) = b$, can be shown to be

$$P(N = n|\mathbf{X}(0) = a, \mathbf{X}(T) = b) = \mathrm{e}^{-\mu T}\frac{(\mu T)^n}{n!}R^n_{ab/}P_{ab}(T). \tag{3}$$

The algorithm proceeds by first sampling the number of state transitions from this distribution. If there are no transitions, or if there is one transition and the states at the endpoints are the same, the algorithm terminates. Otherwise, we drawn $n$ independent uniform values in $[0, T]$ and sort them to obtain the times of state transitions. The state labels at the sorted sequence of times, $\tau_i$, $i = 1, \ldots, n-1$, is then drawn from the discrete distribution with masses given by

$$P(\mathbf{X}(\tau_i)|\mathbf{X}(\tau_{i-1}, \mathbf{X}(T) = b) = \frac{R_{x_{i-1},x_i}(R^{n-i})_{x_i b}}{(R^{n-i+1})_{x_{i-1}b})}. \tag{4}$$

We found that uniformization was preferable to modified rejection sampling when fitting the SEIR model. In this case, modified rejection sampling tended to get hung up when sampling paths in intervals where the endpoints suggested that at least two state transitions occured (which though it seldom occured, significantly slowed down the MCMC). We also note that the transition probability, $P_{ab}(T)$, is computed and cached in carrying out the HMM step of our algorithm. Therefore, there are no additional eigen–decompositions or matrix exponentiations required in using the uniformization algorithm to sample the exact times of state transition.

# S5 Metropolis-Hastings Ratio Details

Our target distribution is $\pi(\mathbf{X}|\mathbf{Y}) \propto \pi(\mathbf{Y}|\mathbf{X})\pi(\mathbf{X})$. Note that $\mathbf{x}^{\text{new}}$ and $\mathbf{x}^{\text{cur}}$ differ only in the path of the $j^{th}$ subject, so $\Lambda^{(-j)}(\mathbf{x}^{\text{cur}}) = \Lambda^{(-j)}(\mathbf{x}^{\text{new}}) = \mathbf{\Lambda}^{(-j)}$. Suppressing the dependence on $\boldsymbol{\theta}$ for clarity, the acceptance ratio is

$$a_{\mathbf{x}^{\text{cur}} \longrightarrow \mathbf{x}^{\text{new}}} = \min\left\{ \frac{\pi(\mathbf{x}^{\text{new}}|\mathbf{y})}{\pi(\mathbf{x}^{\text{cur}}|\mathbf{y})} \frac{q(\mathbf{x}^{\text{cur}}|\mathbf{x}^{\text{new}})}{q(\mathbf{x}^{\text{new}}|\mathbf{x}^{\text{cur}})},\ 1 \right\}$$

Now,

$$\begin{aligned} \pi(\mathbf{x}^{\text{new}}|\mathbf{y}) &\propto \Pr(\mathbf{y}|\mathbf{x}^{\text{new}})\pi(\mathbf{x}^{\text{new}}), \\ \pi(\mathbf{x}^{\text{cur}}|\mathbf{y}) &\propto \Pr(\mathbf{y}|\mathbf{x}^{\text{cur}})\pi(\mathbf{x}^{\text{cur}}), \end{aligned}$$

where $\Pr(\mathbf{y}|\mathbf{x}^{\text{new}})$ and $\Pr(\mathbf{y}|\mathbf{x}^{\text{cur}})$ are binomial probabilities for the measurement process, and $\pi(\mathbf{x}^{\text{new}})$ and $\pi(\mathbf{x}^{\text{cur}})$ are the time–homogenous CTMC densities of the current and the proposed population–level paths that appear in Equation (**??**). Let $\pi(\mathbf{x}_j^{\text{new}}|\mathbf{\Lambda}^{(-j)};\mathcal{I})$ and $\pi(\mathbf{x}_j^{\text{cur}}|\mathbf{\Lambda}^{(-j)};\mathcal{I})$ denote the time–inhomogeneous subject–level CTMC proposal densities given by (**??**). Then,

$$\begin{aligned} q(\mathbf{x}^{\text{new}}|\mathbf{x}^{\text{cur}}) &= \Pr(\mathbf{x}^{\text{new}}|\mathbf{y};\mathbf{\Lambda}^{(-j)}(\mathbf{x}^{\text{cur}}),\mathcal{I}) \\ &= \frac{\pi(\mathbf{x}^{\text{new}},\mathbf{y};\mathbf{\Lambda}^{(-j)}(\mathbf{x}^{\text{cur}}),\mathcal{I})}{\Pr(\mathbf{y};\mathbf{\Lambda}^{(-j)},\mathcal{I})} \\ &= \frac{\Pr(\mathbf{y}|\mathbf{x}^{\text{new}})\pi(\mathbf{x}_j^{\text{new}}|\mathbf{\Lambda}^{(-j)};\mathcal{I})}{\Pr(\mathbf{y};\mathbf{\Lambda}^{(-j)}(\mathbf{x}^{\text{new}}),\mathcal{I})} \end{aligned}$$

and similarly,

$$q(\mathbf{x}^{\text{cur}}|\mathbf{x}^{\text{new}}) = \frac{\Pr(\mathbf{y}|\mathbf{x}^{\text{cur}})\pi(\mathbf{x}_j^{\text{cur}}|\mathbf{\Lambda}^{(-j)};\mathcal{I})}{\Pr(\mathbf{y};\mathbf{\Lambda}^{(-j)}(\mathbf{x}^{\text{cur}}),\mathcal{I})}.$$

Therefore,

$$\begin{aligned} \frac{\pi(\mathbf{x}^{\text{new}}|\mathbf{y})}{\pi(\mathbf{x}^{\text{cur}}|\mathbf{y})} \frac{q(\mathbf{x}^{\text{cur}}|\mathbf{x}^{\text{new}})}{q(\mathbf{x}^{\text{new}}|\mathbf{x}^{\text{cur}})} &= \frac{\Pr(\mathbf{y}|\mathbf{x}^{\text{new}})\pi(\mathbf{x}^{\text{new}})}{\Pr(\mathbf{y}|\mathbf{x}^{\text{cur}})\pi(\mathbf{x}^{\text{cur}})} \frac{\Pr(\mathbf{y}|\mathbf{x}^{\text{cur}})\pi(\mathbf{x}_j^{\text{cur}};\mathbf{\Lambda}^{(-j)})}{\Pr(\mathbf{y}|\mathbf{x}^{\text{new}})\pi(\mathbf{x}_j^{\text{new}};\mathbf{\Lambda}^{(-j)})} \\ &= \frac{\pi(\mathbf{x}^{\text{new}})}{\pi(\mathbf{x}^{\text{cur}})} \frac{\pi(\mathbf{x}_j^{\text{cur}}|\mathbf{\Lambda}^{(-j)};\mathcal{I})}{\pi(\mathbf{x}_j^{\text{new}}|\mathbf{\Lambda}^{(-j)};\mathcal{I})}. \end{aligned}$$

Hence,

$$a_{\mathbf{x}^{\text{cur}} \longrightarrow \mathbf{x}^{\text{new}}} = \min\left\{ \frac{\pi(\mathbf{x}^{\text{new}})}{\pi(\mathbf{x}^{\text{cur}})} \frac{\pi(\mathbf{x}_j^{\text{cur}}|\mathbf{\Lambda}^{(-j)};\mathcal{I})}{\pi(\mathbf{x}_j^{\text{new}}|\mathbf{\Lambda}^{(-j)};\mathcal{I})}, 1 \right\}.$$

# S6 Fitting SEIR and SIRS models via Bayesian data augmentation

## S6.1 SEIR model formulation

The Susceptible–Exposed–Infectious–Recovered (SEIR) model adds an additional latent state to the SIR model in which subjects who are exposed to an infected individual incubate before becoming infectious. As with the SIR model, recovery is assumed to confer lifelong immunity. The structure of this model does not affect any of the machinery involved in the subject–path proposal mechanism, but rather merely redefines the population–level time–homogeneous CMTC for the epidemic process, and the subject–level time–inhomogeneous CTMC used in the subject–path proposals.

Under this model, we suppose that the data are sampled from a latent epidemic process, $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$, that evolves in continuous–time as individuals become exposed, infectious, and recover. The state space of this process is $\mathcal{S} = \{S, E, I, R\}^N$, the Cartesian product of $N$ state labels taking values in $\{S, E, I, R\}$. The state space of a single subject, $\mathbf{X}_j$, is $\mathcal{S}_j = \{S, E, I, R\}$, and a realized subject–path is of the form $\mathbf{x}_j(\tau) = \left(S,\ \tau < \tau_{\mathrm{E}}^{(j)}; E,\ \tau_{\mathrm{E}}^{(j)} \leqslant \tau < \tau_{\mathrm{I}}^{(j)}; I,\ \tau_{\mathrm{I}}^{(j)} \leqslant \tau < \tau_{\mathrm{R}}^{(j)}; R\ ,\ \tau_{\mathrm{R}}^{(j)} \leqslant \tau\right)$ where $\tau_{\mathrm{E}}^{(j)}$, $\tau_{\mathrm{I}}^{(j)}$, and $\tau_{\mathrm{R}}^{(j)}$ are the times at which subject $J$ becomes exposed, infectious, and recovers. As with the SIR model, some or all of these events may not transpire in the observation period $[t_1, t_L]$, or at all. We let $\beta$ be the per–contact infectivity rate, $\gamma$ be the rate at which an exposed individual becomes infectious, and $\mu$ be the rate at which an infectious individual recovers. Furthermore, we write the vector of disease state probabilities as $\mathbf{p}_{t_1} = (p_S, p_E, p_I, p_R)$. The latent epidemic process evolves according to a time–homogeneous CTMC, with transition rate from configuration $\mathbf{x}$ to $\mathbf{x}'$ that differ only in the state of one subject $j$ is given by $\mathbf{\Lambda} = \beta I$ if $\mathbf{X}_j = S$ and $\mathbf{X}'_j = E$, $\gamma$ if $\mathbf{X}_j = E$ and $\mathbf{X}'_j = I$, and $\mu$ if $\mathbf{X}_j = I$ and $\mathbf{X}'_j = R$. Finally, the time–inhomogeneous CTMC rate matrices used in the subject–path proposal distribution have the form

$$\mathbf{\Lambda}_m^{(-j)}(\boldsymbol{\theta}) = \begin{array}{c} \\ S \\ E \\ I \\ R \end{array} \begin{array}{c} \begin{array}{cccc} S & E & I & R \end{array} \\ \begin{pmatrix} -\beta I_{\tau_m}^{(-j)} & \beta I_{\tau_m}^{(-j)} & 0 & 0 \\ 0 & -\gamma & \gamma & 0 \\ 0 & 0 & -\mu & \mu \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}. \tag{5}$$

As is the case with the SIR model, the eigen–values of the CTMC rate matrices for the SEIR model are always real valued. The only computational modification, relative to the SIR model, that we suggest is that times of state transition in inter–event intervals be sampled conditional on the state at the endpoints via uniformization (see Section S4 of the supplement).

## S6.2 SIRS model formulation

The Susceptible–Infected–Recovered–Susceptible (SIRS) model modifies the SIR model to allow for loss of immunity. Again, fitting this model using our Bayesian data augmentation algorithm

does not affect any of the machinery involved in the subject–path proposal mechanism, although the recurrent nature of the disease dynamics increase the computational burden of the algorithm since the disease state at the interval endpoints does absolve us of sampling the path within each inter–event interval where the states at the endpoints are the same.

Under the SIRS model, we suppose that the data are sampled from a latent epidemic process, $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$, that evolves in continuous–time as individuals become exposed, infectious, and recover. The state space of this process is $\mathcal{S} = \{S, I, R\}^N$, the Cartesian product of $N$ state labels taking values in $\{S, I, R\}$. The state space of a single subject, $\mathbf{X}_j$, is $\mathcal{S}_j = \{S, I, R\}$, and a realized subject–path is of the form
$\mathbf{x}_j(\tau) = \Big(S,\ \tau < \tau^{(j)}_{\mathrm{I}_1}; I,\ \tau^{(j)}_{\mathrm{I}_1} \leqslant \tau < \tau^{(j)}_{\mathrm{R}_1}; R\ ,\ \tau^{(j)}_{\mathrm{R}_1} \leqslant \tau < \tau^{(j)}_{\mathrm{L}_1}; S\ ,\ \tau^{(j)}_{\mathrm{L}_1} \leqslant \tau < \tau^{(j)}_{\mathrm{I}_2}; \ldots\Big)$,
where $\tau^{(j)}_{\mathrm{I}_k}$, $\tau^{(j)}_{\mathrm{R}_k}$, and $\tau^{(j)}_{\mathrm{L}_k}$ are times at which subject $J$ becomes infected, recovers, and loses immunity, and are ennumerated by the subscript $k$ as the process may revisit each state multiple time. As with the SIR and SEIR models, some or all of these events may not transpire in the observation period $[t_1, t_L]$, or at all. We let $\beta$ be the per–contact infectivity rate, $\mu$ be the rate at which an infectious individual recovers, and $\gamma$ be the rate at which immunity is lost. Furthermore, we write the vector of disease state probabilities as $\mathbf{p}_{t_1} = (p_S, p_I, p_R)$. The latent epidemic process evolves according to a time–homogeneous CTMC, with transition rate from configuration $\mathbf{x}$ to $\mathbf{x}'$ that differ only in the state of one subject $j$ is given by $\mathbf{\Lambda} = \beta I$ if $\mathbf{X}_j = S$ and $\mathbf{X}'_j = E$, $\mu$ if $\mathbf{X}_j = I$ and $\mathbf{X}'_j = R$, and $\gamma$ if $\mathbf{X}_j = R$ and $\mathbf{X}'_j = S$. Finally, the time–inhomogeneous CTMC rate matrices used in the subject–path proposal distribution have the form

$$\mathbf{\Lambda}^{(-j)}_m(\boldsymbol{\theta}) = \begin{array}{c} \\ S \\ I \\ R \end{array} \begin{array}{c} \begin{array}{ccc} S & I & R \end{array} \\ \begin{pmatrix} -\beta I^{(-j)}_{\tau_m} & \beta I^{(-j)}_{\tau_m} & 0 \\ 0 & -\mu & \mu \\ \gamma & 0 & -\gamma \end{pmatrix} \end{array}. \tag{6}$$

Unlike the SIR and SEIR models, eigenvalues of each CTMC rate matrix may be complex. In order to obtain a real valued transition probability matrix over an interval for which eigen–values of the rate matrix are complex, we must rotate that rate matrix to obtain its real canonical form. This is further discussed in Section S2 of the Supplement.

# S7 Selecting the Number of Subject–Paths to Update per MCMC Iteration

There is no need to re–sample the path of every subject within each MCMC iteration. Indeed, we might suspect that the efficiency of our MCMC could be improved by sampling only a few subject–paths between parameter updates. Successive subject–path proposal tend to be highly autocorrelated, as is the case for traditional DA methods (Roberts and Stramer, 2001), and frequently updating model parameters may help to break this correlation. Although parameter updates tend to also be highly autocorrelated, subject–path proposals are costly compared to updates of model parameters. Therefore, it is reasonable to suspect that the effective sample size (ESS) per CPU time might be improved by sampling only a handful of subject–paths per MCMC iteration.

Many factors, including the SEM dynamics, population size, efficiency of the implementation, and the degree of model misspecification could affect the optimal number subject–path updates per MCMC iteration. It is clearly impossible to disentangle the effects of all of the possible factors that could affect the optimal number of subject–path updates per iteration. In the main paper, we set the number of subject–paths per iteration on the basis of log–posterior effective sample size (ESS) per CPU time in an initial run of 5,000–10,000 iterations (depending on the simulation).

# S8 Prior and Full–Conditional Distributions of SEM parameters

| Parameter | Conjugate Prior Dist. | Prior Hyperparameters | Full Conditional Hyperparameters |
|---|---|---|---|
| $R_0$ | Beta$'$ | $a_\beta,\ a_\mu,\ 1,\ \frac{b_\mu N}{b_\beta}$ | — |
| $\beta$ | Gamma | $a_\beta,\ b_\beta$ | $a_\beta + \sum_{j=1}^{M} \mathbb{I}\left(\tau_j \triangleq I\right),\ b_\beta + \sum_{j=1}^{M} S_{\tau_{j-1}} I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$ |
| $\mu$ | Gamma | $a_\mu,\ b_\mu$ | $a_\mu + \sum_{j=1}^{M} \mathbb{I}\left(\tau_j \triangleq R\right),\ b_\mu + \sum_{j=1}^{M} I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$ |
| $\rho$ | Beta | $a_\rho,\ b_\rho$ | $a_\rho + \sum_{j=1}^{L} Y_{t_j},\ b_\rho + \sum_{j=1}^{L} (I_{t_j} - Y_{t_j})$ |
| $\mathbf{p}_{t_1}$ | Dirichlet | $a_S,\ b_I,\ c_R$ | $a_S + S_{t_1},\ b_I + I_{t_1},\ c_R + R_{t_1}$ |

**Table S1:** Prior and full conditional distributions for SIR model parameters. $\beta$ is the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability, and $\mathbf{p}_{t_1}$ is the vector of initial state probabilities. Gamma priors are parameterized with rates, so a Gamma$(a, b)$ distribution has mean $a/b$. The Beta prime prior for $R_0 = \beta N/\mu$ is the implied prior induced by the prior distributions for $\beta$ and $\mu$. The indicators $\mathbb{I}\left(\tau_j \triangleq I\right)$ and $\mathbb{I}\left(\tau_j \triangleq R\right)$ equal 1 if $\tau_j$ corresponds to a time when an individual becomes infected or recovers.

| Parameter | Conjugate Prior Dist. | Prior Hyperparameters | Full Conditional Hyperparameters |
|---|---|---|---|
| $R_0$ | Beta$'$ | $a_\beta,\ a_\mu,\ 1,\ \frac{b_\mu N}{b_\beta}$ | — |
| $\beta$ | Gamma | $a_\beta,\ b_\beta$ | $a_\beta + \sum_{j=1}^{M} \mathbb{I}\left(\tau_j \triangleq E\right),\ b_\beta + \sum_{j=1}^{M} S_{\tau_{j-1}} I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$ |
| $\gamma$ | Gamma | $a_\gamma,\ b_\gamma$ | $a_\gamma + \sum_{j=1}^{M} \mathbb{I}\left(\tau_j \triangleq I\right),\ b_\gamma + \sum_{j=1}^{M} E_{\tau_{j-1}} (\tau_j - \tau_{j-1})$ |
| $\mu$ | Gamma | $a_\mu,\ b_\mu$ | $a_\mu + \sum_{j=1}^{M} \mathbb{I}\left(\tau_j \triangleq R\right),\ b_\mu + \sum_{j=1}^{M} I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$ |
| $\rho$ | Beta | $a_\rho,\ b_\rho$ | $a_\rho + \sum_{j=1}^{L} Y_{t_j},\ b_\rho + \sum_{j=1}^{L} (I_{t_j} - Y_{t_j})$ |
| $\mathbf{p}_{t_1}$ | Dirichlet | $a_S,\ b_I,\ c_R$ | $a_S + S_{t_1},\ b_I + I_{t_1},\ c_R + R_{t_1}$ |

**Table S2:** Prior and full conditional distributions for SEIR model parameters. $\beta$ is the per–contact infectivity rate, $\gamma$ is the rate at which an exposed individual becomes infectious, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability, and $\mathbf{p}_{t_1}$ is the vector of initial state probabilities. Gamma priors are parameterized with rates, so a Gamma$(a, b)$ distribution has mean $a/b$. The Beta prime prior for $R_0 = \beta N/\mu$ is the implied prior induced by the prior distributions for $\beta$ and $\mu$. The indicators $\mathbb{I}\left(\tau_j \triangleq E\right)$, $\mathbb{I}\left(\tau_j \triangleq I\right)$ and $\mathbb{I}\left(\tau_j \triangleq R\right)$ equal 1 if $\tau_j$ corresponds to a time when an individual becomes exposed, becomes infectious, or recovers.

| Parameter | Conjugate Prior Dist. | Prior Hyperparameters | Full Conditional Hyperparameters |
|---|---|---|---|
| $R_0$ | Beta$'$ | $a_\beta,\ a_\mu,\ 1,\ \frac{b_\mu N}{b_\beta}$ | — |
| $\beta$ | Gamma | $a_\beta,\ b_\beta$ | $a_\beta + \sum_{j=1}^{M} \mathbb{I}\left(\tau_j \triangleq E\right),\ b_\beta + \sum_{j=1}^{M} S_{\tau_{j-1}} I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$ |
| $\mu$ | Gamma | $a_\mu,\ b_\mu$ | $a_\mu + \sum_{j=1}^{M} \mathbb{I}\left(\tau_j \triangleq R\right),\ b_\mu + \sum_{j=1}^{M} I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$ |
| $\gamma$ | Gamma | $a_\gamma,\ b_\gamma$ | $a_\gamma + \sum_{j=1}^{M} \mathbb{I}\left(\tau_j \triangleq L\right),\ b_\gamma + \sum_{j=1}^{M} R_{\tau_{j-1}} (\tau_j - \tau_{j-1})$ |
| $\rho$ | Beta | $a_\rho,\ b_\rho$ | $a_\rho + \sum_{j=1}^{L} Y_{t_j},\ b_\rho + \sum_{j=1}^{L} (I_{t_j} - Y_{t_j})$ |
| $\mathbf{p}_{t_1}$ | Dirichlet | $a_S,\ b_I,\ c_R$ | $a_S + S_{t_1},\ b_I + I_{t_1},\ c_R + R_{t_1}$ |

**Table S3:** Prior and full conditional distributions for SIRS model parameters. $\beta$ is the per–contact infectivity rate, $\mu$ is the recovery rate, $\gamma$ is the rate at which a recovered individual loses immunity, $\rho$ is the binomial sampling probability, and $\mathbf{p}_{t_1}$ is the vector of initial state probabilities. Gamma priors are parameterized with rates, so a Gamma$(a, b)$ distribution has mean $a/b$. The Beta prime prior for $R_0 = \beta N/\mu$ is the implied prior induced by the prior distributions for $\beta$ and $\mu$. The indicators $\mathbb{I}\left(\tau_j \triangleq I\right)$, $\mathbb{I}\left(\tau_j \triangleq R\right)$, and $\mathbb{I}\left(\tau_j \triangleq L\right)$ equal 1 if $\tau_j$ corresponds to a time when an individual becomes infected, recovers, or loses immunity.

