

©Copyright 2018

Jonathan Fintzi

Bayesian Modeling of Partially Observed Epidemic Count Data

Jonathan Fintzi

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

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Vladimir Minin, Chair

Jon Wakefield, Chair

M. Elizabeth Halloran

James Hughes

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Bayesian Modeling of Partially Observed Epidemic Count Data

Jonathan Fintzi

Co-Chairs of the Supervisory Committee:

Co-chair Vladimir Minin

Co-chair Jon Wakefield

Epidemic count data reported by public health surveillance systems reflect the incidence or prevalence of an infectious agent as it spreads through a population. They are a primary source of information for shaping response strategies and predicting how the outbreak will evolve. Incidence and prevalence counts are often the only source of information about historical outbreaks, or outbreaks in resource limited settings, which are of interest for researchers seeking to develop an understanding of disease transmission during “peace time”, with an eye on preparing for future outbreaks. The absence of subject-level information and the systematic underreporting of cases complicate the task of disentangling whether the data arose from a severe outbreak, observed with low fidelity, or a mild outbreak were most cases were detected. The magnitude of the missing data and the high dimensional state space of the latent epidemic process present challenges for fitting epidemic models that appropriately quantify the stochastic aspects of the transmission dynamics. In this dissertation, we develop computational algorithms for fitting stochastic epidemic models to partially observed incidence and prevalence data. Our algorithms are not specific to particular model dynamics, but rather apply to a broad class of commonly used stochastic epidemic models, including models that allow for time-inhomogeneous transmission dynamics. We use our methods to analyze data from an outbreak of influenza in a British boarding school, the 2014–2015 outbreak of Ebola in West Africa, and the 2009–2011 A(H1N1) influenza pandemic in Finland.

TABLE OF CONTENTS

	Page
List of Figures	vi
List of Tables	xiii
Glossary	xvii
Chapter 1: Introduction	1
1.1 Partially Observed Epidemic Count Data	1
1.2 Compartmental Epidemic Models	1
1.3 Motivating Examples	4
1.3.1 Influenza in a British Boarding School	4
1.3.2 Epidemics in Large Populations	5
1.3.3 Pandemic A(H1N1) Influenza in Finland	7
Chapter 2: Technical Background	10
2.1 Mathematical Models for the Spread of Infectious Diseases	10
2.1.1 An Agent-Based Susceptible–Infected–Recovered Model	11
2.1.2 A Population–Level Susceptible–Infected–Recovered Model	14
2.1.3 A Brief Review of CTMCs	18
2.1.4 Large Population Approximations	20
2.1.5 Inference and Computation for Stochastic Epidemic Models	28
2.2 Bayesian Computation and Markov Chain Monte Carlo	31
2.2.1 Markov Chain Monte Carlo	32
2.2.2 Adaptive MCMC	42
Chapter 3: Agent–Based Data Augmentation for Fitting Stochastic Epidemic Models to Prevalence Data	45

3.1	Overview	45
3.2	The Data Augmentation Algorithm for an SIR Model	46
3.2.1	Measurement process and data	46
3.2.2	Latent epidemic process	46
3.2.3	Subject-path proposal framework	48
3.2.4	Parameter Updates and MCMC Scan Order	57
3.2.5	Data augmentation for SEIR dynamics	58
3.2.6	Data augmentation for SIRS dynamics	59
3.3	Simulation results	60
3.3.1	Inference Under a Variety of Epidemic Dynamics	60
3.3.2	Inference under model misspecification	63
3.3.3	Effect of prior specification on posterior inference	66
3.4	Example: Influenza in a British boarding school	69
3.5	Discussion	72
 Chapter 4:	Approximate Inference for Stochastic Epidemic Models of Outbreaks in Large Populations	74
4.1	Overview	74
4.2	Fitting Stochastic Epidemic Models via the Linear Noise Approximation	75
4.2.1	Measurement Process and Data	76
4.2.2	Latent Epidemic Process	77
4.2.3	Tractable Approximations for Intractable Likelihoods	78
4.2.4	Diffusion Approximation	78
4.2.5	Linear Noise Approximation	81
4.2.6	Inference via the Linear Noise Approximation	83
4.2.7	Implementation	90
4.3	Simulations	91
4.3.1	Motivating Use of the LNA — Comparison with Common SEM Approximations	91
4.3.2	Assessing Model Fit	96
4.3.3	A Stratified SEIR Model for a Simulated Outbreak	103
4.4	Application: Modeling the Spread of Ebola	112
4.4.1	Country-Specific and Joint Models for Ebola in West Africa	113

4.4.2	Results	116
4.5	Discussion	126
Chapter 5:	Models with Time-Varying Dynamics for Pandemic A(H1N1) Influenza in Finland	128
5.1	Overview	128
5.2	On the Importance of Allowing for Time-Varying Dynamics	129
5.3	Modeling the Spread of A(H1N1)pdm09 in Finland	133
5.3.1	Data and Vaccination	133
5.3.2	Model Structure	133
5.3.3	Flexible Models for the Force of Infection with Gaussian Markov Ran- dom Fields	140
5.3.4	Computation and Inference	145
5.4	Results	146
5.4.1	Incidence	146
5.4.2	Transmission Dynamics and Vaccination	146
5.4.3	Comparison with Piecewise Homogeneous Dynamics	155
5.5	Discussion	158
Chapter 6:	Discussion and Future Work	162
Bibliography	165
Appendix A:	Appendix to Chapter 3	190
A.1	Computing the matrix exponential	190
A.1.1	Case 1: Λ has real eigenvalues	190
A.1.2	Case 2: Λ has complex eigenvalues	191
A.2	Conjugate priors for SIR, SEIR, and SIRS model parameters	191
A.3	Simulation 1 — Inference Under Various Epidemic Dynamics — Setup and Additional Results	195
A.3.1	Simulation details for the SIR model	195
A.3.2	Additional results and MCMC diagnostics for the SIR model	197
A.3.3	Simulation details for the SEIR model	201
A.3.4	Additional results and MCMC diagnostics for the SEIR model	202

A.3.5	Simulation details for the SIRS model	206
A.3.6	Additional results and MCMC diagnostics for the SIRS model	207
A.3.7	Estimated latent posterior distributions for all models	207
A.4	Simulation 2 — Inference under Model	
	Misspecification — Setup and Additional Results	214
	A.4.1 Simulation setup	214
	A.4.2 Additional results	217
A.5	Simulation 3 — Inference under Population Size	
	Misspecification — Details	222
A.6	Simulation 4 — Effect of Prior Specification on Inference — Setup and Additional Results	223
	A.6.1 Simulation details	223
	A.6.2 Convergence diagnostics	223
A.7	Setup, additional results, and MCMC diagnostics for British boarding school example	228
	A.7.1 Boarding school example — MCMC diagnostics	228
	A.7.2 Supplementary analysis of the British boarding school example under negative binomial emissions	233
Appendix B:	Appendix to Chapter 4	243
B.1	Tuning the Initial Elliptical Slice Sampling Bracket Width	243
B.2	Inference for Initial Compartment Volumes	245
B.3	Choice of Estimation Scale and Implications for Mixing and Convergence . .	247
B.4	Identifiability when Estimating the Effective Population Size	253
B.5	Simulation Details and Additional Results for Section 4.3.1	256
	B.5.1 Simulation Setup and MCMC Details	256
	B.5.2 Additional Coverage Simulation Results	258
B.6	Supplementary Coverage Simulations with Fixed Parameters	262
	B.6.1 Simulation Setup	262
	B.6.2 Fixed Parameter Coverage Results	263
	B.6.3 Fixed Parameter Coverage Simulation Results	265
B.7	Synthetic Ebola Outbreak — Simulation Settings and Additional Results . .	270
	B.7.1 Simulation Details	270
	B.7.2 MCMC Details	273

B.7.3	Additional Results	276
B.8	Modeling Ebola in West Africa — MCMC Details and Supplementary Results	283
	B.8.1 Details and Supplementary Results for Single-Country Models	283
	B.8.2 Details and Supplementary Results for Multi-Country Models	309
Appendix C:	Appendix to Chapter 5	331
C.1	Simulation Details for Results in Section 5.2	331
	C.1.1 Simulation setup	331
	C.1.2 Additional results	333
C.2	Specification of Contact Rates	338
C.3	Mapping Standard Normal Draws to LNA Sample Paths with Forcings . . .	339
C.4	Joint Sampling of LNA Paths and Gaussian Markov Random Fields	340
C.5	Modeling Pandemic A(H1N1) Influenza in Finland — Additional Details and Supplementary Results	342
	C.5.1 MCMC Details	342
	C.5.2 Prior Specification	343
	C.5.3 Additional Results	349
	C.5.4 Sensitivity to Distribution of Initial Numbers of Susceptibles	356

LIST OF FIGURES

Figure Number	Page
1.1 Data from an outbreak of influenza in a British boarding school.	4
1.2 Weekly incidence of confirmed and probable cases of Ebola in Guinea, Liberia, and Sierra Leone.	5
1.3 A(H1N1)pdm09 incidence and vaccination data from Finland, April 15, 2009 — June 5, 2011.	9
2.1 Diagram of a Hidden Markov model.	11
2.2 Diagram of subject-level SIR paths.	12
2.3 Individual and lumped representations of SIR dynamics.	16
2.4 Relationship between various compartmental model representations.	21
3.1 Diagram of subject-level state transitions and HMM structure for an SIR model.	48
3.2 Diagram of subject-path proposals used in fitting models via Bayesian data augmentation.	52
3.3 Estimated latent posteriors for SIR, SEIR, and SIRS models fit to simulated prevalence data.	62
3.4 Posterior estimates of SIR, SEIR, and SIRS model parameters fit to simulated data using Bayesian data augmentation and PMMH.	63
3.5 Simulated outbreak data from an SEIR model with time-varying dynamics.	65
3.6 Latent posterior distributions for SIR and SEIR models fit to simulated data from an SEIR model with time-varying dynamics.	66
3.7 Posterior estimates of SIR model parameters under four prior regimes.	67
3.8 Estimated latent posterior for an SIR model under four prior regimes.	68
3.9 Boarding school data and latent posterior under SIR and SEIR dynamics.	69
3.10 Posterior estimates of SIR and SEIR model parameters fit to British boarding school outbreak data.	71
4.1 Traceplots for an MCMC chain using a centered LNA parameterization.	84
4.2 Traceplots for an MCMC chain using a non-centered LNA parameterization.	86

4.3	Diagram of centered and non-centered parameterizations of an LNA path.	87
4.4	Coverage simulation results for SIR models fit via the LNA, ODE, and MMTL approximations.	94
4.5	SIR model posterior incidence, posterior mean observed incidence, partial posterior predictive, and full posterior distributions.	98
4.6	Model comparison with posterior predictive p-values and relative predictive interval widths.	99
4.7	Posterior distributions of LNA draws for an SIR model.	101
4.8	Posterior predictive distributions of partial autocorrelations at various lags.	102
4.9	Diagram of a stratified SEIR model with country specific outbreak dynamics and cross-country virtual transmission.	105
4.10	Posterior distributions of stratified SEIR model parameters for a simulated Ebola outbreak in three countries.	108
4.11	Posterior mean observed incidence, partial posterior predictive, and full posterior distributions for a stratified SEIR model fit to a simulated Ebola outbreak.	109
4.12	Posterior predictive distributions of partial autocorrelations for a stratified SEIR model fit to a simulated Ebola outbreak.	110
4.13	Posterior distributions of LNA draws for a stratified SEIR model fit to a simulated Ebola outbreak.	111
4.14	Weekly incidence of confirmed and probable cases of Ebola in Guinea, Liberia, and Sierra Leone.	112
4.15	Diagram of single country SEIR models for the Ebola outbreak in West Africa.	114
4.16	Posterior predictive distributions for country-specific SEIR LNA models for the West Africa Ebola outbreak.	122
4.17	Posterior predictive distributions for a stratified SEIR LNA model for the West Africa Ebola outbreak.	122
4.18	Posterior predictive distributions for country-specific SEIR ODE models for the West Africa Ebola outbreak.	123
4.19	Posterior predictive distributions for a stratified SEIR ODE model for the West Africa Ebola outbreak.	123
4.20	Comparison of country-specific and joint SEIR Ebola models with posterior predictive p-values and relative predictive interval widths.	124
4.21	Comparison of SEIR Ebola models fit via the LNA and ODE with posterior predictive p-values and relative predictive interval widths.	125