# S9 Simulation 1 — Inference Under Various Epidemic Dynamics — Setup and Additional Results

## S9.1 Simulation details for the SIR model

We simulated an epidemic in a population of 750 individuals, 90% of whom were initially susceptible and 3% of whom were initially infected. Prevalence was observed with detection probability $\rho = 0.2$ at weekly intervals over a four month period which captured both the exponential growth and decline of the epidemic. The mean infectious period was $1/\mu = 7$ days and the per-contact infectivity rate was 0.00035, which combined to give a basic reproductive number was $R_0 = \beta N/\mu \approx 1.8$.

We ran three chains for 100,000 iterations each, sampling the paths for 75 subjects, chosen uniformly at random, per MCMC iteration. We discarded the first 10 iterations from each chain as burn-in. Priors for the rate parameters (summarized in Table S4) were scaled so that the prior mass spanned a reasonable range of values, but were otherwise mild. Similarly, the prior for the binomial sampling probability reflected a general prior belief that fewer than 40% of cases were detected. The prior for the initial distribution parameters was informative, and was chosen as such because of the paucity of data available for estimation of the initial distribution parameters.

We also fit the SIR model to the data using PMMH. We ran two sets of three MCMC chains with the PMMH algorithm for 50,000 iterations each with 100 particles per chain, and discarded the first 100 iterations as burn-in. The first set of chains simulated particle paths approximately using $\tau$–

| Param. | True Value | Prior distribution |
|---|---|---|
| $R_0$ | 1.8 | Beta$'$(0.3, 1, 1, 6) |
| $\beta$ | 0.00035 | Gamma(0.3, 1000) |
| $\mu$ | 0.14 | Gamma(1, 8) |
| $\mathbf{p}_{t_1}$ | (0.9, 0.03, 0.07) | Dirichlet(90, 2, 5) |
| $\rho$ | 0.2 | Beta(2, 7) |

**Table S4:** Prior distributions for SIR model and measurement process parameters. The prior for $R_0$ is the induced prior implied by $\beta$ and $\mu$. The per–contact infectivity rate is $\beta$, the recovery rate is $\mu$, the binomial sampling probability is $\rho$, and the initial state probabilities are $\mathbf{p}_{t_1}$.

leaping with a time step of two hours, while the second chain simulated paths exactly via Gillespie's direct algorithm. Parameters were updated using random walk Metropolis–Hastings (RWMH) with a proposal covariance matrix estimated from an initial run of 5,000 iterations using an adaptive RWMH algorithm with a target acceptance rate of 23.4%. We updated parameters on transformed scales in order to remove restrictions on the parameter space, applying a log transformation to $\beta$ and $\mu$, a logit transformation to $\rho$, and a generalized logit transformation to $\mathbf{p}_{t_1}$.

## S9.2 Additional results and MCMC diagnostics for the SIR model

| Method | Chain | Hours | ESS | ESS per CPU time |
|---|---|---|---|---|
| BDA | 1 | 9.9 | 87.7 | 8.8 |
| BDA | 2 | 8.7 | 67.9 | 7.8 |
| BDA | 3 | 8.5 | 63.8 | 7.5 |
| PMMH–A | 1 | 0.6 | 1847.4 | 2871.6 |
| PMMH–A | 2 | 0.6 | 1942.2 | 2995.7 |
| PMMH–A | 3 | 0.7 | 1876.6 | 2615.9 |
| PMMH–E | 1 | 26.1 | 1568.3 | 60.1 |
| PMMH–E | 2 | 20.4 | 2123.7 | 104.0 |
| PMMH–E | 3 | 20.5 | 1849.4 | 90.2 |

**Table S5:** Log–posterior run times, effective sample sizes (ESSs), and effective sample sizes per CPU time measure in hours (ESS.per.CPU.time). BDA indicates our Bayesian data augmentation algorithm, PMMH–A indicates PMMH with paths simulated approximately via $\tau$–leaping algorithm, and PMMH–E indicates PMMH with paths simulated exactly using Gillespie's direct algorithm. The BDA chains were run for 100,000 iterations each, while the PMMH chains were run for 50,000 iterations following a tuning run of 5,000 iterations.

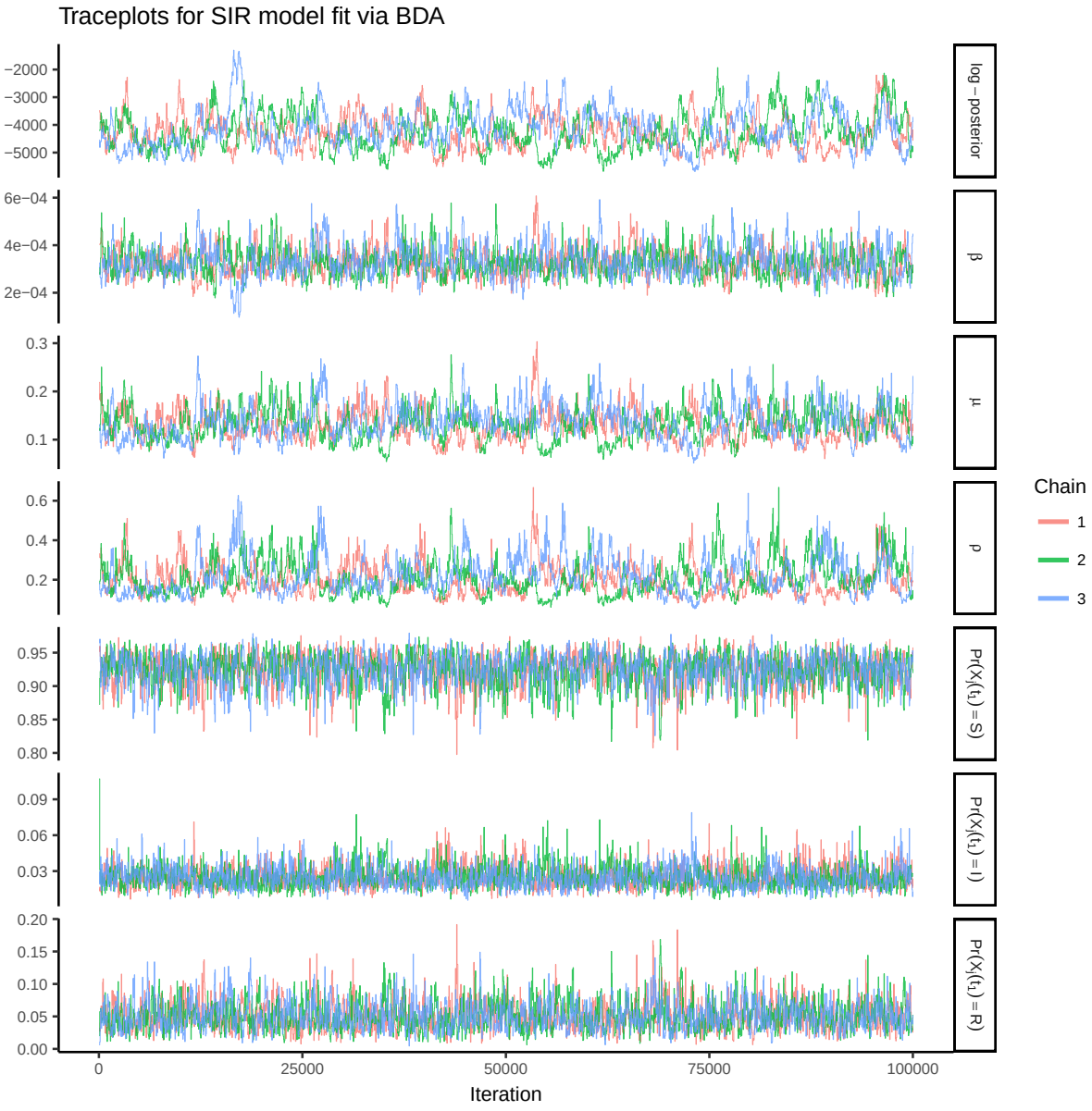


**Figure S2:** Traceplots of the log–posterior and model parameters for the SIR model fit using Bayesian data augmentation following an initial burn–in of 10 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

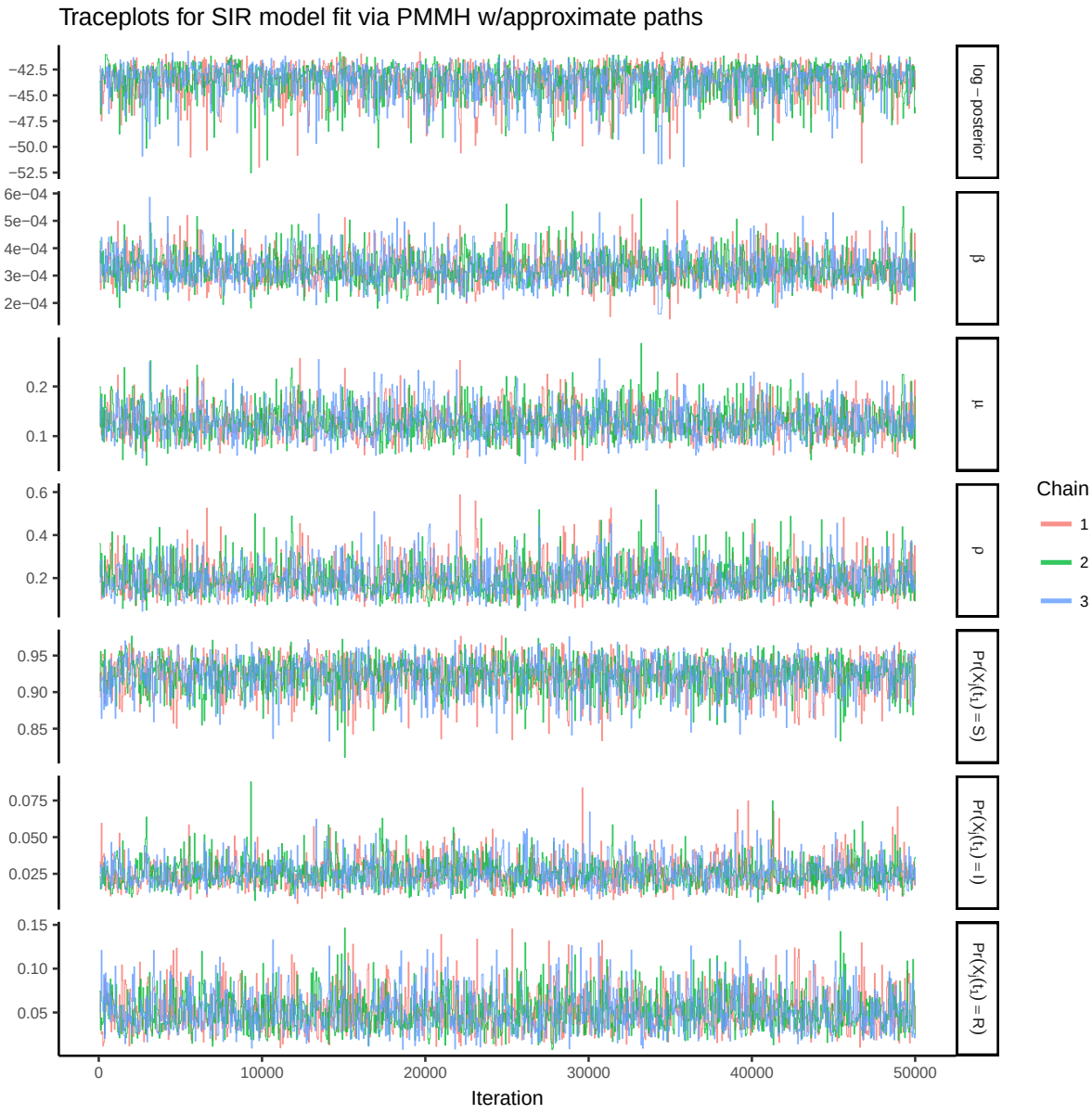


**Figure S3:** Traceplots of the log–posterior and model parameters for the SIR model fit using PMMH with 100 particles and a time step of 8 hours, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH and an initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

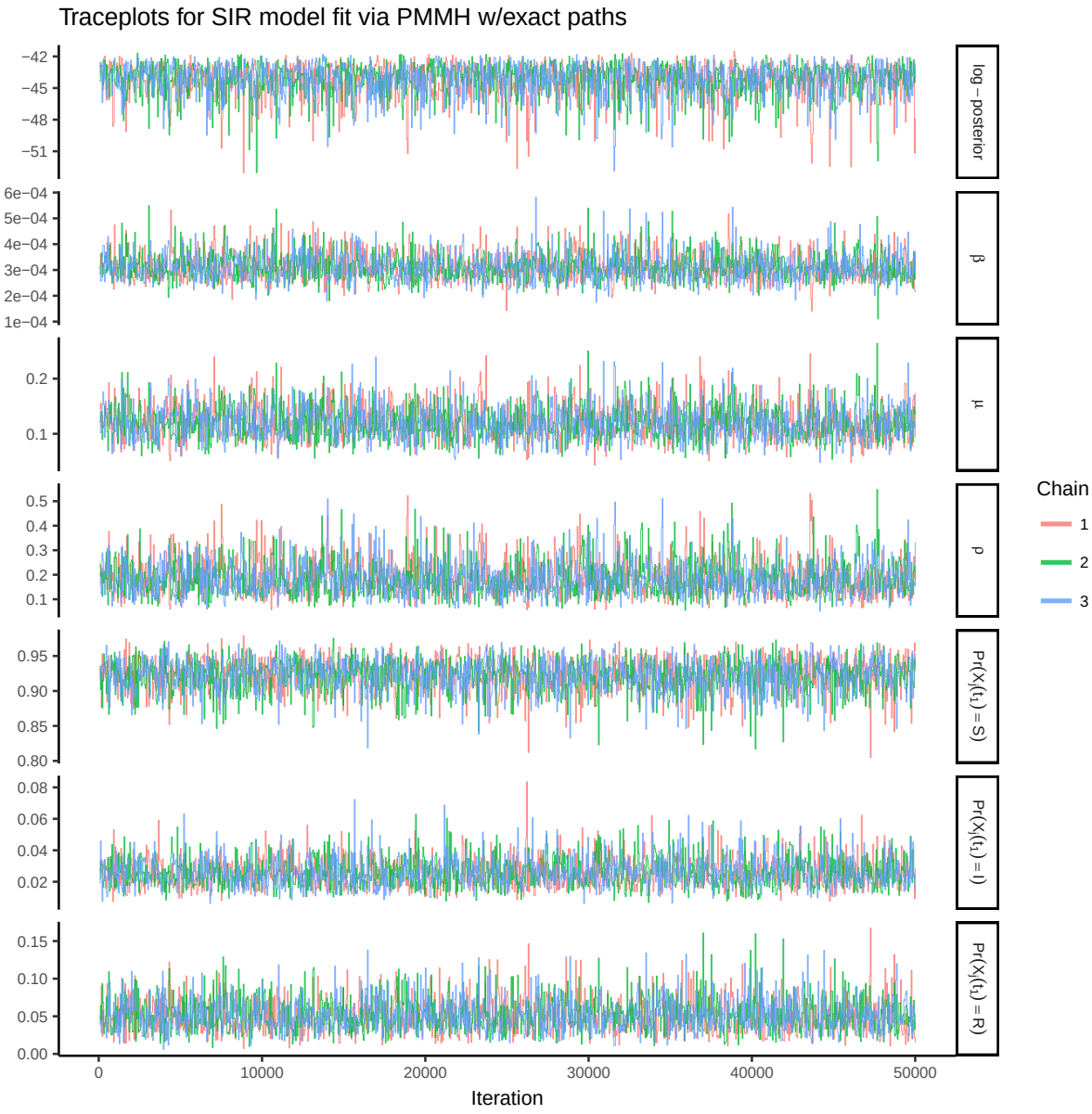


**Figure S4:** Traceplots of the log–posterior and model parameters for the SIR model fit using PMMH with 100 particles, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH and an initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

## S9.3 Simulation details for the SEIR model

We simulated an outbreak under near-endemic SEIR dynamics, with $R_0 = \beta N/\mu = 1.05$, in a population of 500 individuals. The outbreak was initiated by a single infected individual in an otherwise susceptible population, 121 of whom eventually became infected. The mean sojourn time in the exposed state was $1/\gamma = 14$ days, while the mean infectious period duration was $1/\mu = 28$ days. Prevalence was observed at weekly intervals, with detection probability $\rho = 0.3$, over a two year period.

We ran three chains for 100,000 iterations each, sampling the paths for 100 subjects, chosen uniformly at random, per MCMC iteration. We discarded the first 10 iterations from each chain as burn-in. Priors for the rate parameters (summarized in Table S6) were scaled so that the prior mass spanned a reasonable range of values, but were otherwise mild. The prior for the binomial sampling probability was chosen so that 80% of the mass was between roughly 15 and 55 percent. The prior for the initial distribution parameters was informative.

| Param. | True Value | Prior distribution |
|---|---|---|
| $R_0$ | 1.05 | Beta′(1, 3.2, 1, 5) |
| $\beta$ | 0.000075 | Gamma(1, 10000) |
| $\gamma$ | 0.071 | Gamma(1, 11) |
| $\mu$ | 0.036 | Gamma(3.2, 100) |
| $\mathbf{p}_{t_1}$ | (0.998, 0.006, 0.002, 0, 0) | Dirichlet(100, 0.1, 0.4, 0.01) |
| $\rho$ | 0.3 | Beta(3.5, 6.5) |

**Table S6:** Prior distributions for SEIR model and measurement process parameters. The prior for $R_0$ is the induced prior implied by $\beta$ and $\mu$. The per–contact infectivity rate is $\beta$, the rate at which an exposed individual becomes infectious is $\gamma$, the recovery rate is $\mu$, the binomial sampling probability is $\rho$, and the initial state probabilities are $\mathbf{p}_{t_1}$.

We also fit the SEIR model to the data using PMMH. We ran two sets of three MCMC chains with the PMMH algorithm for 50,000 iterations each with 200 particles per chain, and discarded the first 100 iterations as burn-in. The first set of chains simulated particle paths approximately using $\tau$–leaping with a time step of 8 hours, while the second chain simulated paths exactly via Gillespie's direct algorithm. Parameters were updated using random walk Metropolis–Hastings (RWMH) with a proposal covariance matrix estimated from an initial run of 5,000 iterations using an adaptive RWMH algorithm with a target acceptance rate of 23.4%. We updated parameters on transformed scales in order to remove restrictions on the parameter space, applying a log transformation to $\beta$, $\gamma$, and $\mu$, a logit transformation to $\rho$, and a generalized logit transformation to $\mathbf{p}_{t_1}$.

## S9.4 Additional results and MCMC diagnostics for the SEIR model

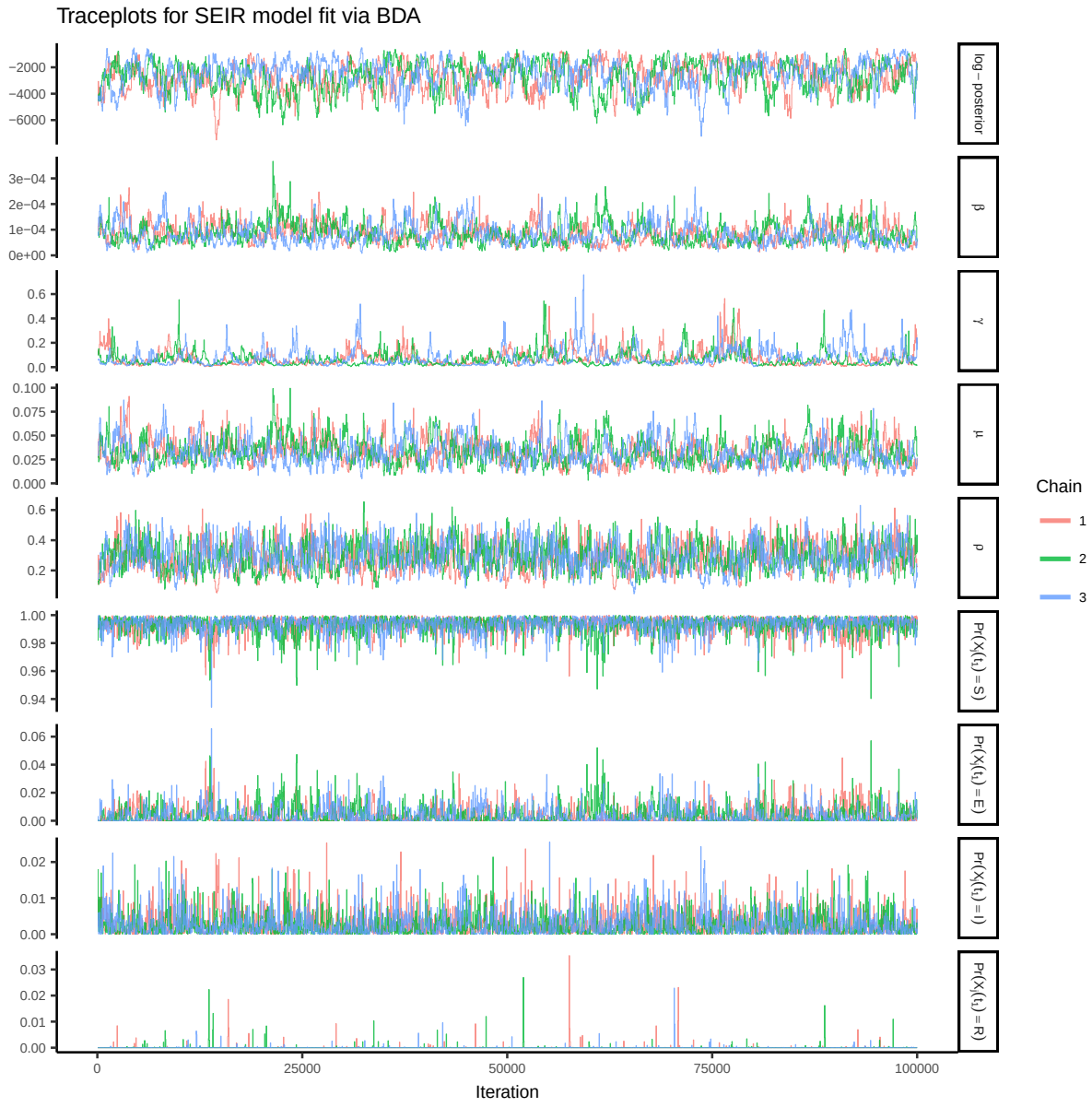


**Figure S5:** Traceplots of the log–posterior and model parameters for the SIR model fit using Bayesian data augmentation following an initial burn–in of 10 iterations. $\beta$ denotes the per–contact infectivity rate, $\gamma$ is the rate at which exposed individuals become infectious, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

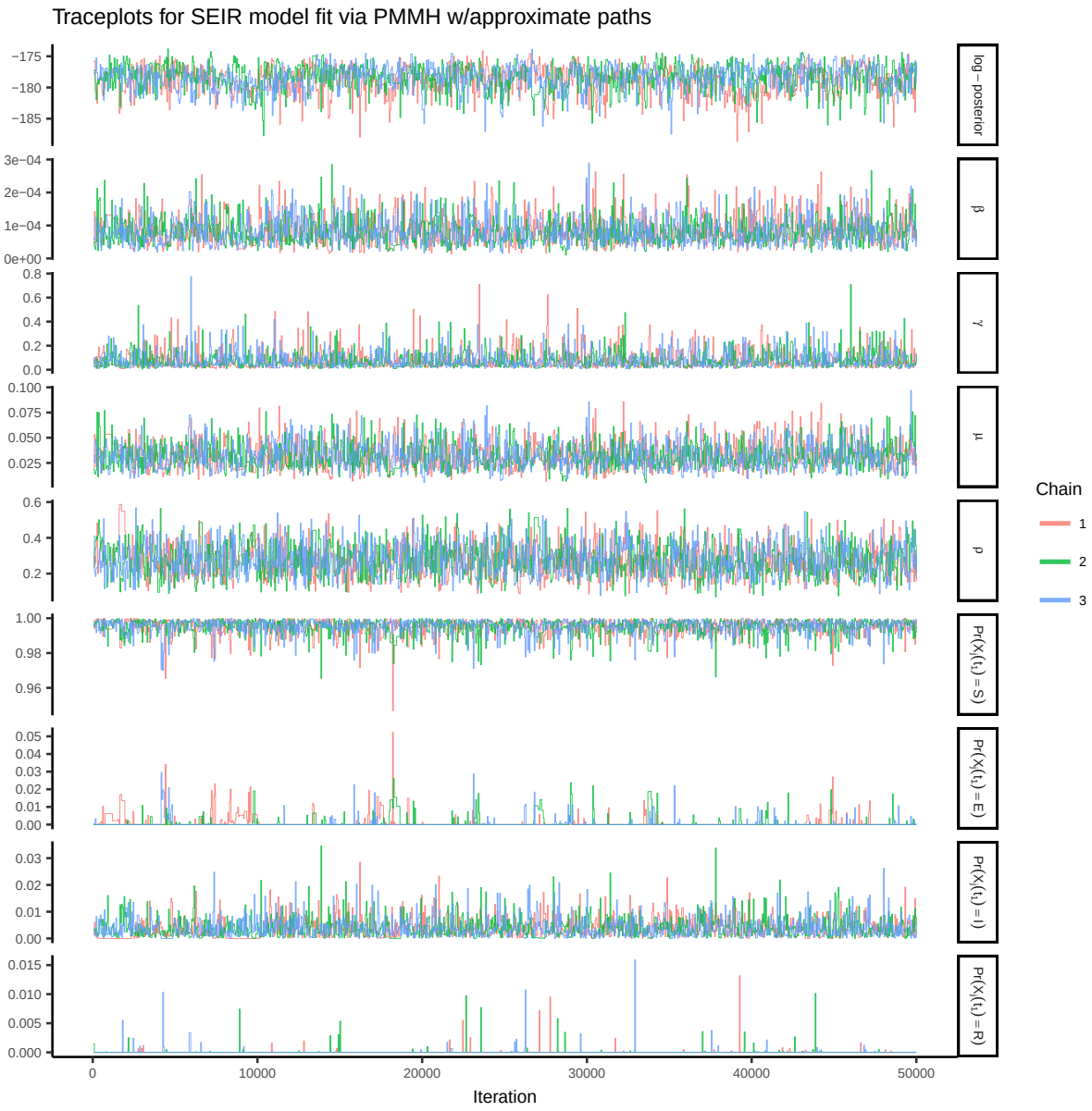


**Figure S6:** Traceplots of the log–posterior and model parameters for the SIR model fit using PMMH with 200 particles and a time step of 8 hours, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH and an initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\gamma$ is the rate at which exposed individuals become infectious, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

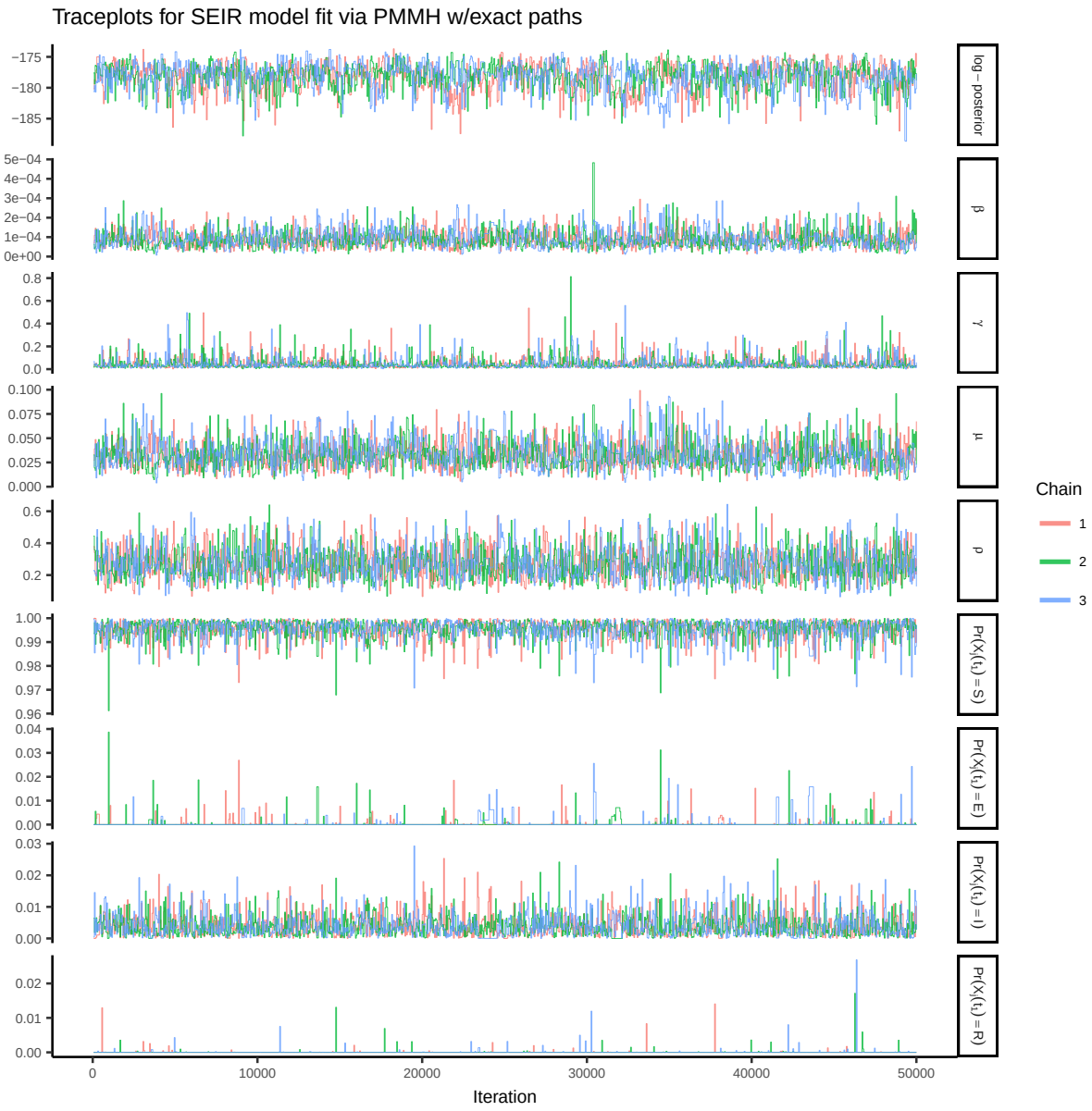


**Figure S7:** Traceplots of the log–posterior and model parameters for the SIR model fit using PMMH with 200 particles, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH and an initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\gamma$ is the rate at which exposed individuals become infectious, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

| Model | Method | Chain | Time | ESS | ESS per CPU time |
|---|---|---|---|---|---|
| SEIR | BDA | 1 | 9.2 | 149.9 | 16.2 |
| SEIR | BDA | 2 | 9.2 | 146.0 | 15.9 |
| SEIR | BDA | 3 | 9.0 | 143.9 | 16.0 |
| SEIR | PMMH - A | 1 | 8.1 | 483.6 | 59.5 |
| SEIR | PMMH - A | 2 | 8.3 | 684.8 | 82.2 |
| SEIR | PMMH - A | 3 | 8.4 | 570.5 | 67.9 |
| SEIR | PMMH - E | 1 | 15.8 | 411.9 | 26.1 |
| SEIR | PMMH - E | 2 | 15.9 | 589.8 | 37.1 |
| SEIR | PMMH - E | 3 | 14.1 | 466.3 | 33.1 |

**Table S7:** Log–posterior run times, effective sample sizes (ESSs), and effective sample sizes per CPU time measure in hours (ESS.per.CPU.time). BDA indicates our Bayesian data augmentation algorithm, PMMH–A indicates PMMH with paths simulated approximately via $\tau$–leaping algorithm, and PMMH–E indicates PMMH with paths simulated exactly using Gillespie's direct algorithm. The BDA chains were run for 100,000 iterations each, while the PMMH chains were run for 50,000 iterations following a tuning run of 5,000 iterations.

## S9.5 Simulation details for the SIRS model

The final outbreak was simulated under SIRS dynamics in a population of 200 individuals, in which $R_0 = \beta N/\mu = 2.52$, the mean infectious period was $1/\mu = 14$ days, and the mean time until loss of immunity was $1/\gamma = 150$ days. One percent of of the population was initially at the time of the first observation and the rest of the individuals were susceptible. Prevalence was observed weekly, with detection probability $\rho = 0.95$, over a one year period that spanned the initial wave of the epidemic as well as most of the second wave of the epidemic.

We ran three chains for 300,000 iterations each, sampling the paths for 3 subjects, chosen uniformly at random, per MCMC iteration. We discarded the first 2,000 iterations from each chain as burn-in. Priors for the rate parameters (summarized in Table S8) were scaled so that the prior mass spanned a reasonable range of values, but were otherwise mild. Similarly, the prior for the binomial sampling probability reflected a general prior belief that more than 60% of cases were detected, but was not otherwise particularly informative. The prior for the initial distribution parameters was informative.

We also fit the SIRS model to the data using PMMH. We ran three MCMC chains with the PMMH algorithm for 50,000 iterations each with 500 particles per chain, and discarded the first 100 iterations as burn-in. We also ran a set of chains with 200 particles but mixing was poor and not all of the chains converged. We attempted to exactly simulate particle paths but ultimately failed due to degeneracies in the algorithm. The time step for the $\tau$–leaping algorithm was 8 hours. Parameters were updated using random walk Metropolis–Hastings (RWMH) with a proposal covariance matrix estimated from an initial run of 5,000 iterations using an adaptive RWMH algorithm with a target acceptance rate of 23.4%. We updated parameters on transformed scales in order to remove restrictions on the parameter space, applying a log transformation to $\beta$, $\mu$, and $\gamma$, a logit transformation to $\rho$, and a generalized logit transformation to $\mathbf{p}_{t_1}$.

| Param. | True Value | Prior distribution |
|---|---|---|
| $R_0$ | 2.52 | Beta$'$(0.1, 1.5, 1, 28) |
| $\beta$ | 0.1 | Gamma$(0.1, 100)$ |
| $\mu$ | 0.036 | Gamma(1.8, 14) |
| $\gamma$ | 0.071 | Gamma$(0.0625, 10)$ |
| $\mathbf{p}_{t_1}$ | (0.99, 0.01, 0) | Dirichlet(90, 1.5, 0.01) |
| $\rho$ | 0.95 | Beta(5, 1) |

**Table S8:** Prior distributions for SIRS model and measurement process parameters. The prior for $R_0$ is the induced prior implied by $\beta$ and $\mu$. The per–contact infectivity rate is $\beta$, the recovery rate is $\mu$, the rate at which immunity is lost is $\gamma$, the binomial sampling probability is $\rho$, and the initial state probabilities are $\mathbf{p}_{t_1}$.

## S9.6 Additional results and MCMC diagnostics for the SIRS model

| Model | Method | Chain | Time | ESS | ESS per CPU time |
|---|---|---|---|---|---|
| SIRS | BDA | 1 | 14.2 | 167.7 | 11.8 |
| SIRS | BDA | 2 | 10.9 | 194.8 | 17.8 |
| SIRS | BDA | 3 | 10.8 | 243.0 | 22.6 |
| SIRS | PMMH - A | 1 | 3.1 | 670.8 | 214.1 |
| SIRS | PMMH - A | 2 | 3.0 | 799.5 | 267.3 |
| SIRS | PMMH - A | 3 | 3.5 | 766.2 | 217.1 |
| SIRS | PMMH - E | 1 | 50.2 | 570.9 | 11.4 |
| SIRS | PMMH - E | 2 | 48.6 | 667.6 | 13.7 |
| SIRS | PMMH - E | 3 | 48.8 | 592.6 | 12.1 |

**Table S9:** Log–posterior run times, effective sample sizes (ESSs), and effective sample sizes per CPU time measure in hours (ESS.per.CPU.time). BDA indicates our Bayesian data augmentation algorithm, PMMH–A indicates PMMH with paths simulated approximately via $\tau$–leaping algorithm, and PMMH–E indicates PMMH with paths simulated exactly using Gillespie's direct algorithm. The BDA chains were run for 100,000 iterations each, while the PMMH chains were run for 50,000 iterations following a tuning run of 5,000 iterations.

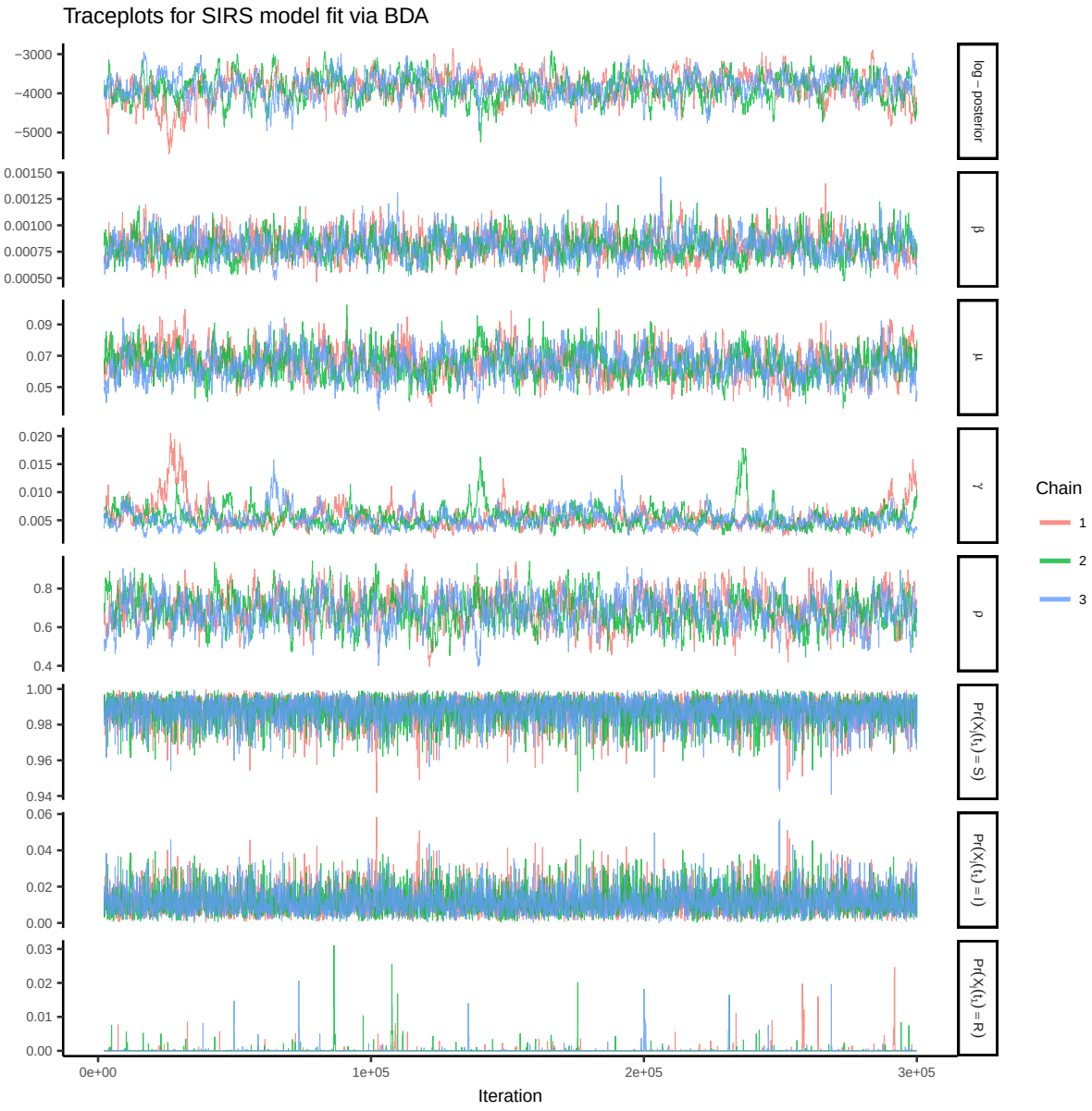


**Figure S8:** Traceplots of the log–posterior and model parameters for the SIR model fit using Bayesian data augmentation following an initial burn–in of 2,000 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, *gamma* is the rate at which immunity is lost, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