5.1	Posterior distributions of SIRS model parameters fit to data from an outbreak with time-varying dynamics.	130
5.2	Time-varying reproduction numbers, latent incidence, and posterior predictive distributions for SIRS models fit to data from an outbreak with time-varying dynamics.	131
5.3	Comparison with posterior predictive p-values and relative predictive interval widths for SIRS models fit to an outbreak with time-varying dynamics.	132
5.4	A(H1N1)pdm09 incidence and vaccination data from Finland, April 15, 2009 — June 5, 2011.	134
5.5	Diagram of state transitions for an age-vaccination stratified SIRS model for influenza.	137
5.6	Estimated A(H1N1) incidence, reproduction numbers, and rates of exogenous infection under time-varying dynamics.	149
5.7	Posterior predictive distributions of a stratified SIRS ODE model with time-varying dynamics for A(H1N1)pdm09 in Finland.	155
5.8	Estimated A(H1N1) incidence, reproduction numbers, and rates of exogenous infection under piecewise homogeneous dynamics.	156
5.9	Posterior predictive distributions of a stratified SIRS ODE model with piecewise homogeneous dynamics over three epochs for A(H1N1)pdm09 in Finland.	157
5.10	Comparison with posterior predictive p-values and relative predictive interval widths for SIRS models with time-varying and piecewise homogeneous dynamics.	158
A.1	Simulation 1 MCMC traceplots for an SIR model fit using Bayesian data augmentation.	198
A.2	Simulation 1 MCMC traceplots for an SIR model fit using PMMH with approximate particle paths.	199
A.3	Simulation 1 MCMC traceplots for an SIR model fit using PMMH with exact particle paths.	200
A.4	Simulation 1 MCMC traceplots for an SEIR model fit using Bayesian data augmentation.	203
A.5	Simulation 1 MCMC traceplots for an SEIR model fit using PMMH with approximate particle paths.	204
A.6	Simulation 1 MCMC traceplots for an SIR model fit using PMMH with exact particle paths.	205
A.7	Simulation 1 MCMC traceplots for an SIR model fit using Bayesian data augmentation.	209

A.8	Simulation 1 MCMC traceplots for an SIRS model fit using PMMH with approximate particle paths with 500 particles per chain.	210
A.9	Simulation 1 MCMC traceplots for an SIRS model fit using PMMH with exact particle paths with 500 particles per chain.	211
A.10	Simulation 1 MCMC traceplots for an SIRS model fit using PMMH with exact particle paths with 200 particles per chain.	212
A.11	Simulation 1 latent posterior distributions for all models.	213
A.12	Simulation 2 MCMC traceplots for SIR model parameters fit using Bayesian data augmentation.	218
A.13	Simulation 2 MCMC traceplots for SIR model parameters fit using PMMH. .	219
A.14	Simulation 2 MCMC traceplots for SEIR model parameters fit using Bayesian data augmentation.	220
A.15	Simulation 2 MCMC traceplots for SEIR model parameters fit using PMMH. .	221
A.16	Simuation 4 MCMC traceplots for SIR model parameters fit under informative priors for all parameters.	224
A.17	Simuation 4 MCMC traceplots for SIR model parameters fit under informative priors for the measurement process and diffuse priors for rate parameters. .	225
A.18	Simuation 4 MCMC traceplots for SIR model parameters fit under diffuse priors for the measurement process and informative priors for rate parameters.	226
A.19	Simuation 4 MCMC traceplots for SIR model parameters fit under diffuse priors for the measurement process and diffuse priors for rate parameters. .	227
A.20	Traceplots for SIR model parameters fit to the boarding school data using Bayesian data augmentation.	229
A.21	Traceplots for SEIR model parameters fit to the boarding school data using Bayesian data augmentation.	230
A.22	Traceplots for SIR model parameters fit to the boarding school data using PMMH.	231
A.23	Traceplots for SEIR model parameters fit to the boarding school data using PMMH.	232
A.24	Comparison of boarding school outbreak data and data simulated using an SEIR model with time-varying dynamics.	235
A.25	Latent posterior estimates of prevalence in a boarding school under negative binomial emissions.	237
A.26	Posterior estimates of SIR and SEIR model parameters for models fit under negative binomial emissions.	238

A.27	Traceplots of SIR model parameters fit to boarding school data using Bayesian data augmentation.	239
A.28	Traceplots of SEIR model parameters fit to boarding school data using Bayesian data augmentation.	240
A.29	Traceplots of SIR model parameters fit to boarding school data using PMMH. .	241
A.30	Traceplots of SEIR model parameters fit to boarding school data using PMMH. .	242
B.1	Distributions of numbers of contractions and accepted for elliptical slice sampling.	244
B.2	Posterior scatterplots for Sierra Leone SEIR model parameters on their natural scales.	248
B.3	Posterior scatterplots for transformed Sierra Leone SEIR model parameters. .	250
B.4	Posterior scatterplots for linear combinations of transformed Sierra Leone SEIR model parameters.	252
B.5	Diagram for a stratified SEIR model with country specific outbreak dynamics and cross-country transmission.	271
B.6	Posterior traceplots for a stratified SEIR model fit to a simulated Ebola outbreak.	279
B.7	Scatterplots of Guinea-specific parameters in a stratified SEIR model for a simulated Ebola outbreak.	280
B.8	Scatterplots of Liberia-specific parameters in a stratified SEIR model for a simulated Ebola outbreak.	281
B.9	Scatterplots of Sierra Leone-specific parameters in a stratified SEIR model for a simulated Ebola outbreak.	282
B.10	Posterior distributions of parameters of the main country-specific SEIR LNA models fit to data from the West Africa Ebola outbreak.	290
B.11	Posterior traceplots for the SEIR LNA model fit to data from the Ebola outbreak in Guinea.	291
B.12	Scatterplots of parameters for the SEIR LNA model fit to data from the Ebola outbreak in Guinea.	292
B.13	Posterior traceplots for the SEIR LNA model fit to data from the Ebola outbreak in Liberia.	293
B.14	Scatterplots of parameters for the SEIR LNA model fit to data from the Ebola outbreak in Liberia.	294
B.15	Posterior traceplots for the SEIR model LNA fit to data from the Ebola outbreak in Sierra Leone.	295

B.16 Scatterplots of parameters for the SEIR LNA model fit to data from the Ebola outbreak in Sierra Leone.	296
B.17 Posterior predictive distributions of partial autocorrelations for the main country-specific SEIR LNA models fit to data from the West Africa Ebola outbreak.	297
B.18 Posterior distributions of LNA draws for the main country-specific SEIR models fit to data from the West Africa Ebola outbreak.	298
B.19 Posterior traceplots for the SEIR ODE model fit to data from the Ebola outbreak in Guinea.	302
B.20 Scatterplots of parameters for the SEIR ODE model fit to data from the Ebola outbreak in Guinea.	303
B.21 Posterior traceplots for the SEIR ODE model fit to data from the Ebola outbreak in Liberia.	304
B.22 Scatterplots of parameters for the SEIR ODE model fit to data from the Ebola outbreak in Liberia.	305
B.23 Posterior traceplots for the SEIR model ODE fit to data from the Ebola outbreak in Sierra Leone.	306
B.24 Scatterplots of parameters for the SEIR ODE model fit to data from the Ebola outbreak in Sierra Leone.	307
B.25 Posterior predictive distributions of partial autocorrelations for the main country-specific SEIR ODE models fit to data from the West Africa Ebola outbreak.	308
B.26 Posterior distributions of parameters of the main joint SEIR LNA model fit via the LNA to data from the West Africa Ebola outbreak.	316
B.27 Posterior traceplots for a stratified SEIR LNA model fit to a simulated Ebola outbreak.	317
B.28 Scatterplots of Guinea-specific parameters in a stratified SEIR LNA model fit to data from the West Africa Ebola outbreak.	318
B.29 Scatterplots of Liberia-specific parameters in a stratified SEIR LNA model fit to data from the West Africa Ebola outbreak.	319
B.30 Scatterplots of Sierra Leone-specific parameters in a stratified SEIR LNA model fit to data from the West Africa Ebola outbreak.	320
B.31 Posterior predictive distributions of partial autocorrelations for a stratified SEIR LNA model fit to data from the West Africa Ebola outbreak.	321
B.32 Posterior distributions of LNA draws for the main stratified SEIR LNA model fit to data from the West Africa Ebola outbreak.	322

B.33	Posterior traceplots for a stratified SEIR ODE model fit to a simulated Ebola outbreak.	326
B.34	Scatterplots of Guinea-specific parameters in a stratified SEIR ODE model fit to data from the West Africa Ebola outbreak.	327
B.35	Scatterplots of Liberia-specific parameters in a stratified SEIR ODE model fit to data from the West Africa Ebola outbreak.	328
B.36	Scatterplots of Sierra Leone-specific parameters in a stratified SEIR ODE model fit to data from the West Africa Ebola outbreak.	329
B.37	Posterior predictive distributions of partial autocorrelations for a stratified SEIR model fit to data from the West Africa Ebola outbreak.	330
C.1	Posterior traceplots for SIRS model parameters with time-varying force of infection.	333
C.2	Posterior histograms and pairwise hexplots for SIRS model parameters with time-varying force of infection.	334
C.3	Posterior traceplots for SIRS model parameters with time-homogeneous force of infection.	335
C.4	Posterior histograms and pairwise hexplots for SIRS model parameters with time-homogeneous force of infection.	336
C.5	Posterior distributions of LNA draws for SIRS models with time-varying and constant force of infection.	337
C.6	Induced prior distribution of for vaccination adjusted basic reproduction numbers A(H1N1)pdm09 influenza in Finland.	348
C.7	Induced prior distribution of for effective numbers of exogenous infections among youths and adults.	348
C.8	Posterior traceplots for a stratified SIRS ODE model with time-varying dynamics fit to data from the A(H1N1) influenza pandemic in Finland.	350
C.9	Histograms and pairwise scatterplots of posterior samples of the parameters from a stratified SIRS ODE model with time varying dynamics fit to data from the A(H1N1) influenza pandemic in Finland.	351
C.10	Posterior traceplots for a stratified SIRS ODE model with piecewise-homogeneous dynamics fit to data from the A(H1N1) influenza pandemic in Finland.	354
C.11	Histograms and pairwise scatterplots of posterior samples of the parameters from a stratified SIRS ODE model with piecewise homogeneous dynamics fit to data from the A(H1N1) influenza pandemic in Finland.	355

LIST OF TABLES

Table Number	Page
1.1 Ebola incidence by country and case type	6
4.1 LNA coverage simulation settings.	93
4.2 Computational performance of the ODE, LNA, and MMTL approximations.	95
4.3 Population sizes for Guinea, Liberia, and Sierra Leone, along with times at which transmission was assumed to begin.	115
4.4 Posterior estimates of outbreak dynamics with country-specific and joint SEIR models fit Ebola outbreak data.	118
5.1 Summary of notation for influenza models.	136
5.2 Estimated A(H1N1) infections and attack rates by season and age stratum. .	147
5.3 Posterior estimates of SIRS model parameters for pandemic A(H1N1) influenza in Finland.	150
5.4 Naive predictions of incidence and attack rates in the absence of vaccination.	153
5.5 Adjusted estimates of predicted incidence and attack rates in the absence of vaccination.	154
A.1 Conjugate priors and full conditionals for SIR model parameters.	192
A.2 Conjugate priors and full conditionals for SEIR model parameters.	193
A.3 Conjugate priors and full conditionals for SIRS model parameters.	194
A.4 Simulation 1 priors for SIR model parameters.	195
A.5 Simulation 1 log-posterior effective sample sizes and run times.	197
A.6 Simulation 1 SEIR model priors.	201
A.7 Simulation 1 SEIR model log-posterior effective sample sizes and run times.	202
A.8 Simulation 1 SIRS model priors.	206
A.9 Simulation 1 SIRS model run times and log-posterior effective sample sizes.	208
A.10 Simulation 2 SEIR model parameters.	215
A.11 Simulation 2 priors for SIR and SEIR models.	216
A.12 Simulation 2 posterior estimates for SIR and SEIR model parameters.	217