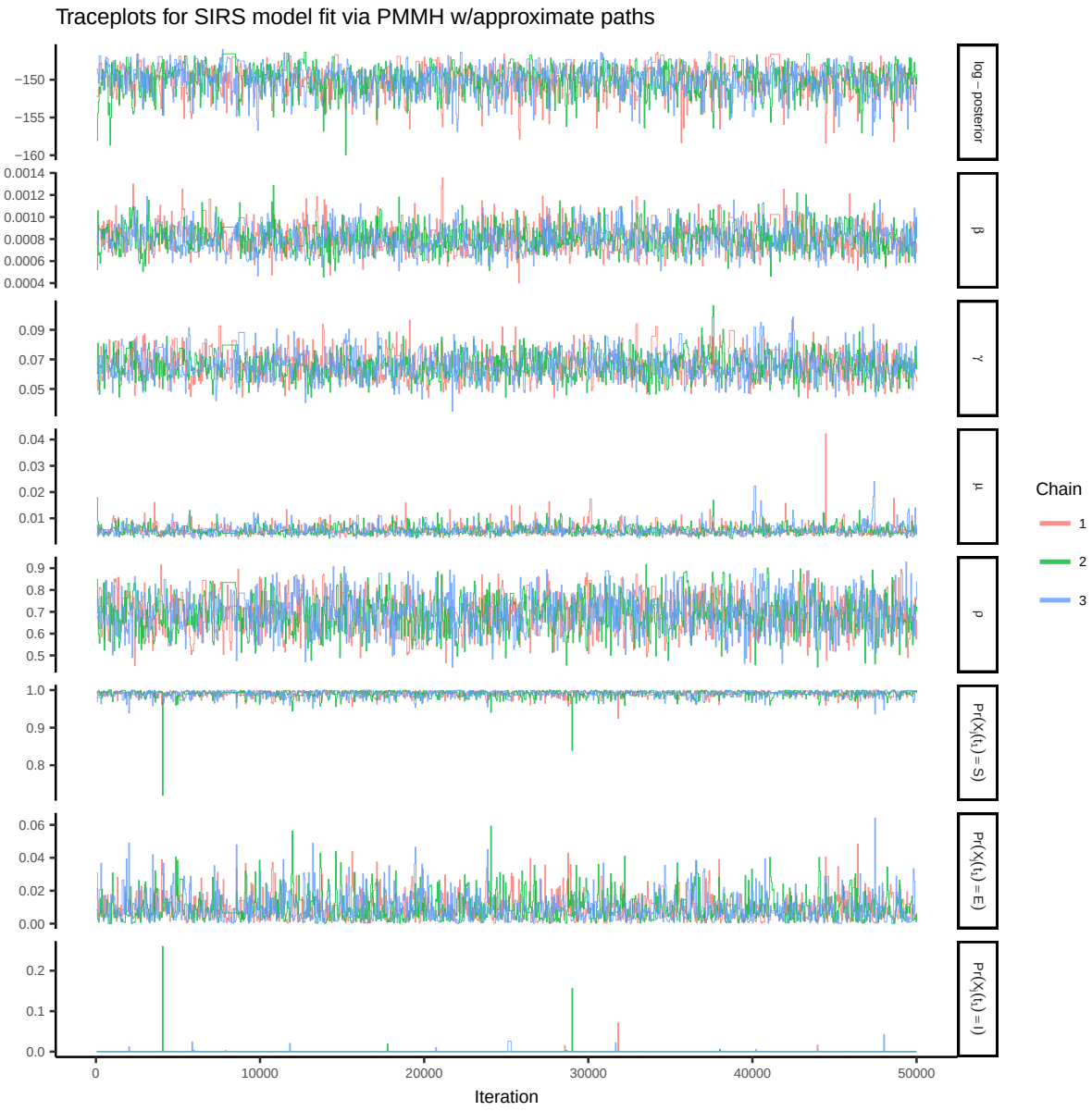


**Figure S9:** Traceplots of the log–posterior and model parameters for the SIR model fit using PMMH with 500 particles per chain and a time–step of 8 hours in the approximate $\tau$–leaping algorithm, following a tuning run of 5,000 iterations to estimate the RWMH covariance matrix and in initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, *gamma* is the rate at which immunity is lost, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

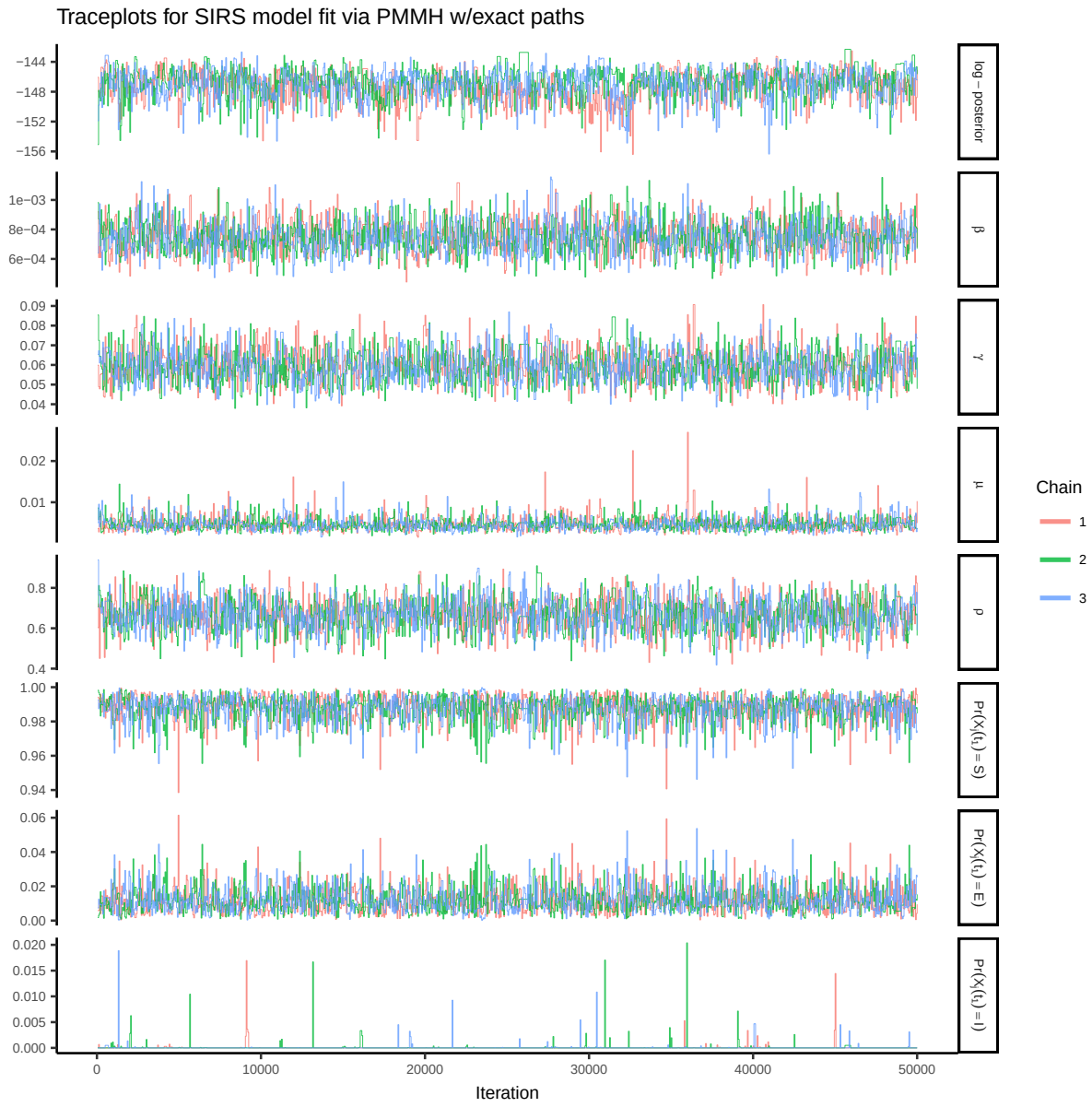


**Figure S10:** Traceplots of the log–posterior and model parameters for the SIR model fit using PMMH with 500 particles per chain and particle paths simulated exactly via Gillespie's direct algorithm, following a tuning run of 5,000 iterations to estimate the RWMH covariance matrix and in initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, *gamma* is the rate at which immunity is lost, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

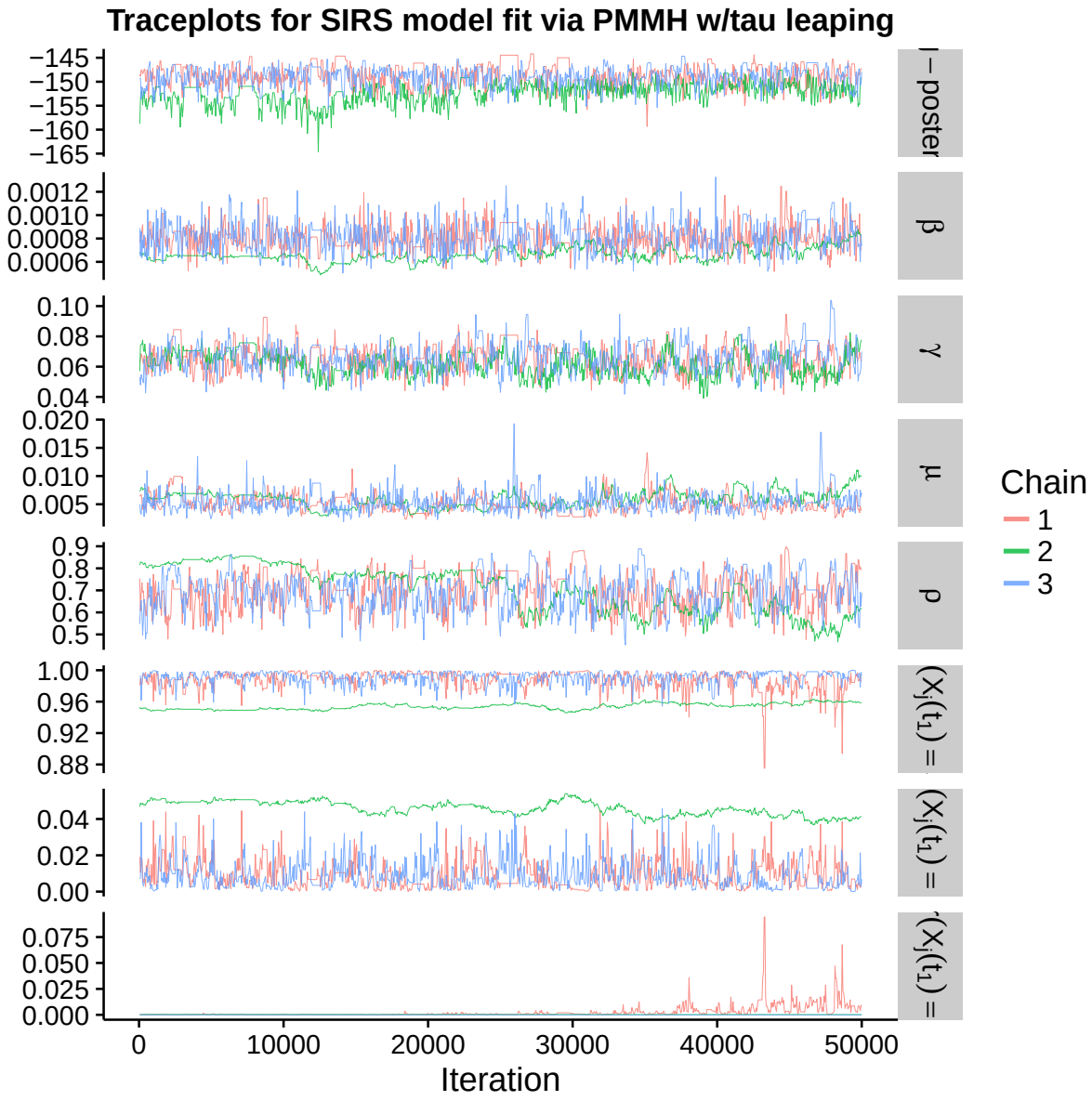


**Figure S11:** Traceplots of the log–posterior and model parameters for the SIR model fit using PMMH with 200 particles per chain and a time–step of 8 hours in the approximate $\tau$–leaping algorithm, following a tuning run of 5,000 iterations to estimate the RWMH covariance matrix and in initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, *gamma* is the rate at which immunity is lost, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

## S9.7 Estimated latent posterior distributions for all models

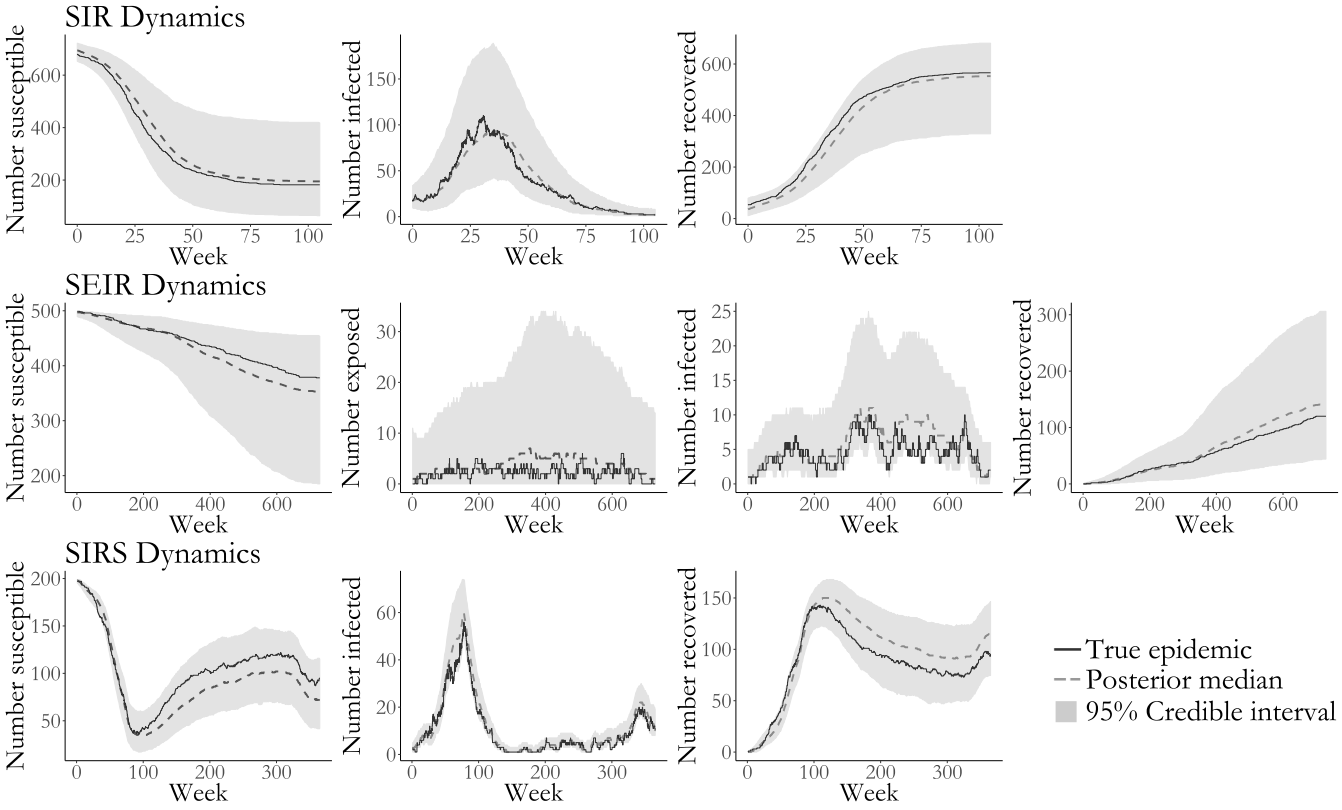


**Figure S12:** Pointwise posterior medians (dashed lines) and pointwise 95% credible intervals for the numbers of individuals in each disease state for the SIR, SEIR, and SIRS models. True compartment counts are shown as solid lines. Estimates are based on a thinned sample, retaining the collection of disease histories at the end of every $250^{th}$ MCMC iteration.

# S10 Simulation 2 — Inference under Model Misspecification — Setup and Additional Results

## S10.1 Simulation setup

We simulated an epidemic in a population of size N=400 with time–varying dynamics using Gillespie's direct algorithm over a four year period. Weekly prevalence counts were binomially distributed with detection probability $\rho = 0.95$. The epidemic dynamics varied over four epochs, based on the parameters given in Table S10. We fit SIR and SEIR models to the data, running three MCMC chains per model, discarding the first 100 iterations as burn–in, and sampling the paths of 150 subjects, chosen uniformly at random, per MCMC iteration. After discarding the burn–in, the resulting samples were combined to form the final sample. We also attempted to fit the models using PMMH. We ran three chains per model, each using 2,500 particles, the paths for which were simulated approximately via $\tau$–leaping with a one day time step. The PMMH chains were plagued by severe particle degeneracy and did not converge.

**Epoch 1: Weeks 0 – 26**

| Param. | True value |
|---|---|
| $\beta$ | 0.00025 |
| $\gamma$ | 1/210 |
| $\mu$ | 1/150 |
| $\rho$ | 0.95 |
| $\mathbf{X}(t_0)$ | $S_0 = 397, E_0 = 2, I_0 = 1, R_0 = 0$ |

**Epoch 2: Weeks 26–105**

| | |
|---|---|
| $\beta$ | 0.0001 |
| $\gamma$ | 1/210 |
| $\mu$ | 1/330 |
| $\rho$ | 0.95 |
| $\mathbf{X}(t_{26})$ | $S_0 = 279, E_0 = 98, I_0 = 20, R_0 = 3$ |

**Epoch 3: Weeks 105–167**

| | |
|---|---|
| $\beta$ | 0.00035 |
| $\gamma$ | 1/90 |
| $\mu$ | 1/300 |
| $\rho$ | 0.95 |
| $\mathbf{X}(t_{105})$ | $S_0 = 1, E_0 = 43, I_0 = 145, R_0 = 211$ |

**Epoch 4: Weeks 167 – 209**

| | |
|---|---|
| $\beta$ | 0.0001 |
| $\gamma$ | 1/180 |
| $\mu$ | 1/70 |
| $\rho$ | 0.95 |
| $\mathbf{X}(t_{167})$ | $S_0 = 0, E_0 = 1, I_0 = 52, R_0 = 347$ |

**Table S10:** Parameter values governing the time–varying SEIR dynamics and binomial emissions process. The epidemic was simulated using Gillespie's direct algorithm and the process was restarted with the new parameter values at the beginning of each epoch.

## S10.2 Additional results

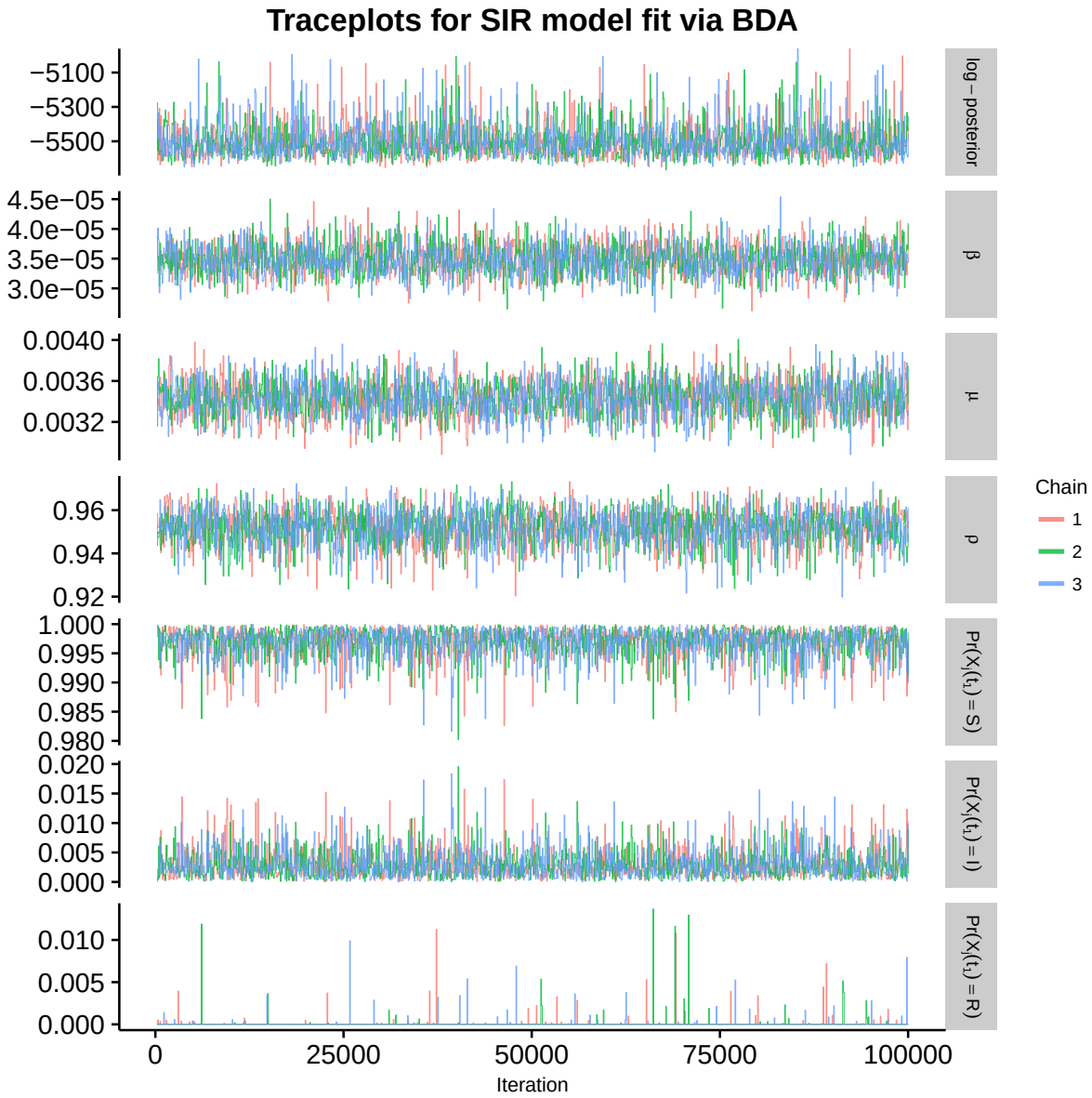


**Figure S13:** Traceplots of the log–posterior and model parameters for the SIR model fit using BDA following an initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

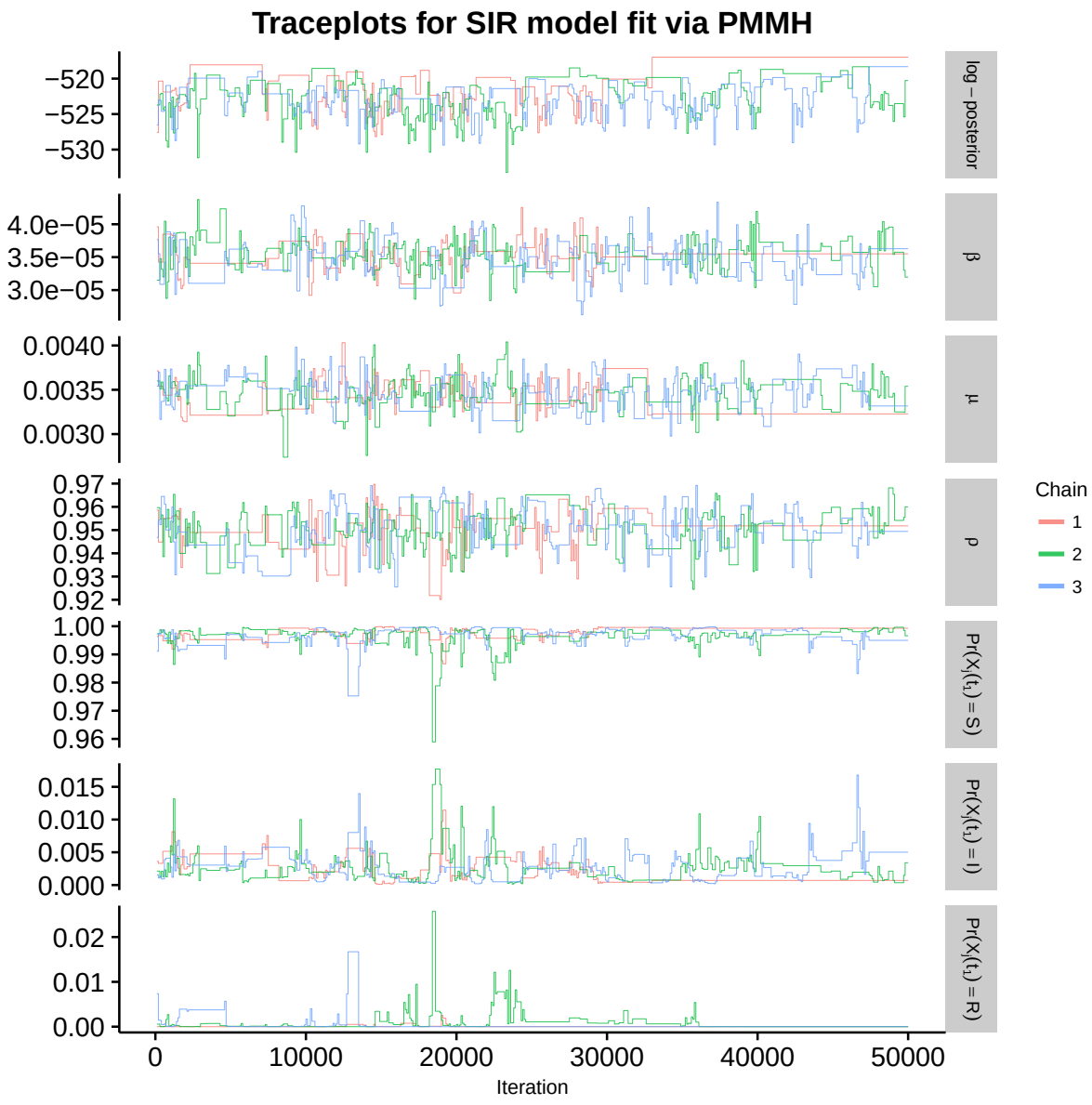


**Figure S14:** Traceplots of the log–posterior and model parameters for the SIR model fit using PMMH with 2,500 particles, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

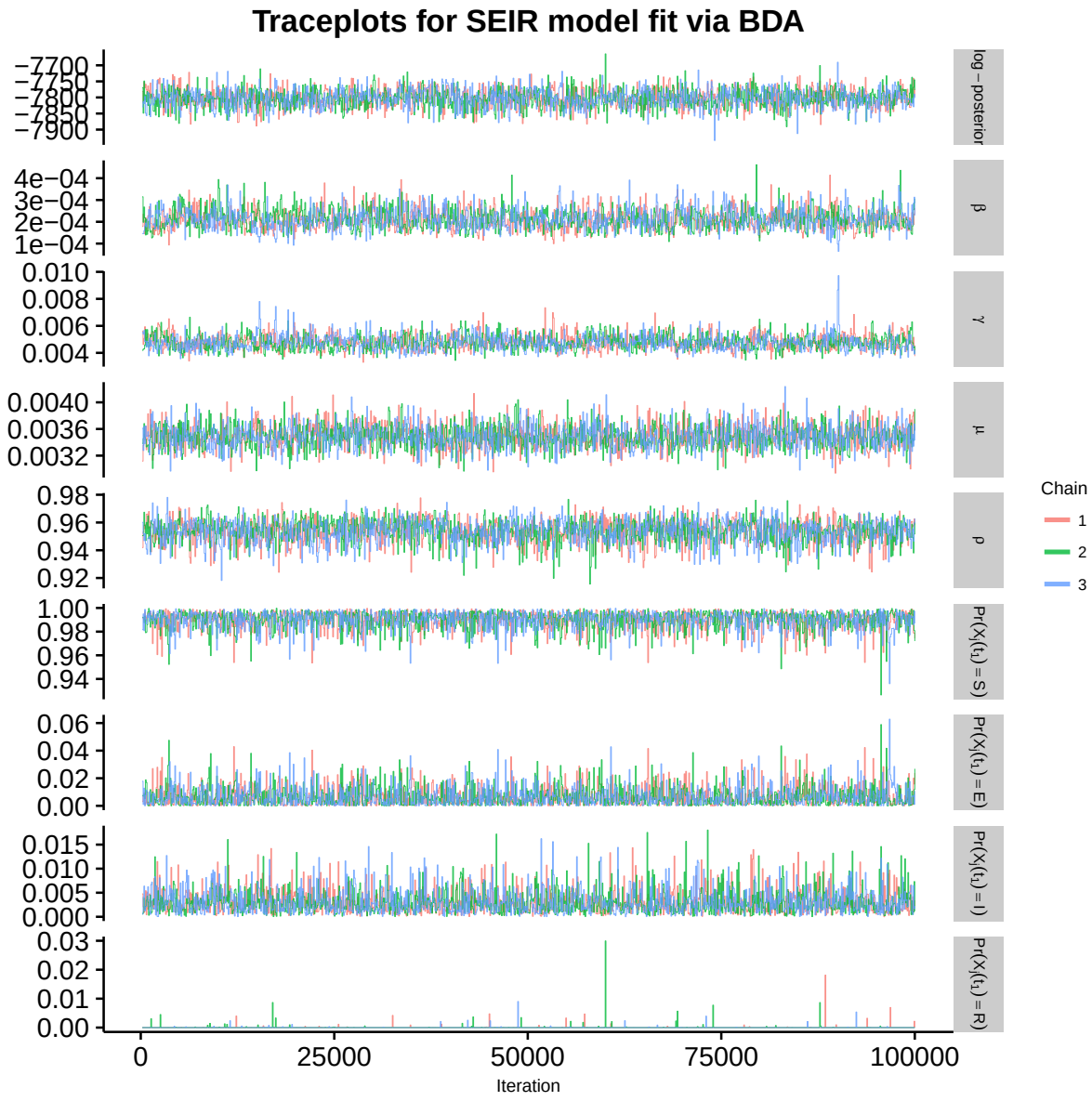


**Figure S15:** Traceplots of the log–posterior and model parameters for the SEIR model fit using BDA following an initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\gamma$ is the rate at which an exposed individual becomes infectious, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

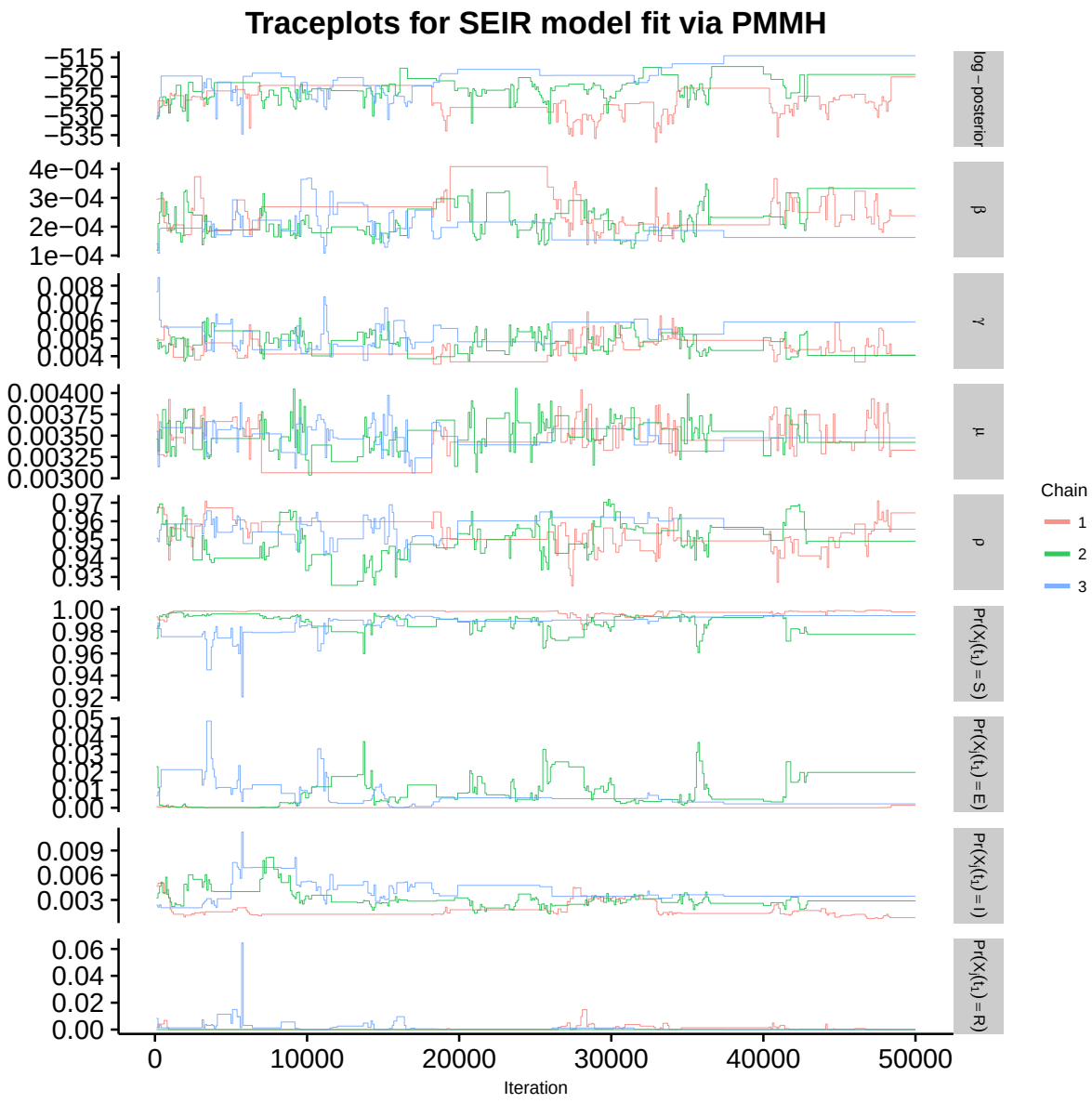


**Figure S16:** Traceplots of the log–posterior and model parameters for the SEIR model fit using PMMH with 2,500 particles, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH. $\beta$ denotes the per–contact infectivity rate, $\gamma$ is the rate at which an exposed individual becomes infectious, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

**SIR model**

| Parameter | Prior distribution |
|---|---|
| $R_0$ | Beta′(0.6, 0.7, 1, 4) |
| $\beta$ | Gamma(0.6, 10000) |
| $\mu$ | Gamma(0.7, 100) |
| $\mathbf{p}_{t_1}$ | Dirichlet(90, 0.5, 0.01) |
| $\rho$ | Beta(10, 1) |

**SEIR model**

| | |
|---|---|
| $R_0$ | Beta′(0.6, 0.7, 1, 4) |
| $\beta$ | Gamma(0.6, 10000) |
| $\gamma$ | Gamma(0.5, 100) |
| $\mu$ | Gamma(0.7, 100) |
| $\mathbf{p}_{t_1}$ | Dirichlet(90, 0.5, 0.5, 0.01) |
| $\rho$ | Beta(10, 1) |

**Table S11:** Prior distributions for the SIR and SEIR model and measurement process parameters for the models fit to the dataset simulated under time–varying SEIR dynamics. The prior for $R_0$ is the induced prior implied by $\beta$ and $\mu$. The per–contact infectivity rate is $\beta$, the rate at which an exposed individual becomes infectious is $\gamma$, the recovery rate is $\mu$, the binomial sampling probability is $\rho$, and the initial state probabilities are $\mathbf{p}_{t_1}$.

**SIR model**

| Parameter | Posterior median (95% Credible interval) |
|---|---|
| $R_0$ | 4.05 (3.40, 4.81) |
| $\beta$ | 0.000035 (0.000030, 0.000040) |
| $\mu$ | 0.0034 (0.0031, 0.0038) |
| $\rho$ | 0.95 (0.93, 0.97) |

**SEIR model**

| Parameter | Posterior median (95% Credible interval) |
|---|---|
| $R_0$ | 23.80 (15.10, 36.98) |
| $\beta$ | 0.00021 (0.00013, 0.00032) |
| $\gamma$ | 0.0047 (0.0038, 0.0061) |
| $\mu$ | 0.0035 (0.0032, 0.0038) |
| $\rho$ | 0.95 (0.94, 0.97) |

**Table S12:** Posterior median estimates and 95% credible intervals for SIR and SEIR model parameters fit under a binomial emission distribution to the epidemic simulated with time–varying SEIR dynamics.

# S11 Simulation 3 — Inference under Population Size Misspecification — Details

We simulated an outbreak under SIR dynamics, with $R_0 = \beta N/\mu = 3.5$, in a population of 1,250 individuals. Roughly 0.2% of the population was initially infected, and 95% were initially suscep-

tible. The mean infectious period was $1/\mu = 7$ days. Prevalence was observed at weekly intervals, with detection probability $\rho = 0.3$, over a one year period.

We ran three chains for 100,000 iterations each under the following assumed population sized: 150, 300, 500, 900, 1100, 1200, 1250, 1300, 1400. We sampled the paths for 10% of the subjects, chosen uniformly at random, per MCMC iteration. We discarded the first 500 iterations from each chain as burn-in. Diffuse priors were specified for all model parameters, with the prior for the per–contact infectivity rate depending on the assumed population size (summarized in Table S13).

| Param. | Prior distribution |
|---|---|
| $R_0$ | Beta′(0.00042 $\times \frac{1250}{N}$, 0.35, 1, 2 / N) |
| $\beta$ | Gamma(0.00042 $\times \frac{1250}{N}$, 1) |
| $\mu$ | Gamma(0.35, 2) |
| $\mathbf{p}_{t_1}$ | Dirichlet(100, 1, 5) |
| $\rho$ | Beta(1,1) |

**Table S13:** Prior distributions for SIR model and measurement process parameters. The prior for $R_0$ is the induced prior implied by $\beta$ and $\mu$. The per–contact infectivity rate is $\beta$, the recovery rate is $\mu$, the binomial sampling probability is $\rho$, and the initial state probabilities are $\mathbf{p}_{t_1}$. The prior for $\beta$ was scaled in accordance with the assumed population size.

# S12 Simulation 4 — Effect of Prior Specification on Inference — Setup and Additional Results

## S12.1 Simulation details

We ran three MCMC chains for each of the SIR models fit under the prior regimes that are specified in Table S14 along with the true parameter values under which the data were simulated. Each chain was run for 100,000 MCMC iterations with 75 subject–paths per iteration. The first 100 iterations of each were discarded as burn–in, after which the samples from all three chains for each model were combined to form the posterior sample.

| | Prior Distribution | | | |
|---|---|---|---|---|
| Parameter | Regime 1 | Regime 2 | Regime 3 | Regime 4 |
| $R_0 = 1.84$ | Beta$'$(3, 3, 1, 1.526) | Beta$'$(0.3, 0.1, 1, 0.6) | Beta$'$(3, 3, 1, 1.526) | Beta$'$(0.3, 0.1, 1, 0.6) |
| $\beta = 0.00035$ | Gamma(3, 10000) | Gamma(0.3, 1000) | Gamma(3, 10000) | Gamma(0.3, 1000) |
| $\mu = 0.14$ | Gamma(3, 20) | Gamma(0.1, 0.8) | Gamma(3, 20) | Gamma(0.1, 0.8) |
| $\rho = 0.2$ | Beta(21, 75) | Beta(21, 75) | Beta(1,1) | Beta(1,1) |

**Table S14:** True parameter values and prior distributions under four different prior regimes. The prior for $R_0$ is the implied prior induced by the priors for $\beta$ and $\mu$. In regimes one and three, the central 80% of the prior mass for $R_0$ lay between 1.25 and 4.56, while in regimes two and four, 80% of the prior mass lay between $3.8 \times 10^{-4}$ and $2.7 \times 10^{4}$. In regimes one and two, 80% of the prior mass for $\rho$ lay between 0.17 and 0.27, while in regimes three and four the prior mass for $\rho$ was uniformly distributed between 0 and 1. We used the same mildly informative Dirichlet(9, 0.2, 0.5) prior for $\mathbf{p}_{t_1}$ in all prior regimes.