A.13	Simulation 3 SIR model priors.	222
A.14	Simulation 4 prior regimes.	223
A.15	Prior distributions for an SIR model fit to the British boarding school outbreak data.	233
A.16	Prior distributions for an SEIR model fit to the British boarding school outbreak data.	234
B.1	SEIR model parameter and their interpretation on their natural scales.	247
B.2	SEIR model parameter and their interpretation on a possible set of estimation scales.	249
B.3	SEIR model parameters and their interpretation on a possible set of estimation scales.	253
B.4	Parameters and priors for a coverage simulation when estimating the effective population size.	255
B.5	Coverage simulation results when estimating the effective population size	256
B.6	Run times for coverage simulation SIR models fit via the LNA, ODE, and MMTL approximations.	258
B.7	Small population coverage results for SIR models fit via the LNA, ODE, and MMTL approximations.	259
B.8	Medium population coverage results for SIR models fit via the LNA, ODE, and MMTL approximations.	260
B.9	Large population coverage results for SIR models fit via the LNA, ODE, and MMTL approximations.	261
B.10	Fixed parameter coverage simulation setup.	264
B.11	Run times for fixed parameter coverage simulations.	265
B.12	Slow dynamics, low detection probability regime fixed parameter coverage simulation results.	266
B.13	Fast dynamics, low detection probability regime fixed parameter coverage simulation results.	267
B.14	Slow dynamics, high detection probability regime fixed parameter coverage simulation results.	268
B.15	Fast dynamics, high detection probability regime fixed parameter coverage simulation results.	269
B.16	Week zero, and true and effective population sizes for a simulated Ebola outbreak in West Africa.	270
B.17	Parameters and priors for a simulated Ebola outbreak in West Africa.	272

B.18	Priors for initial compartment volumes for a simulated Ebola outbreak in West Africa	273
B.19	Transmission rates for the stratified SEIR model with migration of infecteds for Ebola in West Africa.	274
B.20	Transmission rates for the stratified SEIR model with virtual migration of infecteds for Ebola in West Africa.	275
B.21	Posterior estimates of initial numbers of exposed, infected, and recovered individuals for a stratified SEIR model fit to simulated Ebola data.	276
B.22	Posterior parameter estimates for full and approximate stratified SEIR models fit to a simulated Ebola outbreak.	277
B.23	Effective sample sizes and potential scale reduction factors for the stratified SEIR model with virtual migration of infecteds fit to a simulated Ebola outbreak.	278
B.24	Transmission rates for single country SEIR models for Ebola in West Africa.	283
B.25	Priors for initial compartment volumes for single country models fit to data from the Ebola outbreak in West Africa.	284
B.26	Parameters and priors for single country SEIR models fit to the West Africa Ebola outbreak.	285
B.27	Parameters and priors for supplementary country-specific SEIR models fit to the West Africa Ebola outbreak.	286
B.28	Posterior estimates of initial numbers of exposed, infected, and recovered individuals for country-specific SEIR LNA models fit to Ebola outbreak data.	287
B.29	Posterior parameter estimates for country-specific SEIR LNA models fit Ebola outbreak data.	288
B.30	Effective sample sizes and potential scale reduction factors for the main stratified SEIR LNA model with virtual migration of infecteds fit to data from the West Africa Ebola outbreak under tight priors.	289
B.31	Posterior estimates of initial numbers of exposed, infected, and recovered individuals for country-specific SEIR ODE models fit to Ebola outbreak data.	299
B.32	Posterior parameter estimates for country-specific SEIR ODE models fit Ebola outbreak data.	300
B.33	Effective sample sizes and potential scale reduction factors for the main stratified SEIR ODE model with virtual migration of infecteds fit to data from the West Africa Ebola outbreak under tight priors.	301
B.34	Priors for initial compartment volumes for a stratified SEIR model fit to data from the Ebola outbreak in West Africa.	310

B.35	Parameters and priors for a country-stratified SEIR model fit to the West Africa Ebola outbreak.	311
B.36	Loose priors used in a sensitivity analysis for a stratified SEIR model for Ebola in West Africa.	312
B.37	Posterior estimates of initial numbers of exposed, infected, and recovered individuals for a stratified SEIR LNA model fit to Ebola outbreak data.	313
B.38	Posterior parameter estimates for stratified SEIR LNA models fit Ebola outbreak data.	314
B.39	Effective sample sizes and potential scale reduction factors for the main stratified SEIR LNA model with virtual migration of infecteds fit to data from the West Africa Ebola outbreak under tight priors.	315
B.40	Posterior estimates of initial numbers of exposed, infected, and recovered individuals for a stratified SEIR ODE model fit to Ebola outbreak data.	323
B.41	Posterior parameter estimates for stratified SEIR ODE model fit Ebola outbreak data.	324
B.42	Effective sample sizes and potential scale reduction factors for the main stratified SEIR ODE model with virtual migration of infecteds fit to data from the West Africa Ebola outbreak under tight priors.	325
C.1	Parameters and priors for models fit two SIRS models to data from a simulated outbreak with sinusoidal FOI.	331
C.2	Effective fraction of the population that is susceptible for different reproduction numbers and dection rates.	344
C.3	Parameters and priors for age-vaccination stratified SIRS ODE models fit to the A(H1N1)pdm09 influenza outbreak in Finland.	345
C.4	Priors for initial values of GMRFs for the numbers of exogenous infection in each age stratum.	346
C.5	Priors for standard deviations of GMRF increments of reproduction numbers and rates of exogenous infection.	347
C.6	Parameters and priors for age-vaccination stratified piecewise homogeneous SIRS ODE models fit to the A(H1N1)pdm09 influenza outbreak in Finland.	353
C.7	Estimated A(H1N1) infections and attack rates by season and age stratum.	357
C.8	Posterior estimates of SIRS model parameters for pandemic A(H1N1) influenza in Finland — Sensitivity analysis to prior distribution for the effective population size.	358

GLOSSARY

ACF: Autocorrelation function.

BDA: Bayesian data augmentation

CDF: Cumulative distribution function.

CLE: Chemical Langevin equation.

CLT: Central limit theorem.

CP: Centered parameterization.

CTMC: Continuous-time Markov chain.

DA: Data augmentation.

DFE: Disease free equilibrium.

ELIPTSS: Elliptical slice sampler.

ESS: Effective sample size.

EVD: Ebola virus disease.

FOI: Force of infection.

GMRF: Gaussian Markov random field.

ILI: Influenza-like illness.

LNA: Linear noise approximation.

MCSE: Monte Carlo standard error.

MJP: Markov jump process.

MMTL: Multinomial modification of the τ –leaping algorithm.

MVNSS: Multivariate normal slice sampler.

NGM: Next generation matrix.

PACF: Partial autocorrelation function.

PMCMC: Particle Markov chain Monte Carlo.

PMMH: Particle marginal Metropolis–Hastings.

PPI: Posterior predictive interval.

PPP: Posterior predictive p–value.

PSRF: Potential scale reduction factor.

NCP: Non–centered parameterization.

SDE: Stochastic differential equation.

SEM: Stochastic epidemic model.

SIR: Susceptible–infected–recovered model.

TPM: Transition probability matrix.

VE: Vaccine efficacy.

WHO: World Health Organization.

ACKNOWLEDGMENTS

To be written.

DEDICATION

To be written.

Chapter 1

INTRODUCTION

1.1 *Partially Observed Epidemic Count Data*

Outbreak data often consist of aggregate counts reported by public health surveillance systems. The data, which typically reflect the incidence or prevalence of cases in the population, are used to develop an understanding of the outbreak dynamics, assess the effectiveness of interventions, and quantify uncertainty about how the outbreak is likely to evolve. The task of modeling an outbreak is complicated by the limited extent of epidemiological data, which are recorded at discrete observation times, commonly describe just one aspect of the disease process, e.g., infections, and usually capture only a fraction of cases. Complete subject-level data, which would consist of full transmission chains and exact times at which individuals transition through disease states, are rarely available. Asymptomatic cases and imperfect surveillance result in systematic underreporting, which makes 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 challenges involved in characterizing the outbreak dynamics are often further exacerbated by changes in the outbreak dynamics or surveillance over time, which can reduce the effective amount of available information. Hence, fitting models in the absence of complete data is a complicated missing data/latent variable problem.

1.2 *Compartmental Epidemic Models*

Compartmental epidemic models are used to describe the dynamics of an outbreak, to estimate how features of the population or environment affect its severity, and to understand

how interventions might help to disrupt transmission. A compartmental model represents the time–evolution of an outbreak in terms of the infection histories of individuals as they transition between discrete states, or model compartments. When we use a compartmental model to describe the *transmission* dynamics of an outbreak, the compartments encode structural information about how individuals at different infection states interact to transmit the infectious agent. In contrast, states in a compartmental model for *disease* dynamics typically correspond to discrete stages in the natural history of within–host disease progression without necessarily referring to a host’s transmissive potential.

A compartmental model is referred to as a mechanistic model when it prescribes physical laws that govern the transmission dynamics of an outbreak. For example, the model might specify the rates of infectious contact between classes of individuals, or with an environmental reservoir that is a vector for transmission. The mechanistic structure of the model specifies a functional form for the temporal, and possibly spatial, spread of the epidemic process that is coarsely reflected in incidence and prevalence data. In contrast with their mechanistic counterparts, phenomenological models describe the incidence process without explicitly referring to the physical system that is being observed [48, 50]. The price we pay for adopting a mechanistic approach is that our models will be obviously “wrong”, at least in the sense that all models are wrong, so it will be our responsibility to justify their reasonableness. Our reward is that mechanistic models provide us with interpretable, generative descriptions of outbreak dynamics. Moreover, the ubiquity of mechanistic models in the epidemiological literature enables us to incorporate information from previous studies into our own models to inform our prior beliefs about specific aspects of the transmission dynamics. Historical overviews of mechanistic models for disease transmission may be found in [8, 28, 119, 139].

Mechanistic compartmental models can be specified at varying levels of fidelity to the underlying epidemic process. Our models will range in granularity, from agent–based models in which we explicitly track the disease histories of individuals, to population–level models defined by the aggregate flow of individuals between model compartments. We refer to a mechanistic model as a stochastic epidemic model (SEM) when we model stochasticity

in the epidemic process. Compartmental models with complex dynamics have historically represented the epidemic process as deterministic systems of ordinary or partial differential equations [119]. All of the models in this dissertation will treat the epidemic process as evolving continuously in time, but observed at discrete times. The decision to work with continuous-time models is advantageous when observation times are not evenly spaced, or when various sub-processes evolve on different time scales [90, 192]. Discrete time models generally produce results that are not independent of the choice time-step, and can be problematic when the census interval of the data and the generation time of the epidemic are misaligned (see [192] for examples). One (very) compelling reason to prefer discrete time models is that the computational effort required to fit discrete time models can be quite low in settings where it is possible to approximate the transition densities of the epidemic process with common statistical distributions [75, 103].