## S12.2 Convergence diagnostics

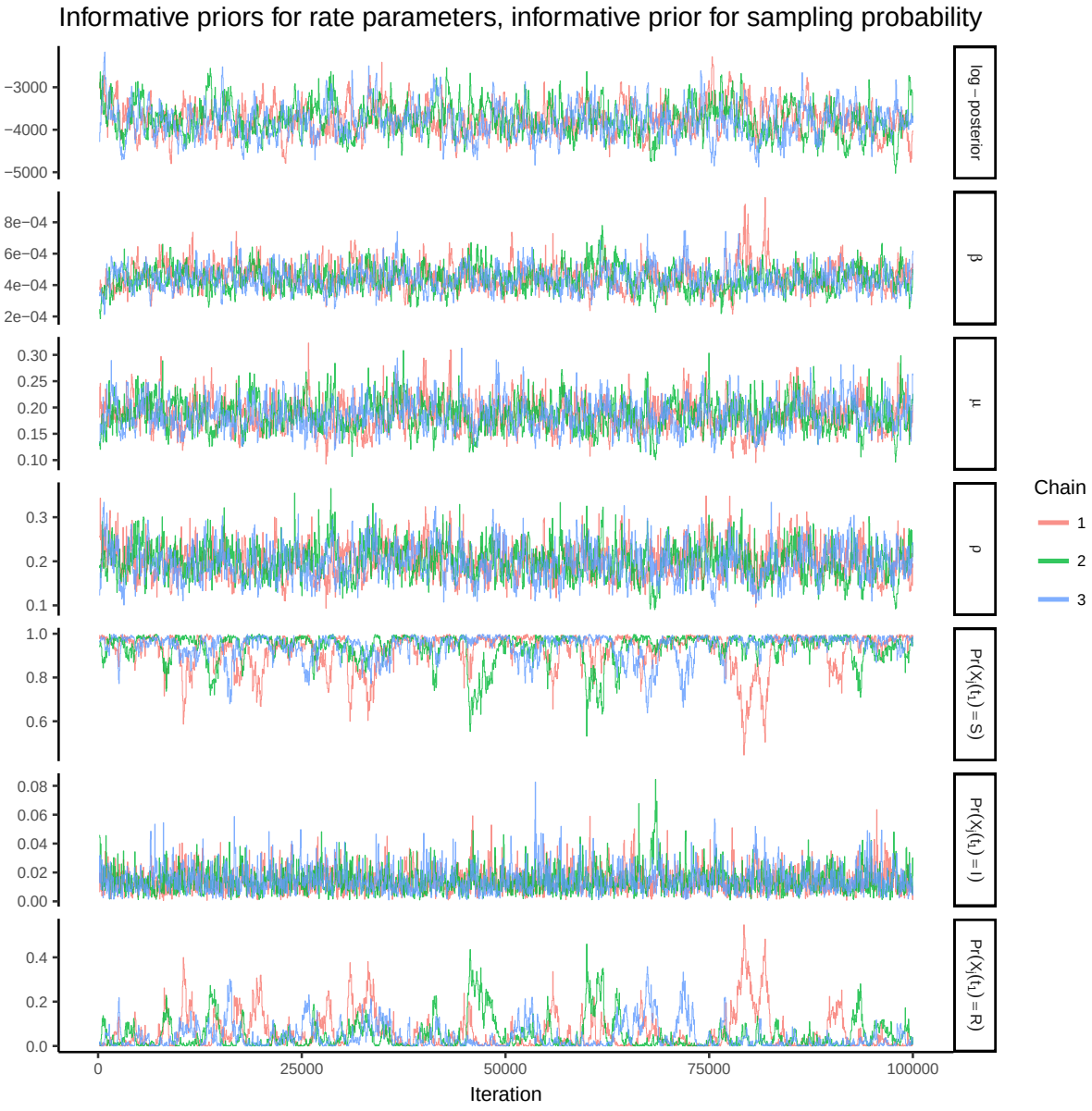


**Figure S17:** Traceplots of the log–posterior and model parameters for the SIR model fit under informative priors for all model parameters. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

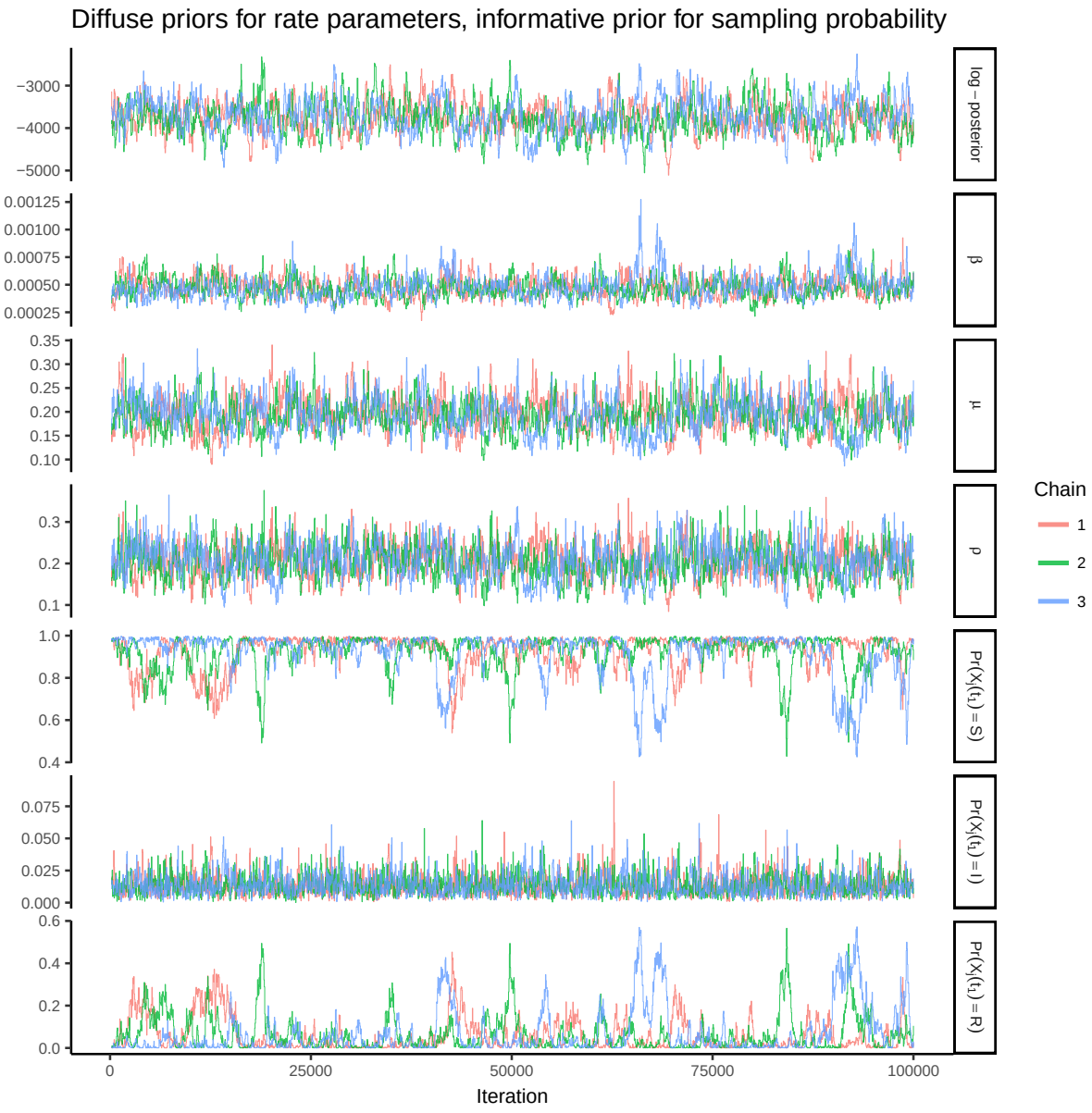


**Figure S18:** Traceplots of the log–posterior and model parameters for the SIR model fit under diffuse priors for the rate parameters and an informative prior for the binomial sampling probability. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

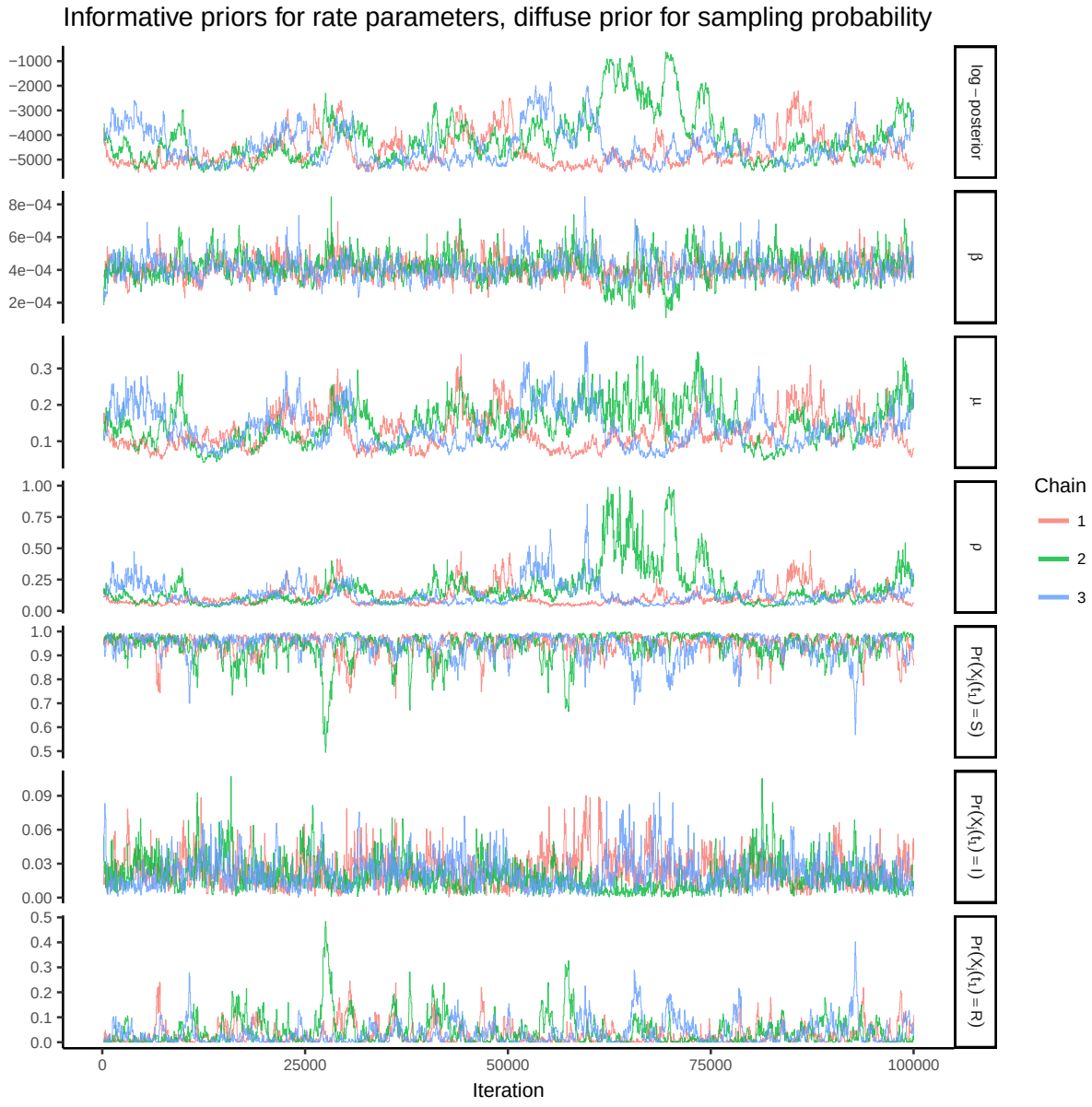


**Figure S19:** Traceplots of the log–posterior and model parameters for the SIR model fit under informative priors for the rate parameters and a diffuse prior for the binomial sampling probability. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

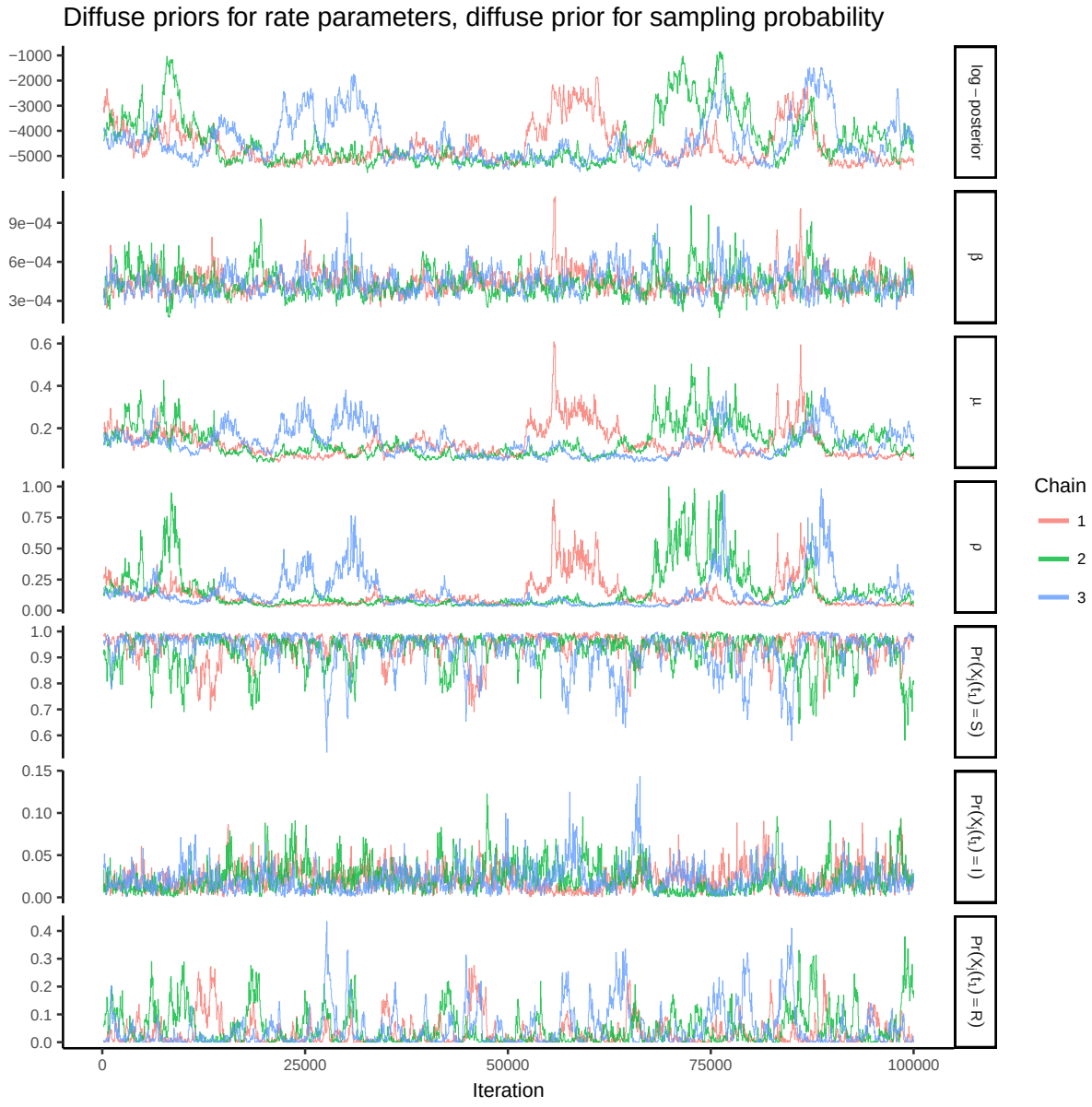


**Figure S20:** Traceplots of the log–posterior and model parameters for the SIR model fit under diffuse priors for all model parameters. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

# S13 Setup, additional results, and MCMC diagnostics for British boarding school example

We ran three MCMC chains per model to fit the SIR and SEIR models to the British boarding school dataset, for 100,000 iterations per chain. We sampled the paths for 100 subjects, chosen uniformly at random, per MCMC iteration, and discarded, as burn-in, the first 100 iterations of each chain for the SIR model, and the first 5,000 iterations of each chain for the SEIR model. Prior distributions, along with posterior medians and credible intervals are given in tables S15 and S16. The induced prior for $R_0$ is highly diffuse due to the diffuse prior on the per–contact infectivity rate. The prior distribution for the recovery rate and the rate at which exposed individuals became infectious reflected prior knowledge of the natural history of influenza. The prior for the detection probability has roughly 90% of its mass above 0.3, but is arguably quite diffuse given that it is known that over 90% of the boys were eventually infected.

We also fit the SIR and SEIR models using PMMH with paths for 5,000 particles simulated approximately via a multinomial modification of $\tau$–leaping over two hour increments. The same priors were used as for the chains fit using BDA. Parameters were updated via random walk Metropolis–Hastings on transformed scales with a proposal covariance matrix that was estimated from an initial run of 2,000 MCMC iterations. We applied a log transformation to the rate parameters, a logit transformation to the binomial sampling probability, and a generalized logit transformation to the initial state probabilities. Results for PMMH are not reported since the MCMC never converged (see traceplots below).

| Parameter | Prior Distribution | Posterior Median (95% BCI) |
|---|---|---|
| $R_0$ | Beta′(0.001, 1, 1, 1526) | 3.89 (3.40, 4.47) |
| $\beta$ | Gamma(0.001, 1) | 0.0024 (0.0021, 0.0026) |
| $\mu$ | Gamma(1,2) | 0.46 (0.42, 0.50) |
| $\rho$ | Beta(1,2) | 0.98 (0.92, 1.00) |
| $\Pr(X_j(t_1) = S)$ | | 0.99 (0.98, 0.99) |
| $\Pr(X_j(t_1) = I)$ | Dirichlet(900,3,9) | 0.003 (0.001, 0.007) |
| $\Pr(X_j(t_1) = R)$ | | 0.009 (0.004, 0.017) |

**Table S15:** Prior distributions and posterior estimates for parameters of the SIR model with binomial emissions fit to the British boarding school outbreak data. The per–contact infectivity rate is $\beta$, the recovery rate is $\mu$, and the binomial sampling probability is $\rho$. The prior for $R_0$ is the implied prior induced by the priors for $\beta$ and $\mu$. Effective sample size were $\beta$: 11,304; $\mu$: 16,238; $\rho$: 3,920; $p_{S_{t_1}}$: 26,989; $p_{I_{t_1}}$: 284,431; $p_{R_{t_1}}$: 22,761.

## S13.1 Boarding school example — MCMC diagnostics

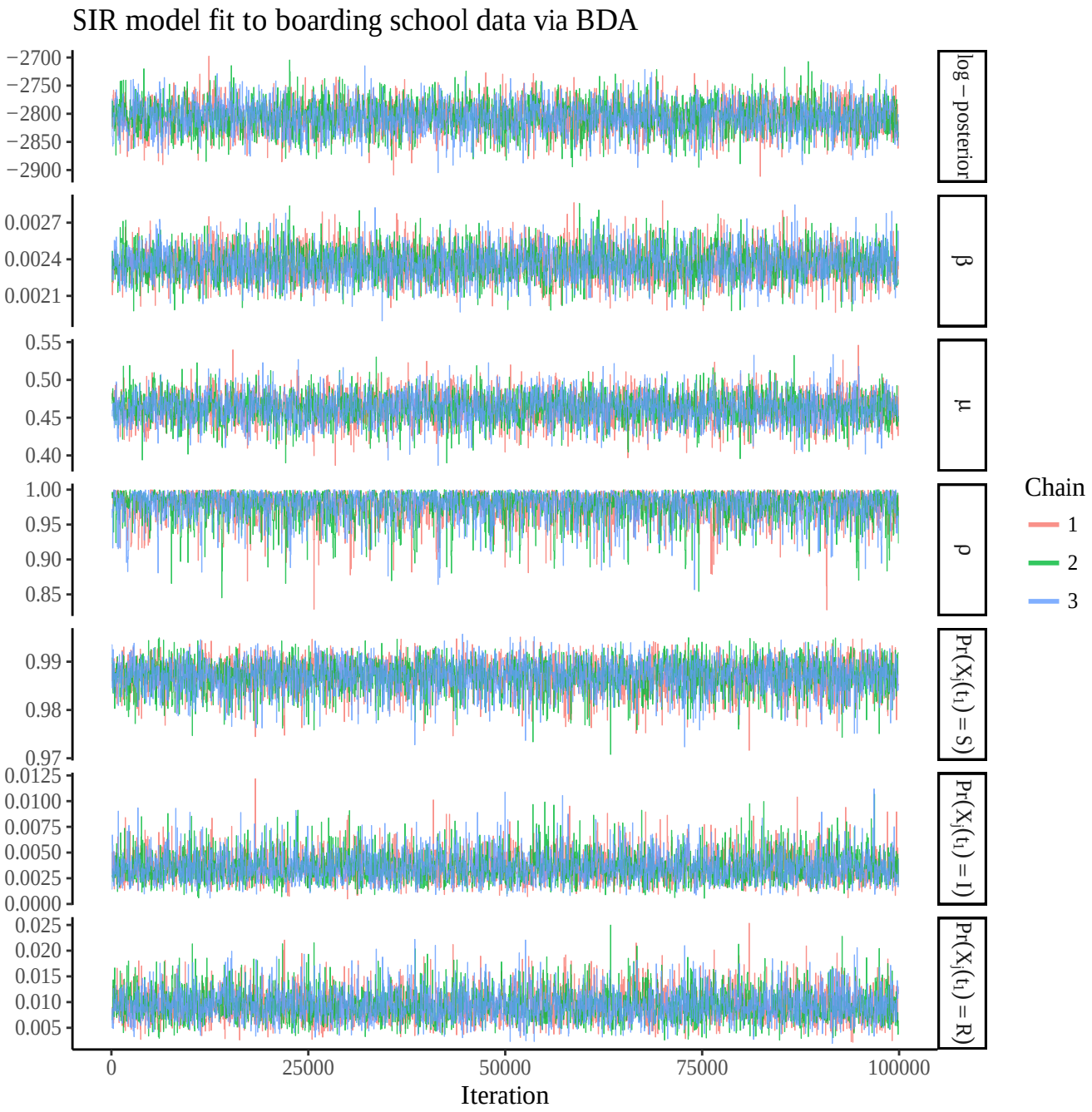


**Figure S21:** Traceplots of the log–posterior and model parameters for the SIR model fit under binomial emissions using BDA following an initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