The statistical questions and computational challenges involved in fitting compartmental models are different in small and large population settings. It is particularly important to account for stochasticity in the epidemic process when the population size is small since randomness in the disease process at the subject level can substantially affect the propensity of an infectious agent to spread. As we shall see, fitting a SEM to partially observed count data requires that we enumerate the possible epidemic paths from which the data could have arisen. This can be a daunting task since the set of possible epidemic paths is potentially huge, and the stochastic processes used to model outbreaks in small populations are not always easily amenable to approximation. In contrast, there are many methods for approximating the behavior of outbreaks in large, well-mixed populations that enable us to fit models with complex dynamics. As we shall see, various aspects of these models will interact in complicated ways, so we must think carefully about computation and parameterization.

In this dissertation, we contribute computational methods for fitting continuous time SEMs to partially observed epidemic count data in both small and large populations. In spite of the rather obvious ubiquity and importance of this data setting, there is a dearth of computational tools for fitting SEMs that are simple, broadly applicable, and computa-

tionally robust. The remainder of this chapter provides an overview of the datasets that motivate our methods, which will also serve to outline the organization of this dissertation.

1.3 Motivating Examples

1.3.1 Influenza in a British Boarding School

In our first application, we will analyze data from an outbreak of influenza in a British boarding school [12, 58]. This outbreak took place shortly after the Easter term began in January 1978, and was estimated to eventually infect roughly 90% of the 763 boys aged 10–18. Daily counts of the boys who were confined to the infirmary from January 22nd through February 4th were accessed via the `pomp` package in R [124].

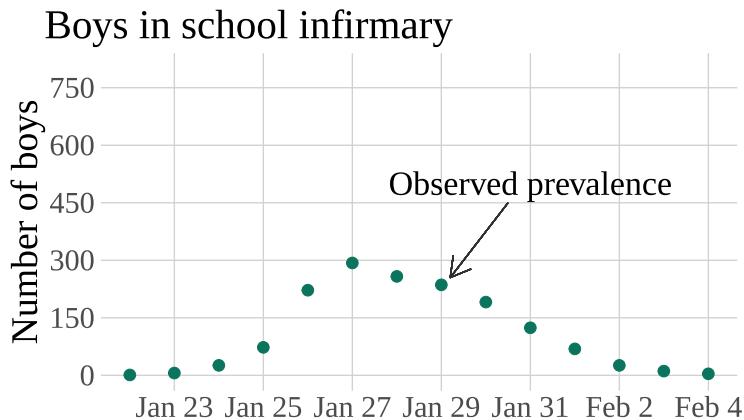


Figure 1.1: Data from an outbreak of influenza in a British boarding school.

Although this dataset is small in scale and duration, reconstructing the true path of the outbreak and describing its dynamics are highly non-trivial inferential tasks. Complicating matters is the lack of individual-level information of any sort, and the possibility that not all of the boys who were infectious on a given day were successfully detected and isolated. In Chapter 3, we present a computationally efficient algorithm for fitting stochastic epidemic models to partially observed prevalence data in small to moderate size populations. This work has been published in Fintzi, et al.[74].

1.3.2 Epidemics in Large Populations

Ebola in West Africa, 2013–2015

Starting in December 2013, the West African countries of Guinea, Liberia, and Sierra Leone experienced an outbreak of Ebola that was unprecedented in its size and duration compared to previous Ebola outbreaks. The first cases were thought to have occurred in December 2013 in the Guédéckou prefecture in Guinea, while the first cases in Liberia and Sierra Leone were detected on March 30 and June 12, 2014, respectively. By March 2016, a total of 28,646 suspected, probable, and confirmed cases had been reported along with 11,323 deaths [223]. The outbreak was exacerbated by a number of factors including insufficient outbreak response infrastructure, highly transient populations, and failures to engage with communities early on in the outbreak to implement infection prevention and control measures [52, 63].

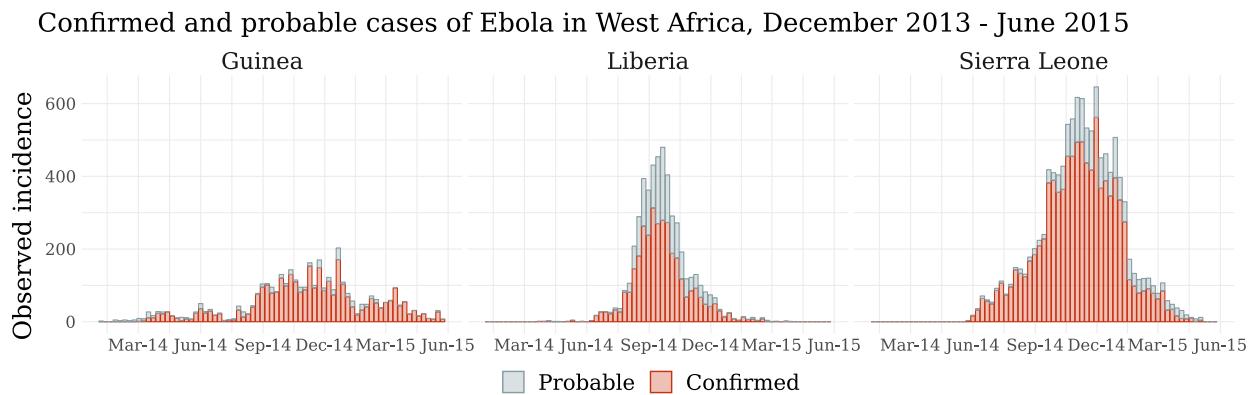


Figure 1.2: Weekly incidence of confirmed and probable cases of Ebola in Guinea, Liberia, and Sierra Leone.

There are several reasons to suspect that the true incidence is significantly under-counted in the dataset. First, suspected cases accounted for sizable fractions of the total case counts, particularly in Liberia, but were not available as part of this dataset (Table 1.1). Furthermore, analyses based on phylogenetic data [189] and case fatality ratios [13, 80] suggest that many cases may have gone undetected, particularly in the early stages, and during the peak,

of the outbreak when health systems were overwhelmed.

Table 1.1: WHO cumulative Ebola incidence by country and case type through May 24, 2015 [224]. Suspected cases were computed as the difference between the official CDC total [45], which included all three case types, and the WHO confirmed plus probable cases.

	Guinea	Liberia	Sierra Leone	Total
Confirmed ¹	3,210	3,342	9,494	16,044
Probable ¹	419	1,652	1,823	3,894
Suspected ²	12	5,672	1,389	7,073
Total	3,641	10,666	12,706	27,013

¹ Dataset used in the current analysis, cases through 05–24–2015.

² Not available as part of the dataset for this analysis.

Mathematical models were critical to informing decisions about resource allocation and interventions throughout the outbreak [52]. A review of 66 modeling studies comprising 125 models for the population-level spread of the outbreak found that modelers typically relied on pre-existing publicly available epidemiological data, which most often consisted of aggregate count data such as weekly incidence counts published by the WHO or national ministries of health [51]. Modelers working with incidence data were most often interested in describing the transmission dynamics, particularly the basic reproductive number. Other objectives included forecasting the possible time-evolution of the outbreak, and constructing models to assess the effects of various interventions on the transmission dynamics. Although modeling teams working with publicly available data were not always involved in the policy-making process, the importance of developing outbreak models using real-world data, during what for those teams would be “peace time”, has been emphasized as a critical exercise in preparing for future outbreaks [213].

Analyzing data from the Ebola outbreak presents, in some ways, a fundamentally different challenge compared to the boarding school data. As will become clear, the population sizes

of the three countries are far too large for the small population methods in Chapter 3 to be used. In Chapter 4, we develop a framework that builds on a linear noise approximation (LNA) for a broad class of compartmental SEMs [68, 128, 187, 222], and that uses state of the art Markov chain Monte Carlo (MCMC) tools to fit SEMs to partially observed incidence data in large population settings.

1.3.3 Pandemic A(H1N1) Influenza in Finland

The emergence of a pandemic influenza A strain, A(H1N1)pdm09, in the spring of 2009 led to widespread concern that it would result in high mortality and excessive stress on public health systems around the world. The strain, commonly referred to as “swine flu”, was a triple reassortment of human, avian, and swine viruses, and was of particular concern because of its similarity to the 1918 pandemic strain that infected up to a third of the world’s population and led to an estimated 50 million deaths [46]. While the burden imposed by the 2009 pandemic was ultimately comparable to that of seasonal influenza [114], it nevertheless resulted in an estimated 110,000–400,000 respiratory deaths and 50,000–180,000 cardiovascular deaths [59]. The age-standardized cumulative incidence was estimated from serological samples in 19 countries to be between 20%–27% of the total population in those countries [211]. There was substantial variability in attack rates and transmission dynamics by age, and outbreaks were consistently estimated to be more severe among children and adolescents in terms of cumulative incidence than among adults [169, 197, 211, 226].

We will analyze surveillance data, displayed in Figure 1.3, from the first and second waves of the epidemic in Finland, where the pandemic A(H1N1) strain was first detected in May, 2009, and where the first wave of the major outbreak began in October, 2009. The dataset, previously analyzed in [194, 195], consists of weekly counts of laboratory-confirmed A(H1N1) cases, aggregated into sixteen age strata, that were culled from a national surveillance system [144]. For computational considerations, we will further aggregate the data into two age groups, individuals ages 0–19 and those of ages 20+, chosen on basis of differences in mixing patterns, susceptibility, vaccination priority, and observed attack rates [121, 169, 197]. All

cases in our dataset were of mild severity, i.e., cases that did not require hospitalization. Following [194, 195], we treat all A(H1N1) cases as A(H1N1)pdm09 as nearly all A(H1N1) cases were of the pandemic strain. It is critical to note that the data are not marked by vaccination status.

An important aspect of the pandemic response in Finland was the execution of a concerted vaccination campaign with the adjuvanted monovalent vaccine Pandemrix, produced by GlaxoSmithKline. Each resident was offered one vaccine dose, free of charge, starting in October, 2009. Vaccination priority was given to health care workers, vulnerable individuals, and young people under the age of 20 [199]. Coverage levels and the timing of vaccination administration varied by age (Figure 1.3). Youths under the age of 20 tended to be vaccinated much sooner than adults, and coverage levels among children under the age of 15 were greater than 70%. Young adults between the ages of 20–29 had the lowest vaccine coverage rates at approximately 30%. In other settings, vaccine efficacy was found to be higher among children than adults [135]. Data were not available on vaccine coverage for a trivalent seasonal influenza vaccine administered during the 2010–2011 season.

In Chapter 5, we will use methods for approximate inference for SEMs developed and explored in Chapter 4 to fit models with time-varying dynamics. We will then develop an age–vaccination stratified model for the spread of pandemic A(H1N1) in Finland that we will use to pursue three lines of inquiry. First, we will quantify the transmission dynamics and effectiveness of disease surveillance during the first and second seasons. The most important aspect of this involves estimating time-varying reproduction numbers that are interpretable as thresholds for sustained transmission. Second, we will estimate age-specific attack rates and unobserved incidence curves, accounting for underreporting, so that we may better understand the true burden of the pandemic. Finally, we will attempt to understand what effect, if any, the vaccination campaign had in mitigating the severity of the outbreak.

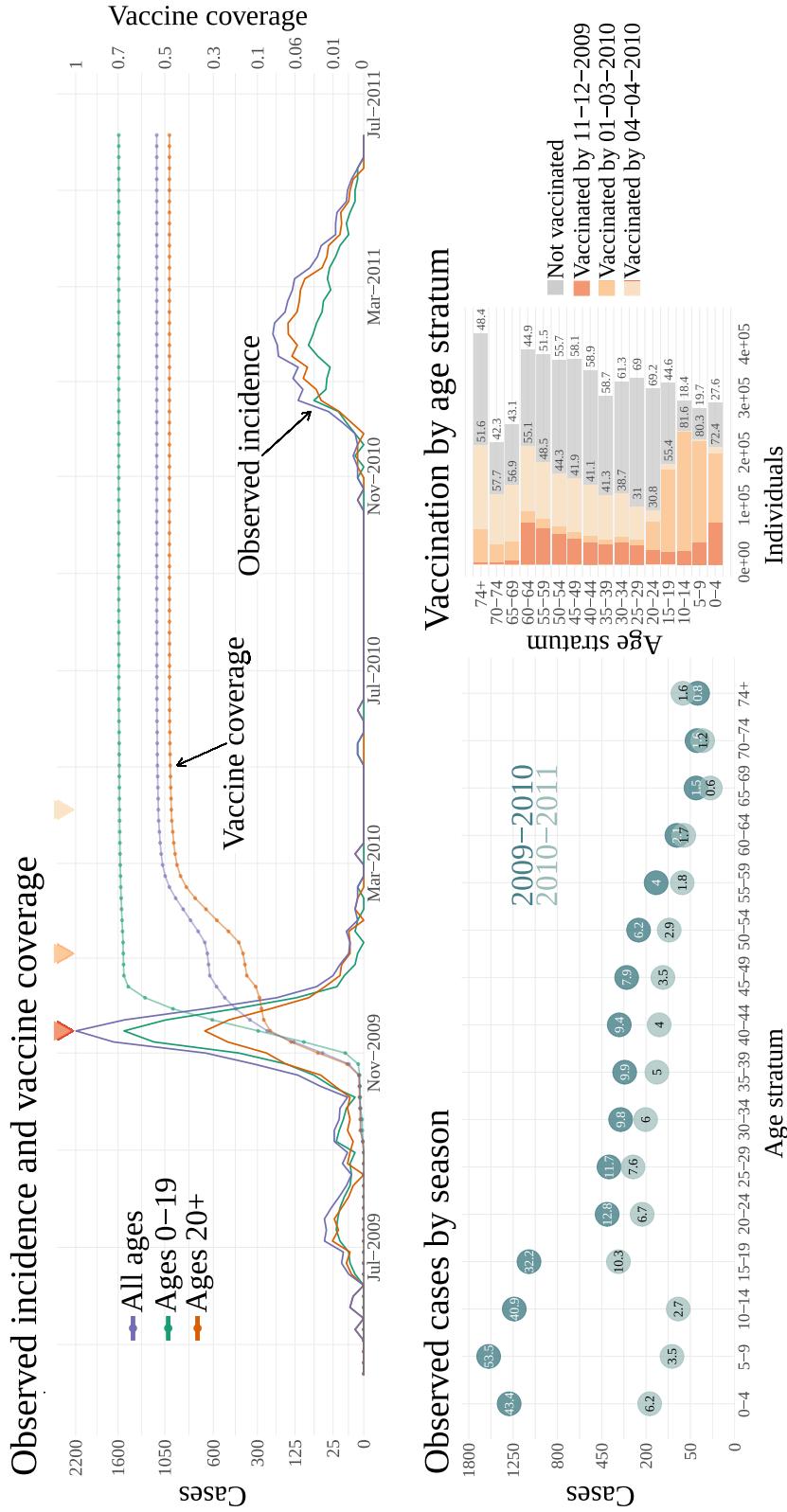


Figure 1.3: Influenza data summaries. *Top:* Observed incidence (solid lines) and vaccine coverage (lines with points). *Bottom left:* Observed cases by season and age stratum. The 2009–2010 season (dark green) corresponds to the period from April 15, 2009 through April 4, 2010. The 2010–2011 season (light green) corresponds to the period from September 12, 2010 through June 5, 2011. Numbers in points give the attack rate within each stratum for the corresponding season. *Bottom right:* Vaccination coverage by age stratum, colored by vaccine coverages at the times of peak incidence in the first season, tail of the major outbreak in the first season, and end of the vaccination campaign. The numbers inside and outside the histograms denote the percentage of individuals in each stratum that were vaccinated and unvaccinated, respectively, by the end of the vaccination campaign. The times at which vaccine coverages are summarized, denoted by colors of histogram bars, are also identified by corresponding triangles above the top figure.

Chapter 2

TECHNICAL BACKGROUND

2.1 Mathematical Models for the Spread of Infectious Diseases

Our objective is to estimate the posterior distribution of parameters, $\boldsymbol{\theta}$, that govern the dynamics of an outbreak, given the data $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_L)$, which are accrued at a set of discrete observation times, $\{t_\ell : \ell = 1, \dots, L\}$. The data are not independent and typically exhibit complicated temporal (and spatial) dependencies. Therefore, the observed data likelihood in the posterior is not a simple product of independent densities, i.e.,

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{Y}) &\propto L(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &= \prod_{\ell=1}^L \pi(\mathbf{Y}_\ell|\mathbf{Y}_1, \dots, \mathbf{Y}_{\ell-1}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \neq \prod_{\ell=1}^L \pi(\mathbf{Y}_\ell|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).\end{aligned}$$

Furthermore, the transition density, $\pi(\mathbf{Y}_\ell|\mathbf{Y}_1, \dots, \mathbf{Y}_{\ell-1})$, involves a high dimensional integral over an unobserved epidemic process, \mathbf{X} ,

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) \propto \int \prod_{\ell=1}^L \pi(\mathbf{Y}_\ell|\mathbf{Y}_1, \dots, \mathbf{Y}_{\ell-1}, \mathbf{X}, \boldsymbol{\theta})\pi(\mathbf{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\mathbf{X}. \quad (2.1)$$

This integral is analytically intractable when the state space of \mathbf{X} is of even moderate size. Simply put, it is difficult to enumerate how the data could have arisen, given the vastness of possibilities for how an unseen outbreak might have evolved.

In this dissertation, we will describe the latent epidemic process using continuous-time models that possess the Markov property, meaning that the forward-time evolution of the latent process depends on its history only through its current state. This choice reduces the structure of (2.1) to that of a hidden Markov model (HMM) wherein the data are conditionally independent given the latent epidemic process (diagrammed in Figure 2.1).

The posterior is

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) \propto \int \prod_{\ell=1}^L L(\mathbf{Y}_\ell|\mathbf{X}(t_\ell), \boldsymbol{\theta}) \pi(\mathbf{X}(t_\ell)|\mathbf{X}(t_{\ell-1}), \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\mathbf{X}. \quad (2.2)$$

On its own, this might seem to solve nothing. The integral in (2.2) is still analytically intractable, and the state space of \mathbf{X} is still too large to easily work with the augmented posterior,

$$\pi(\boldsymbol{\theta}, \mathbf{X}|\mathbf{Y}) \propto L(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) \pi(\mathbf{X}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}). \quad (2.3)$$

However, the Markov property of the latent process also induces symmetries in the ways that individuals interact and contribute to the outbreak dynamics. This additional structure is critical in facilitating efficient computation. In the following sections, we present a number of representations for the latent epidemic process, and conclude with a brief overview of computational approaches for inference.

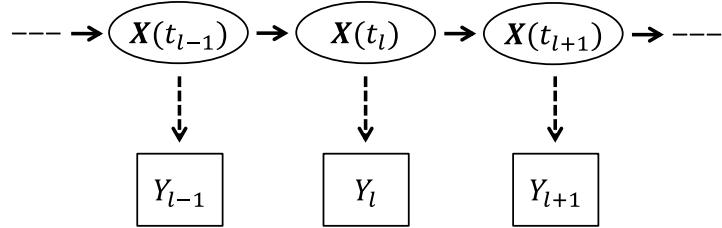


Figure 2.1: Diagram of a Hidden Markov model. The data, \mathbf{Y} , are conditionally independent given the latent epidemic process, \mathbf{X} .

2.1.1 An Agent-Based Susceptible–Infected–Recovered Model

For clarity of exposition, we will present the technical background on epidemic models in terms of the susceptible–infected–recovered (SIR model). Formal treatments that deal with this material in greater generality can be found in [9, 27, 33, 78, 97, 222]. The SIR model classifies individuals in a population of size N into one of three infection states: susceptible (S), infected (I), and recovered (R). Individuals are assumed to become infectious immediately

upon entering the infected state, and acquire lasting immunity upon recovery. To simplify matters, we will assume the population is closed, meaning that there are no demographic changes or immigration, and that individuals are exchangeable. This latter assumption implies that individuals mix homogeneously and are alike in their infection dynamics.

The SIR model defines an epidemic process, $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$, that collects the subject-level subprocesses, \mathbf{X}_j , $j = 1, \dots, N$, each of which takes values in the state space of disease state labels, $\mathcal{S}_j = \{S, I, R\}$. A realized subject-path is of the form

$$\mathbf{x}_j = \begin{cases} S, & \tau < \tau_I^{(j)}, \\ I, & \tau_I^{(j)} \leq \tau < \tau_R^{(j)}, \\ R, & \tau_R^{(j)} \leq \tau, \end{cases} \quad (2.4)$$

where $\tau_I^{(j)}$ and $\tau_R^{(j)}$ are the infection and recovery times for subject j , and are possibly infinite (Figure 2.2). The state space of \mathbf{X} is $\mathcal{S} = \{S, I, R\}^N$, the Cartesian product of subject-level state labels. We denote by $\mathbf{X}(\tau) = (\mathbf{X}_1(\tau), \dots, \mathbf{X}_N(\tau))$ the state of \mathbf{X} at time τ , and by $\mathbf{X}(\tau^+)$ the state just after time τ .

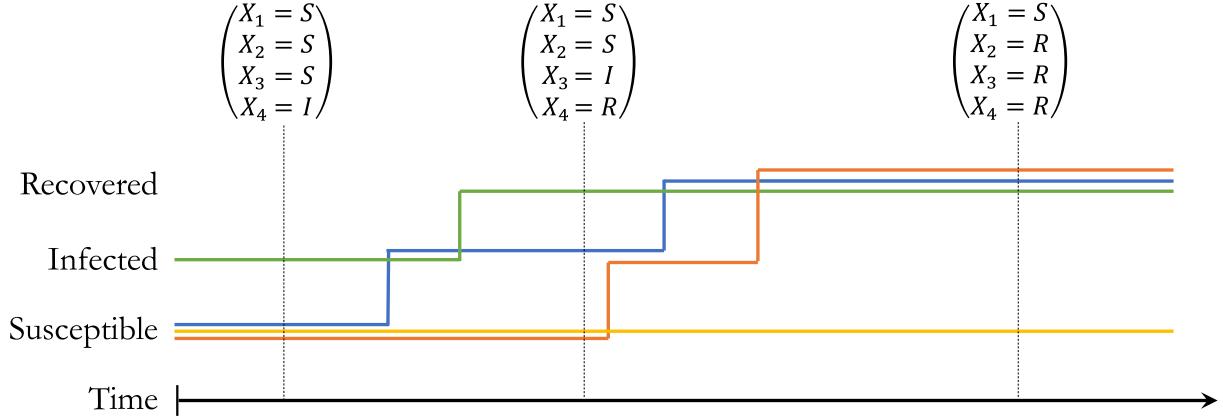


Figure 2.2: Diagram of subject-level paths (colored lines) for an SIR model in a population of size $N=4$. Individuals transition through infection states continuously in time. The epidemic process is defined in terms of the infection states of individuals in the population.