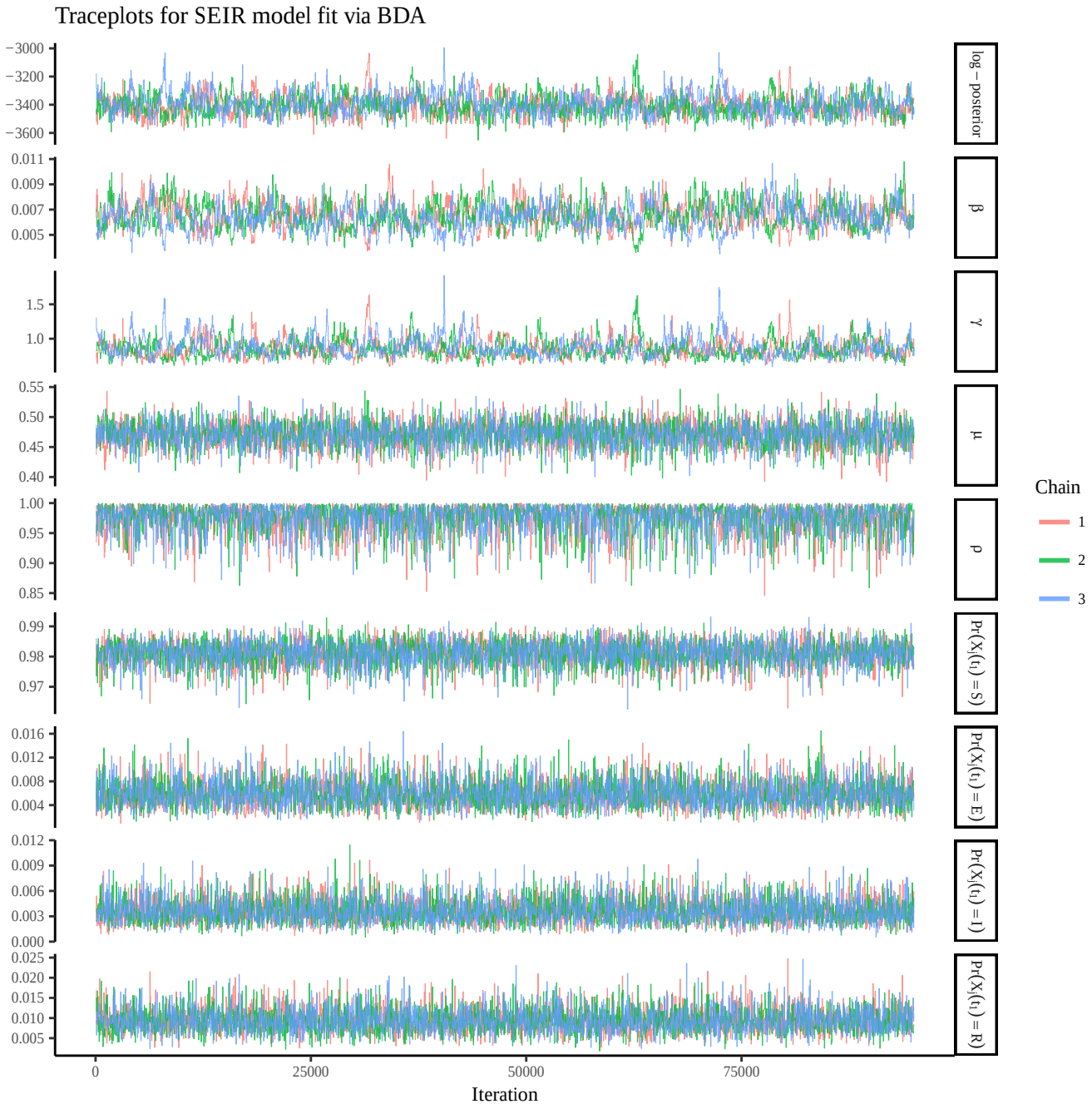


**Figure S22:** Traceplots of the log–posterior and model parameters for the SEIR model fit under binomial emissions via BDA following an initial burn–in of 5,000 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\gamma$ is the rate at which immunity is lost, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

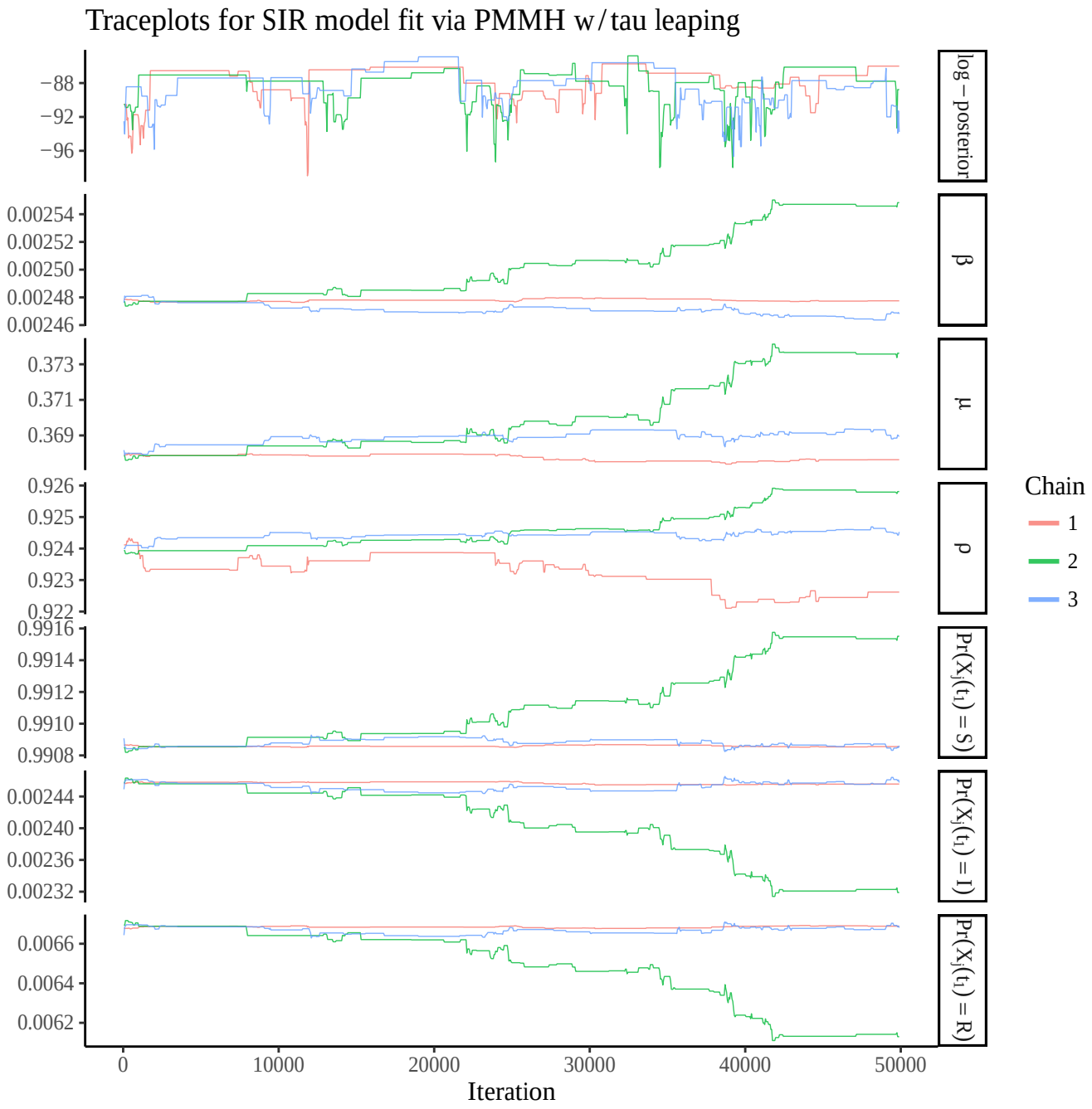


**Figure S23:** Traceplots of the log–posterior and model parameters for the SIR model fit under binomial emissions using PMMH with 5,000 particles per chain and a time–step of 2 hours in the approximate $\tau$–leaping algorithm, following a tuning run of 2,000 iterations to estimate the RWMH covariance matrix and in initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

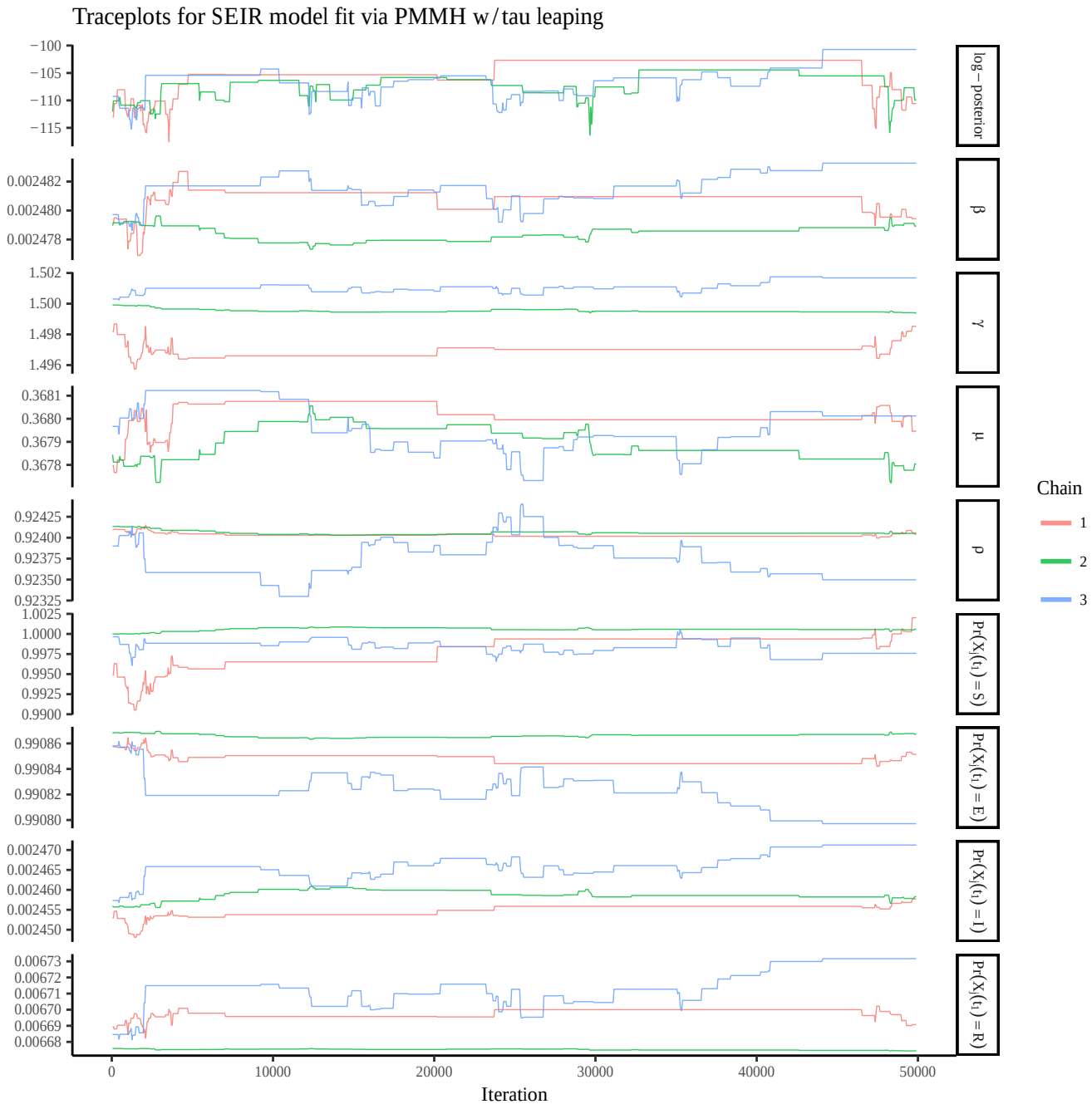


**Figure S24:** Traceplots of the log–posterior and model parameters for the SEIR model fit under binomial emissions using PMMH with 5,000 particles per chain and a time–step of 2 hours in the approximate $\tau$–leaping algorithm, following a tuning run of 2,000 iterations to estimate the RWMH covariance matrix and in initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\gamma$ is the rate at which immunity is lost, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

| Parameter | Prior Distribution | Posterior Median (95% BCI) |
|---|---|---|
| $R_0$ | Beta$'$(0.001, 1, 1, 1526) | 3.89 (3.40, 4.47) |
| $\beta$ | Gamma(0.001, 1) | 0.0064 (0.0046, 0.0086) |
| $\gamma$ | Gamma(0.001, 1) | 0.84 (0.66, 1.19) |
| $\mu$ | Gamma(1,2) | 0.47 (0.43, 0.51) |
| $\rho$ | Beta(1,2) | 0.98 (0.91, 1.00) |
| $\Pr(X_j(t_1) = S)$ | | 0.98 (0.97, 0.99) |
| $\Pr(X_j(t_1) = E)$ | Dirichlet(900, 6,3,9) | 0.006 (0.002, 0.01) |
| $\Pr(X_j(t_1) = I)$ | | 0.003 (0.001, 0.007) |
| $\Pr(X_j(t_1) = R)$ | | 0.009 (0.004, 0.016) |

**Table S16:** Prior distributions and posterior estimates for parameters of the SEIR model with binomial emissions fit to the British boarding school outbreak data. The per–contact infectivity rate is $\beta$, the rate at which an exposed individual becomes infectious is $\gamma$, the recovery rate is $\mu$, and the binomial sampling probability is $\rho$. The prior for $R_0$ is the implied prior induced by the priors for $\beta$ and $\mu$. Effective sample size were $\beta$: 679; $\gamma$: 658; $\mu$: 10,069; $\rho$: 3,244; $p_{S_{t_1}}$: 26,868; $p_{I_{t_1}}$: 26,168; $p_{R_{t_1}}$: 273,613.

**Table S17:**

## S13.2 Supplementary analysis of the British boarding school example under negative binomial emissions

The PMMH MCMC runs in which SIR and SEIR models were fit to the boarding school data under a binomial emission distribution were plagued by severe particle degeneracy (Figures S23 and S24). The binomial emission distribution requires that the latent prevalence always be at least as great as the observed prevalence. However, this seemed to be a very stringent criterion with such a high case detection rate. That this criterion was so stringent is suggestive of non–trivial model misspecification. We attempted to confirm this by simulating a dataset that resembled the boarding school data. One possible data generating mechanism that yielded to similar prevalence counts resulted from an outbreak evolving under SEIR dynamics that varied over three epochs (Figure S25). That even a model with this simple set of time–varying dynamics would undoubtedly still be misspecified with respect to the real world circumstances in the boarding school is suggestive of a non–trivial level of model misspecification for both the simple SIR and SEIR models that we attempted to fit. Still, the inability of PMMH to fit simple, easily interpretable, SEMs to this data under binomial emissions is a severe limitation.

We fit an alternative set of SIR and SEIR models to the data using BDA and PMMH in which the observed prevalence was modeled as a negative binomial sample of the true prevalence, parameterized by its mean and overdispersion. This is a somewhat unrealistic emission distribution because it allows for the observed prevalence to be greater than the true prevalence. That this tended to occur more often in the later parts of the epidemic when boys were being discharged from the infirmary was particularly odd. However, the negative binomial emission distribution allows us to avoid degeneracy in the collection of PMMH particles by doing away with the constraint that the latent prevalence be no smaller than the observed prevalence. Parameters were assigned the same priors

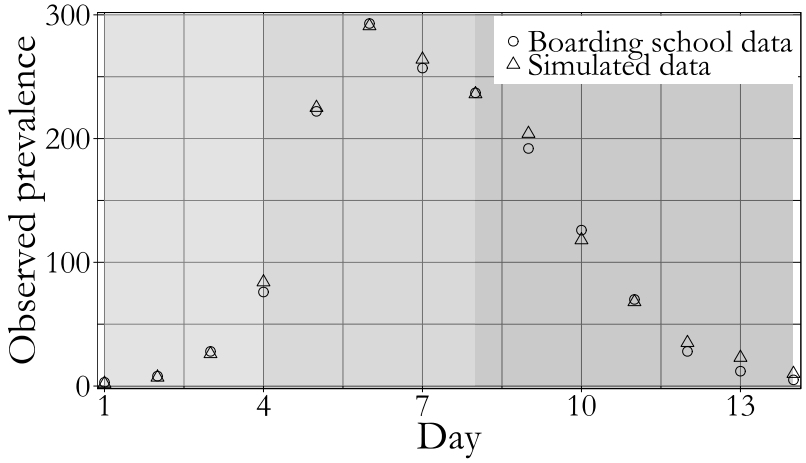


**Figure S25:** British boardings school data and data simulated under SEIR dynamics with time–varying dynamics over three epochs (indicated by different shaded regions). The simulated dataset was generated using the following parameters: in the first epoch (days 1-4), $\beta = 0.0035$, $\gamma = 1.25$, $\mu = 0.3$. In the second epoch (days 4-8), $\beta = 0.065$, $\gamma = 0.51$, $\mu = 0.41$. In the third epoch (days 8-14), $\beta = 0.06$, $\gamma = 2.5$, $\mu = 0.54$. The data were a binomial sample of the true prevalence with detection probability $\rho = 0.98$. There were three exposed individuals and two infected individuals at the beginning of day 1.

given in Tables S15 and S16, and the negative binomial overdispersion parameter, $\phi$, was assigned a Gamma(1, 0.1) prior parameterized by rate. When fitting the model with BDA, we sampled new values for the rate parameters and initial state probabilities from their univariate full conditional distributions via Gibbs sampling. New values for the negative binomial sampling probability and the overdispersion parameter were sampled using multivariate random walk Metropolis–Hastings on the logit scale for $\rho$ and on the log scale for $\phi$. An empirical covariance matrix for the RWMH was estimated from an initial run of 10,000 iterations and scaled until the acceptance rate was between 15%–50%. We ran three chains per model for 100,000 iterations each, updating the paths of 100 subjects per MCMC iteration, and discarding the first 10,000 iterations as burn–in. We also ran three chains for 50,000 iterations each using PMMH for each of the models, with 500 particles per chain for the SIR model and 5,000 particles per chain for the SEIR model. Particle paths were simulated approximately using $\tau$–leaping over a time step of 2 hours. Parameters were updated via multivariate RWMH whose covariance matrix was estimated from an initial tuning run of 2,000 iterations. Rate parameters and the overdispersion parameter were updated on the log scale, the negative binomial sampling probability was updated on the logit scale, and the initial state probabilities were updated on the generalized logit scale. We discarded the first 1,000 of each PMMH chain as burn–in.

Although the posterior median estimates under binomial and negative binomial emissions for the SIR and SEIR dynamics and detection rate are generally quite similar, the posterior credible intervals are considerably wider when the data modeled as a negative binomial sample of the true prevalence. This manifests both in the widths of the credible intervals for the latent process (Figure S26), and the credible intervals for the model parameters (Figure S27). This is not unexpected given that the negative binomial distribution is substantially more flexible than the binomial distribution. In comparing the posterior estimates obtained using BDA and PMMH under negative binomial emissions, we find that the estimates are essentially identical for the SIR model. For the SEIR

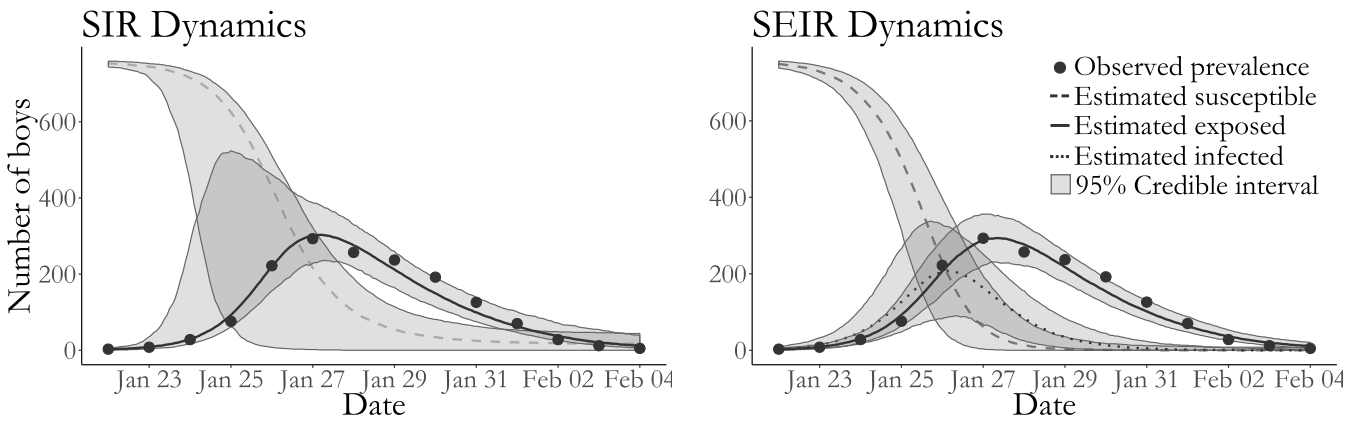


**Figure S26:** Boarding school data, pointwise posterior median estimates and pointwise 95% credible intervals under negative binomial emissions (grey shaded areas) for the numbers of infected boys (solid line) and susceptible boys (dashed line). Posterior estimates based on a thinned sample, with every $250^{th}$ configuration retained.

model, estimates of the dynamics are generally similar, though not to the same degree as those for the SIR model. We notice that the credible intervals for the mean infectious period and the negative binomial detection probability obtained using PMMH are substantially wider than those obtained using BDA. Upon closer inspection of the traceplots of the model parameters, it is clear that the negative binomial overdispersion parameter in the PMMH chains did not converge (Figure S29).