The waiting times between subject-level transition events are typically modeled as being exponentially distributed. This will allow us to take advantage of several useful properties of exponential random variables (List 2.1, see [222] for proofs). Critically, this choice also implies that \mathbf{X} evolves as a continuous-time Markov chain (CTMC) with transition rate from configuration \mathbf{X} to \mathbf{X}' , differing only in the state of a single subject, given by

$$\lambda_{\mathbf{x}, \mathbf{x}'} = \begin{cases} \beta I, & \text{if } \mathbf{X} \text{ and } \mathbf{X}' \text{ differ only in subject } j, \text{ with } \mathbf{X}_j = S, \text{ and } \mathbf{X}'_j = I, \\ \mu, & \text{if } \mathbf{X} \text{ and } \mathbf{X}' \text{ differ only in subject } j, \text{ with } \mathbf{X}_j = I, \text{ and } \mathbf{X}'_j = R, \\ 0, & \text{for all other configurations } \mathbf{X} \text{ and } \mathbf{X}'. \end{cases} \quad (2.5)$$

with per-contact infection rate, β , and recovery rate, μ . The quantity, $1/\mu$, is interpreted as the mean infectious period duration. That \mathbf{X} is a *Markov* process means that its forward-time evolution depends on its history only through its current state, i.e.,

$$\Pr(\mathbf{X}(\tau + d\tau) = \mathbf{x}' | \{\mathbf{X}(\tau) = \mathbf{x}, \tau \in [0, \tau]\}, \boldsymbol{\theta}) = \Pr(\mathbf{X}(\tau + d\tau) = \mathbf{x}' | \mathbf{X}(\tau) = \mathbf{x}, \boldsymbol{\theta}), \quad (2.6)$$

where $\mathbf{x}', \mathbf{x} \in \mathcal{S}$. \mathbf{X} is *time-homogeneous* since the rates of transition between configurations in the state space of \mathbf{X} are constant over time.

Let $\boldsymbol{\tau} = \{\tau_0, \dots, \tau_{K+1}\}$, be the (ordered) set of K infection and recovery times of all individuals along with the endpoints of the time period $[\tau_0, \tau_{K+1}]$. Let $\mathbb{1}_{\{\tau_k \triangleq I\}}$ and $\mathbb{1}_{\{\tau_k \triangleq R\}}$ indicate whether τ_k is an infection or recovery time, and let $\boldsymbol{\theta} = (\beta, \mu, \mathbf{X}_0)$ denote the vector of parameters, including the initial state of \mathbf{X} at time τ_0 . The CTMC likelihood of \mathbf{X} over $[\tau_0, \tau_{K+1}]$ is a product of exponential waiting time densities,

$$\begin{aligned} L(\mathbf{X} | \boldsymbol{\theta}) &= \prod_{k=1}^K \left\{ [\beta I_{\tau_k} \times \mathbb{1}_{\{\tau_k \triangleq I\}} + \mu \times \mathbb{1}_{\{\tau_k \triangleq R\}}] \exp[-(\tau_k - \tau_{k-1})(\beta I_{\tau_k} S_{\tau_k} + \mu I_{\tau_k})] \right\} \\ &\quad \times \exp\left[-(t_L - \tau_K)(\beta I_{\tau_K^+} S_{\tau_K^+} + \mu I_{\tau_K^+})\right]. \end{aligned} \quad (2.7)$$

List 2.1: Useful properties of exponential random variables.

- (Memoryless property) If $Z \sim \text{Exp}(\lambda)$, then $\forall t, dt \geq 0$ we have

$$\Pr(Z > t + dt | Z > t) = \Pr(Z > dt). \quad (2.8)$$

- (Racing exponentials) If $Z_i \sim \text{Exp}(\lambda_i)$, $i = 1, \dots, n$, are independent, then

$$\min_i(Z_i) \sim \text{Exp}\left(\lambda = \sum_i \lambda_i\right). \quad (2.9)$$

- (Index of minimum) If $Z_i \sim \text{Exp}(\lambda_i)$, $i = 1, \dots, n$, are independent, then the index k of the minimum of Z_i is a random variable with probability mass function

$$\Pr(k | Z_k = \min(Z_1, \dots, Z_n)) = \frac{\lambda_k}{\sum_j \lambda_j}. \quad (2.10)$$

- (Minimum is independent of its index) If $Z_i \sim \text{Exp}(\lambda_i)$ are independent, then $M = \min_i(Z_i)$, which is the minimum of Z_i , and I , the index of the minimum taking value $I = k$ if $Z_k = M$, are independent.

2.1.2 A Population–Level Susceptible–Infected–Recovered Model

The subject–level SIR model is equivalent to an aggregated SIR model, expressed in terms of compartment counts [5, 9]. This equivalence derives from two properties, *lumpability* and *commutativity*, possessed by the Markov processes used to construct the above subject–level SEMs. The population–level SIR model is usually presented for computational reasons since discarding the subject labels associated with infections and recoveries substantially reduces the computational burden of caching subject–level paths. We refer to [203] for a more formal

presentation of the following discussion.

Given a Markov process, \mathbf{X} with state space $\mathcal{S} = \{s_1, \dots, s_P\}$ and initial probability vector π , we define another process, $\bar{\mathbf{X}}$ on the state space $\bar{\mathcal{S}} = \{S_1, \dots, S_L\}$, which is a *partition* of \mathcal{S} . In our setting, the partitioning will map a configuration in \mathcal{S} to a set in $\bar{\mathcal{S}}$ by counting the number of people in each model compartment. The jump chain of the new process is obtained by partitioning the jump chain of the complete process. We want to establish conditions under which $\bar{\mathbf{X}}$ is stochastically coupled to \mathbf{X} .

Suppose the initial distribution of $\bar{\mathbf{X}}(t_0)$, induced by the distribution of $\mathbf{X}(t_0)$, is

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

and that its transition probabilities are

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

where $\bar{\mathbf{x}}(t')$ and $\mathbf{x}(t')$ denote the paths of the complete Markov process and the new process. We say that the complete Markov process is *lumpable* with respect to a partition, $\bar{\mathcal{S}}$, of its state space if lumping the complete process results in a Markov process with respect to the lumped state space for every choice of π whose transition probabilities do not depend on π . We refer to the process obtained by lumping as the *lumped* Markov jump process (MJP). We say that the complete process is *commutative* with respect to lumping if the set of jump chains obtained by applying the lumping to jump chains of the complete MJP are equal in distribution to jump chains of the lumped MJP.

Let S_A and S_B be elements of $\bar{\mathcal{S}}$, i.e., sets of states $s_j \in \mathcal{S}$. The rate matrix of a CTMC is lumpable if

$$\sum_{s_b \in S_B} \lambda_{s_a, s_b} = \sum_{s_b \in S_B} \lambda_{s_c, s_b}$$

for any pair of sets $S_A, S_B \in \bar{\mathcal{S}}$ and any pair of states $(s_a, s_c) \in S_A$. A Markov process \mathbf{X} is lumpable with respect to a partition of its state space if and only if its rate matrix is lumpable. Lumpability implies that transition probabilities for the lumped chain can be computed based on the lumped rate matrix [203].

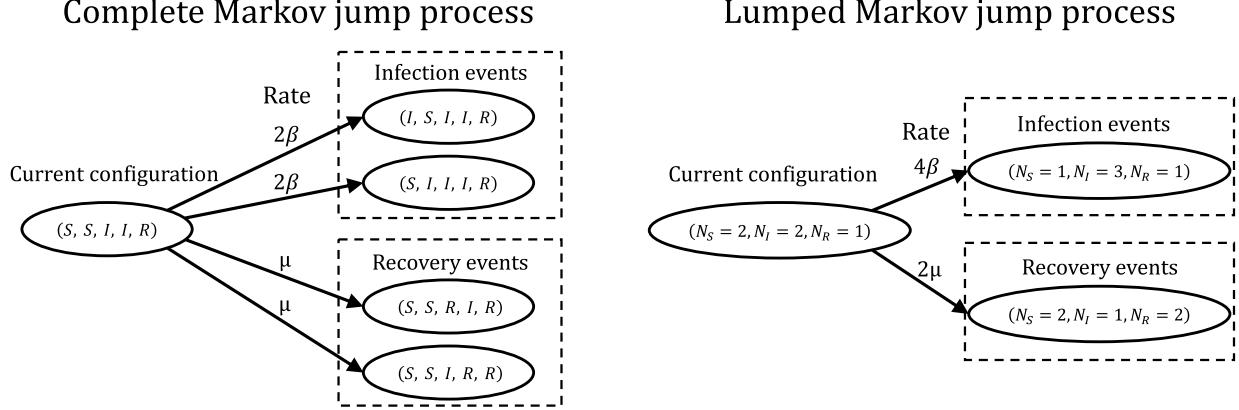


Figure 2.3: Complete and lumped representations of SIR dynamics in a population of five individuals. The per-contact infectivity rate, β , and the recovery rate, μ , parameterize exponential waiting time distributions between transition events. The complete Markov jump process evolves on the state space 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, $\bar{\mathcal{S}} = \{N_S, N_I, N_R : N_S + N_I + N_R = N\}$, 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 $\tau_{\min} = \min(\tau_1, \tau_2) \sim \exp(\lambda_1 + \lambda_2)$.

Suppose \mathbf{X} is lumpable with respect to the partition $\bar{\mathcal{S}}$ of \mathcal{S} . Then \mathbf{X} is commutative with respect to the partition if and only if the rate matrix of \mathbf{X} satisfies

$$\lambda_{s_a, s_b} = 0, \text{ if } s_a, s_b \in S_A, \text{ and } s_a \neq s_b.$$

In our context, commutativity means that the transition rate is zero between different configurations \mathbf{x}_a and \mathbf{x}_b with equal compartment counts. Commutativity implies that lumped quantities of interest for \mathbf{X} , such as transition rates or transition probabilities, can be equivalently computed based on \mathbf{X} , or based on the lumped process, $\bar{\mathbf{X}}$ [203].

Turning back to the SIR model, we defined the epidemic process, $\mathbf{X}(\tau) = (\mathbf{X}_1, \dots, \mathbf{X}_N)$, with state space $\mathcal{S} = \{S, I, R\}^N$. Let $\mathbf{x} = (x_1, \dots, x_N)$ denote a configuration of the state

labels (e.g. $\mathbf{x} = (S, I, S, R, I)$), and let

$$\bar{\mathbf{x}} = h(\mathbf{x}) = \left(l = \sum_{i=1}^N \mathbb{1}_{\{x_i=S\}}, m = \sum_{i=1}^N \mathbb{1}_{\{x_i=I\}}, n = \sum_{i=1}^N \mathbb{1}_{\{x_i=R\}} \right)$$

be the corresponding vector of compartment counts. The lumped state space is

$$\bar{\mathcal{S}} = \{\bar{\mathbf{x}} = (l, m, n) : l, m, n \in \{0, \dots, N\}, l + m + n = N\},$$

which partitions \mathcal{S} by summing the number of individuals in each disease state.

The population-level SIR model expressed in terms of compartment counts, $\bar{\mathbf{X}} = (S, I, R) \in \bar{\mathcal{S}}$ (depicted in Figure 2.3), evolves as a CTMC on the lumped state space $\bar{\mathcal{S}}$ with transition rates

<u>Transition</u>	<u>Lumped Rate</u>
$(S, I, R) \longrightarrow (S - 1, I + 1, R)$	$\beta SI,$
$(S, I, R) \longrightarrow (S, I - 1, R + 1)$	$\mu I.$

To see how we arrive at these rates, note that in a population with S susceptibles, each of whom is independently infected at rate βI , the time until the first infection is exponentially distributed with rate $\lambda_{SI} = \beta SI$ by the racing exponentials property (2.9). Similarly, the time to the first recovery in a population with I infected individuals is exponentially distributed with rate $\lambda_{IR} = \mu I$. Note that the mean infectious period duration is still $1/\mu$, as it was in the case of the subject-level SIR model.

Again, let $\boldsymbol{\tau} = \{\tau_0, \dots, \tau_{K+1}\}$, be the (ordered) set of K infection and recovery times, along with the endpoints of the time period $[\tau_0, \tau_{K+1}]$ over which the outbreak is modeled. We indicate by $\mathbb{1}_{\{\tau_k \leq I\}}$ and $\mathbb{1}_{\{\tau_k \leq R\}}$ whether τ_k is an infection or recovery time, and let $\boldsymbol{\theta} = (\beta, \mu, \bar{\mathbf{X}}_0)$ denote the vector of parameters, including the initial state of $\bar{\mathbf{X}}$ at time τ_0 . The CTMC likelihood of $\bar{\mathbf{X}}$ over $[\tau_0, \tau_{K+1}]$ is a product of exponential waiting time densities,

$$\begin{aligned} L(\bar{\mathbf{X}} | \boldsymbol{\theta}) &= \prod_{k=1}^K \left\{ [\beta S_{\tau_k} I_{\tau_k} \times \mathbb{1}_{\{\tau_k \leq I\}} + \mu I_{\tau_k} \times \mathbb{1}_{\{\tau_k \leq R\}}] \exp [-(\tau_k - \tau_{k-1})(\beta I_{\tau_k} S_{\tau_k} + \mu I_{\tau_k})] \right\} \\ &\quad \times \exp \left[-(t_L - \tau_K) \left(\beta I_{\tau_K^+} S_{\tau_K^+} + \mu I_{\tau_K^+} \right) \right]. \end{aligned} \tag{2.11}$$

2.1.3 A Brief Review of CTMCs

We briefly digress from our discussion of epidemic models to review some basic properties of CTMCs with more generality. The following discussion is not intended to be comprehensive, but rather to provide an overview of some results that will be useful in this dissertation. We refer to [29, 78, 99, 222] for more complete and rigorous discussions of the following material. For simplicity, we will focus on CTMCs with finite state spaces.

The forward-time evolution of a CTMC is described by its transition kernel

$$\Pr(\mathbf{X}(\tau') = \mathbf{x}' | \mathbf{X}(\tau) = \mathbf{x}) = \mathbf{P}_{\mathbf{x}, \mathbf{x}'}(\tau, \tau'),$$

where $\mathbf{x}, \mathbf{x}' \in \mathcal{S}$ and $0 \leq \tau \leq \tau'$. The transition kernel depends only on the time elapsed when the process is time-homogeneous $\mathbf{P}_{\mathbf{x}, \mathbf{x}'}(\tau, \tau') = \mathbf{P}_{\mathbf{x}, \mathbf{x}'}(|\tau' - \tau|)$. We will use the shorthand notation $\mathbf{P}(\tau)$ to simply denote the transition kernel for a time-homogeneous CTMC over a time interval of length τ , and $P_{ij}(\tau)$ to denote its i, j element. In matrix form, the transition kernel, $\mathbf{P}(\tau)$, is an $r \times r$ stochastic matrix, where $r = |\mathcal{S}|$, the rows of $\mathbf{P}(\tau)$ sum to one, and $P_{ii}(\tau) = -\sum_{j \neq i} P_{ij}(\tau)$. Trivially, $\mathbf{P}(0) = \mathbf{I}$, i.e., there are no state changes in a time interval of zero length, almost surely. By the Markov property, the probability of transitioning from state i to state k in time $s + t$ is

$$P_{ik}(s + t) = \sum_{j \in \mathcal{S}} P_{ij}(s) P_{jk}(t), \quad (2.12)$$

or in matrix form,

$$\mathbf{P}(s + t) = \mathbf{P}(s)\mathbf{P}(t).$$

Equation (2.12) is known as the *Chapman–Kolmogorov* master equation.

The transition rate matrix (or *infinitesimal generator*) of \mathbf{X} is defined as the derivative of $\mathbf{P}(\tau)$ at $\tau = 0$, i.e.,

$$\begin{aligned} \boldsymbol{\Lambda} &= \lim_{d\tau \rightarrow 0} \frac{\mathbf{P}(d\tau) - \mathbf{P}(0)}{d\tau} \\ &= \lim_{d\tau \rightarrow 0} \frac{\mathbf{P}(d\tau) - \mathbf{I}}{d\tau}, \end{aligned}$$

which implies that the infinitesimal transition matrix is

$$\mathbf{P}(d\tau) = \mathbf{I} + \boldsymbol{\Lambda}d\tau.$$

For example, the infinitesimal transition probabilities for the population–level SIR model are

$$\Pr(\mathbf{X}(\tau + d\tau) = (S - 1, I + 1, R) | \mathbf{X}(\tau) = (S, I, R)) = \beta SI d\tau + o(d\tau),$$

$$\Pr(\mathbf{X}(\tau + d\tau) = (S, I - 1, R + 1) | \mathbf{X}(\tau) = (S, I, R)) = \mu I d\tau + o(d\tau).$$

The transition probability matrix over an interval of arbitrary length, τ , solves the matrix differential equation,

$$\frac{d}{d\tau} \mathbf{P}(\tau) = \boldsymbol{\Lambda} \mathbf{P}(\tau), \quad s.t. \quad \mathbf{P}(0) = \mathbf{I}. \quad (2.13)$$

Thus, $\mathbf{P}(\tau) = \exp(\boldsymbol{\Lambda}\tau)$, where $\exp(\cdot)$ denotes the matrix exponential. There are many ways to compute the matrix exponential [155]. We will typically do so by diagonalizing $\boldsymbol{\Lambda}$ and exponentiating the eigenvalues. Section A.1 outlines this procedure in two separate cases: when $\boldsymbol{\Lambda}$ has real-valued eigenvalues, and also when the eigenvalues are complex. Equation (2.13) is known as the *forward* Kolmogorov equation, and is obtained by computing the derivative

$$\begin{aligned} \frac{d}{d\tau} \mathbf{P}(\tau) &= \frac{\mathbf{P}(\tau + d\tau) - \mathbf{P}(\tau)}{d\tau} \\ &= \frac{\mathbf{P}(d\tau)\mathbf{P}(\tau) - \mathbf{P}(\tau)}{d\tau} \\ &= \frac{\mathbf{P}(d\tau) - \mathbf{I}}{d\tau} \mathbf{P}(\tau) \\ &= \boldsymbol{\Lambda} \mathbf{P}(\tau). \end{aligned}$$

The *backward* Kolmogorov equation is similarly obtained by

$$\begin{aligned} \frac{d}{d\tau} \mathbf{P}(\tau) &= \frac{\mathbf{P}(\tau + d\tau) - \mathbf{P}(\tau)}{d\tau} \\ &= \frac{\mathbf{P}(\tau)\mathbf{P}(d\tau) - \mathbf{P}(\tau)}{d\tau} \\ &= \mathbf{P}(\tau) \frac{\mathbf{P}(d\tau) - \mathbf{I}}{d\tau} \\ &= \mathbf{P}(\tau) \boldsymbol{\Lambda}. \end{aligned} \quad (2.14)$$

2.1.4 Large Population Approximations

It is often infeasible to work with the CTMC formulation of a SEM when modeling an outbreak in a large population, particularly when working within a Bayesian MCMC framework. The cardinality of the model's state space grows polynomially in the population size. This makes it difficult to efficiently sample from the posterior or to explore the likelihood surface, even when fitting SEMs with relatively simple dynamics. A second challenge is that as the population size grows, so too do the numbers of transition events. Hence, repeatedly evaluating the likelihoods (2.7) and (2.11) is prohibitively expensive.

Two commonly used approximations of the MJP representation of a SEM are through a system of ordinary differential equations (ODEs) and through a system of stochastic differential equations (SDEs). We will not go into detail on the derivations of these representations, except to give some intuition about the conditions under which they are appropriate, and refer to [6, 97] for more detailed overviews.

We presented the MJP representation of the SIR model as an *extensive*, \mathbf{X} , where transition events led to jumps of size one in the compartment counts. We could have equivalently defined the model in terms of an *intensive* process, $\tilde{\mathbf{X}} = \mathbf{X}/N = (S/N, I/N, R/N)$, where transition events led to jumps of size $1/N$. As $N \rightarrow \infty$, it becomes reasonable to consider approximating $\tilde{\mathbf{X}}$ by a process with continuous sample paths. The large population stochastic *approximation* is the solution to an Itô diffusion whose sample paths are continuous but nowhere differentiable. The infinite population deterministic functional *limit* solves a system of ODEs and has smooth sample paths. We will present extensive forms of the SDE and ODE models, which are obtained by simply rescaling the intensive processes from which the approximations derive.

Deterministic representation as a system of ODEs

In the infinite population limit, sample paths of the MJP converge pointwise to their deterministic functional limit, which is given by the solution to a system of ordinary differential

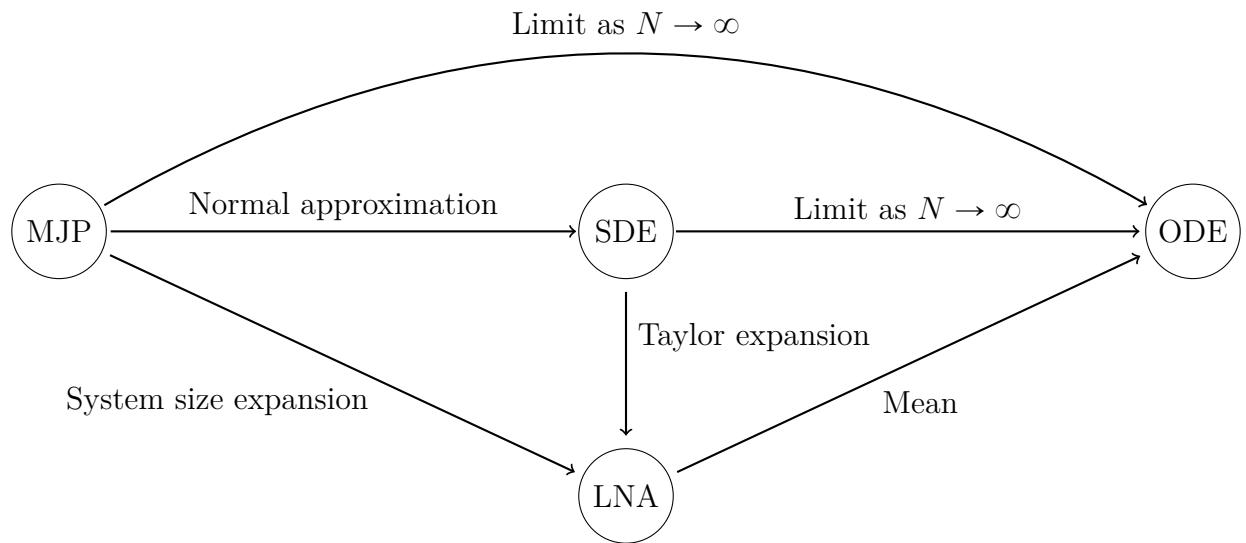


Figure 2.4: Relationship between various compartmental model representations. The Markov jump process (MJP) is typically the model of interest, and in large populations can be approximated by a system of stochastic differential equations (SDE). The linear noise approximation (LNA) is a tractable approximation obtained by Taylor expansion of the SDE about its deterministic limit, given by a system of ordinary differential equations (ODE), or by a system size expansion of the backward Kolmogorov equation of the MJP.

equations [97, 132]. For the SIR model, $\mathbf{X}(t) = \lim_{N \rightarrow \infty} \mathbf{E}(N\tilde{\mathbf{X}}(t))$ is the solution to the following system of ODEs, subject to the initial constraint that $\mathbf{X}(\tau_0) = \mathbf{x}_0$:

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - \mu I, \quad \frac{dR}{dt} = \mu I.$$

The state space of the ODE representation of the SIR model is

$$\mathcal{S}^R = \{(j, k, l) : j, k, l \in [0, N], j + k + l = N\}.$$

Note that the ODE model implies that if $I(\tau) > 0$ at any time τ , then $I(\tau) > 0 \forall \tau \in [0, \infty)$. Therefore, the ODE model is not appropriate if we are interested in answering questions about the stochastic emergence or extinction of an outbreak.