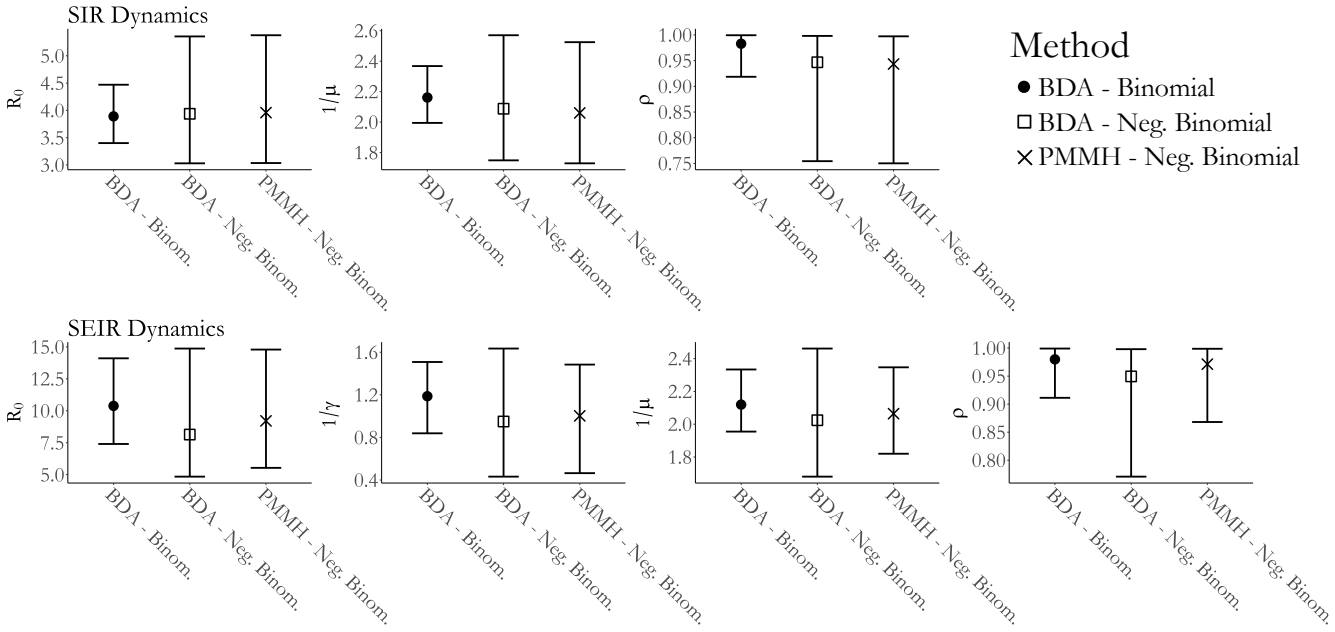


**Figure S27:** Posterior medians and 95% credible intervals for SIR and SEIR models fit with BDA and PMMH to the British boarding school data under binomial and negative binomial emission distributions.

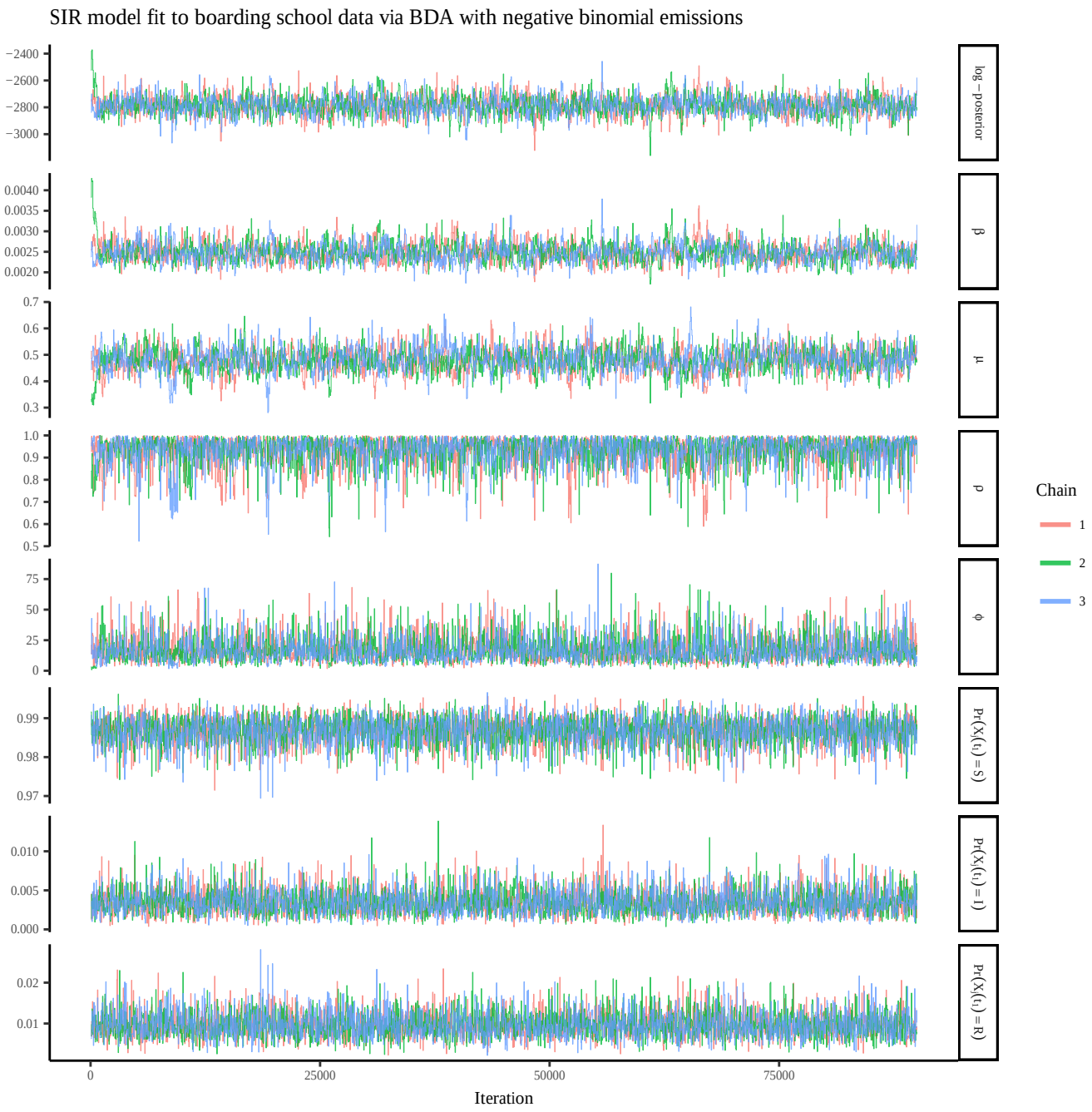


**Figure S28:** Traceplots of the log–posterior and model parameters for the SIR model fit under negative binomial emissions using BDA following an initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

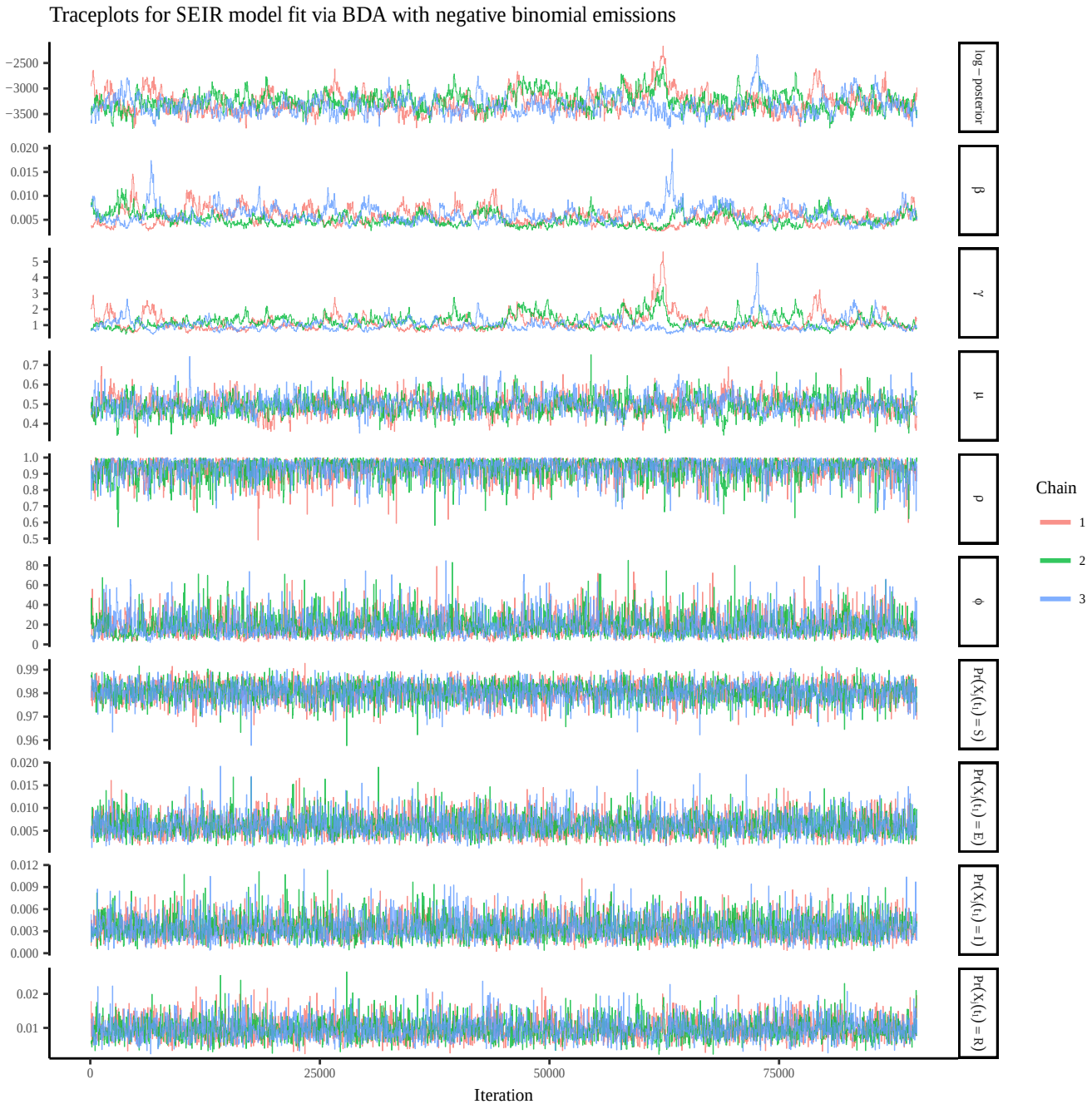


**Figure S29:** Traceplots of the log–posterior and model parameters for the SEIR model fit under negative binomial emissions via BDA following an initial burn–in of 5,000 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\gamma$ is the rate at which immunity is lost, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

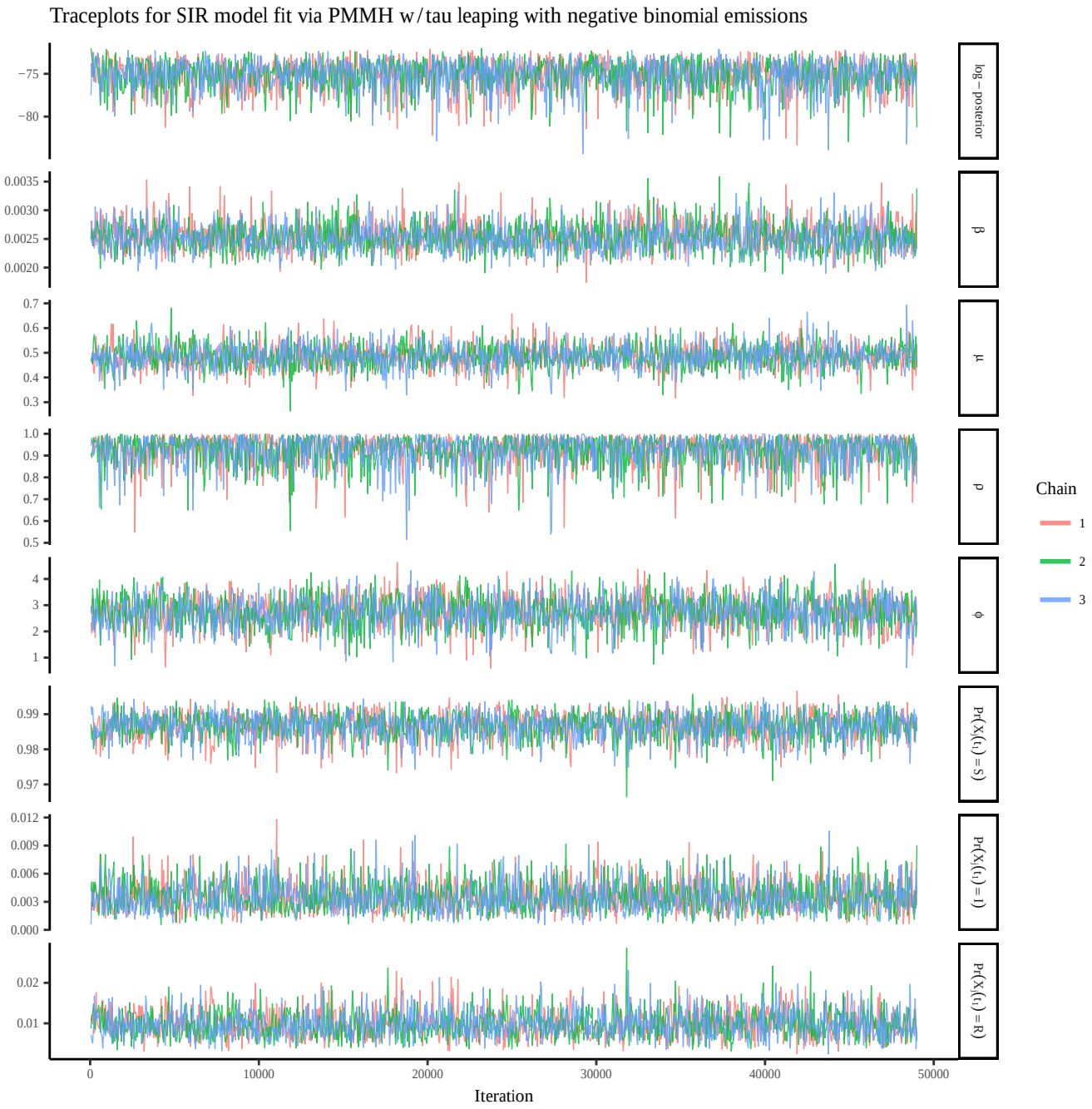


**Figure S30:** Traceplots of the log–posterior and model parameters for the SIR model fit under negative binomial emissions using PMMH with 5,000 particles per chain and a time–step of 2 hours in the approximate $\tau$–leaping algorithm, following a tuning run of 2,000 iterations to estimate the RWMH covariance matrix and in initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

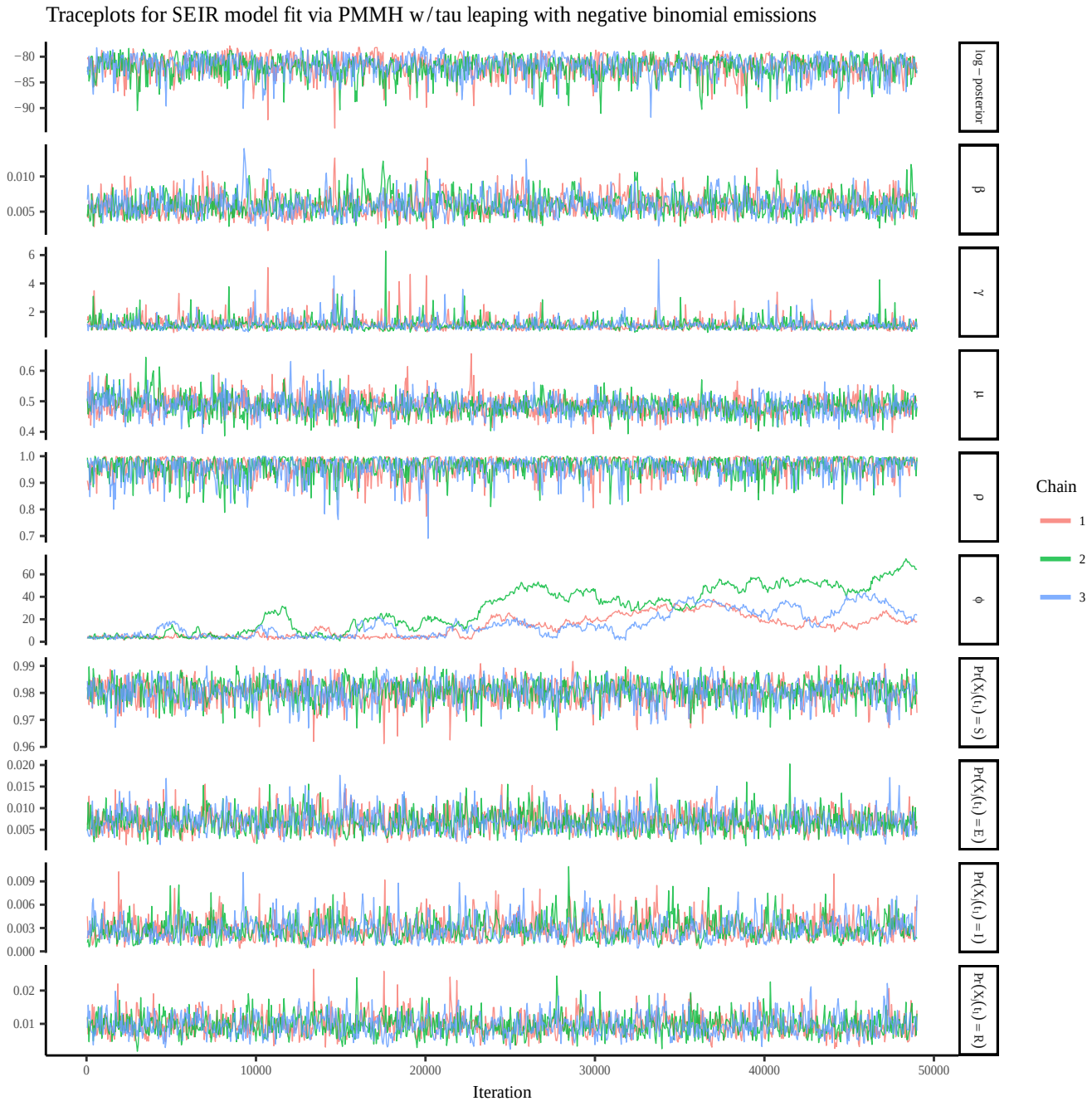


**Figure S31:** Traceplots of the log–posterior and model parameters for the SEIR model fit under negative binomial emissions using PMMH with 5,000 particles per chain and a time–step of 2 hours in the approximate $\tau$–leaping algorithm, following a tuning run of 2,000 iterations to estimate the RWMH covariance matrix and in initial burn–in of 100 iterations. $\beta$ denotes the per–contact infectivity rate, $\mu$ is the recovery rate, $\gamma$ is the rate at which immunity is lost, and $\rho$ is the binomial sampling probability. Traceplots are thinned to display every $50^{\text{th}}$ iteration.

# References


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

J. Fintzi. *ECctmc: Simulation from Endpoint-Conditioned Continuous Time Markov Chains*, 2016. URL `https://github.com/fintzij/ECctmc`. R package, version 0.2.2.

M.W. Hirsch, S. Smale, and R.L. Devaney. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press. Academic Press, Waltham, 2013.

A. Hobolth and E.A. Stone. Simulation from endpoint-conditioned, continuous-time Markov chains on a finite state space, with applications to molecular evolution. *The Annals of Applied Statistics*, 3:1204–1231, 2009.

W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 115, pages 700–721. The Royal Society, 1927.

C. Moler and C. Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45:3–49, 2003.

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

S.L. Scott. Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association*, 97:337–351, 2002.

J.P. Tian and D. Kannan. Lumpability and commutativity of Markov processes. *Stochastic Analysis and Applications*, 24:685–702, 2006.

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