ODE formulations of SEMs are particularly useful, in part, because they more easily lend themselves mathematical analysis. One important quantity that is straightforwardly derived for ODE models is the basic reproduction number, R_0 , which is interpreted as the expected number of secondary infections attributable to a single typical index case in an otherwise susceptible population (of infinite size). R_0 can be calculated as the spectral radius of the next generation matrix (NGM) for the linearized system of ODEs at the disease free equilibrium (DFE) [62, 208]. Other methods exist, e.g., by analyzing the survival function or by fitting phenomenological models when the growth rate of the outbreak is observed to be subexponential [208]. The formulation of R_0 is more complicated for stochastic models, though it can be derived from a branching process approximation for the early behavior of an outbreak [5]. When the epidemic process is stochastic, the basic reproduction number can be related to both the final size distribution and to the probability of a major outbreak [6, 97, 152].

As an example, we will demonstrate the NGM method for calculating R_0 for an outbreak with SIR dynamics where a fraction of the population, p_v , is vaccinated and vaccine efficacy (VE) modifies the per-contact rate of infection (VE for susceptibility: ν_s), the infectiousness of carriers (VE for infectiousness: ν_i), and the rate of recovery (VE for recovery: ν_r). We will write $N^u = (1 - p_v)N$ and $N^v = p_v N$. The DFE is

$$\mathbf{X}_{DFE} = \left(S_{DFE}^{(u)} = (1 - p_v)N, I_{DFE}^{(u)} = 0, R_{DFE}^{(u)} = 0, S_{DFE}^{(v)} = p_v N, I_{DFE}^{(v)} = 0, R_{DFE}^{(v)} = 0 \right).$$

The linearized system of ODEs at the DFE is

$$\frac{dI^{(u)}}{dt} = \beta (I^{(u)} + \nu_i I^{(v)}) N^{(u)} - \mu I^{(u)}, \quad \frac{dI^{(v)}}{dt} = \beta \nu_s (I^{(u)} + \nu_i I^{(v)}) N^{(v)} - \mu \nu_r I^{(v)}.$$

The NGM is constructed as $\mathbf{K} = -\mathbf{T}\boldsymbol{\Sigma}^{-1}$, where bT gives the rates of infectious contact between states at infection, $I^{(u)}$ and $I^{(v)}$, and $\boldsymbol{\Sigma}$ contains the recovery rates out of states at infection. Here,

$$\begin{aligned} \mathbf{T} &= \begin{matrix} I^{(u)} & I^{(v)} \\ I^{(u)} & I^{(v)} \end{matrix} \left(\begin{matrix} \beta N^u & \nu_i \beta N^u \\ \nu_s \beta N^v & \nu_s \nu_i \beta N^v \end{matrix} \right), \quad \boldsymbol{\Sigma} = \begin{matrix} I^{(u)} & I^{(v)} \\ I^{(v)} & I^{(v)} \end{matrix} \left(\begin{matrix} -\mu & 0 \\ 0 & -\mu \nu_r \end{matrix} \right), \\ \mathbf{K} &= \begin{matrix} I^{(u)} & I^{(v)} \\ I^{(u)} & I^{(v)} \end{matrix} \left(\begin{matrix} \frac{\beta N^u}{\mu} & \frac{\nu_i \beta N^u}{\nu_r \mu} \\ \frac{\nu_s \beta N^v}{\mu} & \frac{\nu_s \nu_i \beta N^v}{\nu_r \mu} \end{matrix} \right). \end{aligned}$$

Thus,

$$R_0 = \sigma(K) = \frac{1}{2} \left(Tr(K) + \sqrt{Tr(K)^2 - 4 \det(K)} \right) = \frac{\beta N^u}{\mu} + \frac{\nu_s \nu_i \beta N^v}{\nu_r \mu}.$$

Diffusion approximations of Markov Jump Processes

We will give the diffusion approximation for the SIR model after a, somewhat colloquial, review of diffusions based on material in [78, 164, 190, 222]. We begin with the SDE

$$\frac{d\mathbf{X}_t}{dt} = \boldsymbol{\mu}(\mathbf{X}_t, t) + \boldsymbol{\Sigma}(\mathbf{X}_t, t)\mathbf{Z}_t, \quad t \geq s; \quad \mathbf{X}_x = \mathbf{x}, \quad (2.15)$$

with *drift vector* $\boldsymbol{\mu} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and *diffusion matrix* $\boldsymbol{\Sigma} : \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d$, which are interpretable as the infinitesimal first and second moments of the process innovations. \mathbf{Z}_t is d -dimensional Gaussian white noise. We denote by $d\mathbf{W}_t = \mathbf{Z}_t dt$ standard d -dimensional Brownian motion (see [164] for a formal definition). An *Itô diffusion* is a stochastic process, $(\mathbf{X}_t)_{t \geq 0}$, that satisfies the Itô stochastic integral equation

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \boldsymbol{\mu}(\mathbf{X}_s, s) ds + \int_0^t \boldsymbol{\Sigma}(\mathbf{X}_s, s) d\mathbf{W}_s, \quad (2.16)$$

which we can express equivalently in differential form,

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t, t)dt + \boldsymbol{\Sigma}(\mathbf{X}_t, t)d\mathbf{W}_t. \quad (2.17)$$

The SDE (2.17) can be interpreted as the limit of a difference equation, $\Delta\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t, t)\Delta t + \boldsymbol{\Sigma}(\mathbf{X}_t, t)\Delta\mathbf{W}_t$, with infinitely small time steps, $\Delta t \rightarrow 0$. Hence, we can simulate approximate sample paths using an *Euler–Maruyama* scheme by taking $\Delta t = \epsilon > 0$ and sampling the innovations, $\Delta\mathbf{X}_t$, according to

$$\Delta\mathbf{X}_t \sim MVN(\boldsymbol{\mu}(\mathbf{X}_t, t)\Delta t, \boldsymbol{\Sigma}(\mathbf{X}_t, t)\boldsymbol{\Sigma}(\mathbf{X}_t, t)^T\Delta t).$$

Given a d -dimensional Itô process of the form (2.17), *Itô’s lemma* gives a formula for the SDE satisfied by a transformed process [164]. Let $g(\mathbf{x}, t) = (g_1(\mathbf{x}, t), \dots, g_p(\mathbf{x}, t))$ be a twice continuously differentiable map from $\mathbb{R}^d \times [0, \infty) \rightarrow \mathbb{R}^p$. Then the transformed process, $\mathbf{Y}(\mathbf{W}, t) = g(\mathbf{X}_t, t)$, is also an Itô process with component Y_k , $k = 1, \dots, p$, given by

$$dY_k = \frac{\partial g_k}{\partial t}(\mathbf{X}, t) + \sum_i \frac{\partial g_k}{\partial x_i}(\mathbf{X}, t) dX_i + \frac{1}{2} \sum_i \sum_j \frac{\partial^2 g_k}{\partial x_i \partial x_j}(\mathbf{X}, t) dX_i dX_j,$$

where $dW_i dW_j = dt dW_i = dW_i dt = 0$ and $dW_i dW_i = dt$.

The SDE approximation of a density dependent MJP can be rigorously derived in a number of ways. These include proving the convergence of the MJP Kolmogorov master equation to its SDE counterpart, convergence of the infinitesimal generator, and through a number of different system size expansions applied to the master equation [78]. An intuitive approach yielding an equivalent result, referred to as the *Langevin approach*, postulates that the SDE approximation is obtained by matching the infinitesimal moments of the diffusion to those of the MJP [78, 88, 216].

We proceed by approximating the numbers of infections and recoveries in a small time interval, $(t, t + dt]$. Suppose that we can choose dt so that the following *leap* conditions hold:

1. dt is sufficiently *small* that the \mathbf{X}^c is essentially unchanged over $(t, t + dt]$, so that the rates of infections and recoveries, $\boldsymbol{\lambda}(\mathbf{X}^c(t)) = (\beta S(t)I(t), \mu I(t))$, are approximately constant over the interval, i.e.,

$$\boldsymbol{\lambda}(\mathbf{X}^c(t')) \approx \boldsymbol{\lambda}(\mathbf{x}^c(t)), \quad \forall t' \in (t, t + dt]. \quad (2.18)$$

2. dt is sufficiently *large* that we can expect many disease state transitions of each type:

$$\boldsymbol{\lambda}(\mathbf{x}^c(t)) \gg \mathbf{1}. \quad (2.19)$$

Condition (2.18) can be trivially satisfied by choosing dt to be infinitesimally small, and implies that the numbers of infections and recoveries in $(t, t + dt]$ are essentially independent of one another since the rates at which they occur are approximately constant within the interval [88]. Furthermore, (2.18) implies that the numbers of infections and recoveries in the interval are independent Poisson random variables with rates $\boldsymbol{\lambda}(\mathbf{x}^c(t)dt)$. For the SIR model, the numbers of infections and recoveries in an infinitesimal time increment are $N_{SI}(dt) \sim \text{Poisson}(\beta S(t)I(t)dt)$ and $N_{IR}(dt) \sim \text{Poisson}(\mu I(t)dt)$. Condition (2.19), which we can reasonably expect to be satisfied in large populations where there are many infections and recoveries [216], implies that the Poisson distributed increments can be well approximated by independent Gaussian random variables.

We now give the SDE approximation for the SIR model. Let $\boldsymbol{\lambda}(\mathbf{X}) = (\lambda_{SI}, \lambda_{IR})$ denote the rates at which individuals become infected and recover, and let \mathbf{A} denote the matrix whose rows specify changes in counts of susceptible, infected, and recovered individuals corresponding to one infection or recovery event:

$$\mathbf{A} = \begin{array}{ccc} & S & I & R \\ \xrightarrow{S \rightarrow I} & -1 & 1 & 0 \\ \xrightarrow{I \rightarrow R} & 0 & -1 & 1 \end{array}.$$

The SDE for the SIR model is

$$d\mathbf{X}(t) = \mathbf{A}^T \boldsymbol{\lambda}(\mathbf{X}) dt + \sqrt{\mathbf{A}^T \text{diag}(\boldsymbol{\lambda}(\mathbf{X})) \mathbf{A}} d\mathbf{W}_t. \quad (2.20)$$

This SDE is referred to as the *chemical Langevin equation* (CLE). If we wanted to change the model dynamics, the expressions for \mathbf{A} and $\boldsymbol{\lambda}$ would differ, but the diffusion approximation would be of the same form (with the caveat that $\boldsymbol{\lambda}$ must satisfy certain Lipschitz conditions to ensure existence of the SDE, see [78, 164]).

Linear noise approximation

The intractability of the CLE transition density is problematic when the CLE is used as a basis for inference. Absent additional simplifying assumptions, e.g., as in [44], inference with SDEs typically relies on simulation based methods, e.g., [64, 92, 94]. An attractive alternative, dating to at least the 1970s [129, 130], but that has received attention in recent years, involves approximating the CLE by a Gaussian state space model, the moments of which are solutions to systems of ODEs that are related to the transition rates of the MJP and its SDE approximation. This approximation is known as the *linear noise approximation* (LNA), and it will be the methodological basis for Chapters 4 and 5 of this dissertation. Our informal derivation of the LNA follows [92, 222]. Rigorous derivations can be found in [67, 132, 210, 216].

We take the SDE (2.20) as our starting point and express the transition rates in terms of intensive concentrations. Letting $\tilde{\mathbf{X}} \equiv \tilde{\mathbf{X}}(t) = \mathbf{X}(t)/N$ and suppressing the dependence on $\boldsymbol{\theta}$, we have $\boldsymbol{\lambda}(\mathbf{X}) = N\lambda(\tilde{\mathbf{X}})$. The rescaled CLE is

$$\begin{aligned} d\mathbf{X}_t &= N\mathbf{A}^T \boldsymbol{\lambda}(\tilde{\mathbf{X}}_t)dt + \sqrt{N\mathbf{A}^T \text{diag}(\boldsymbol{\lambda}(\tilde{\mathbf{X}}_t))\mathbf{A}}d\mathbf{W}_t \\ \implies d\tilde{\mathbf{X}}_t &= \mathbf{A}^T \boldsymbol{\lambda}(\tilde{\mathbf{X}}_t)dt + \frac{1}{\sqrt{N}}\sqrt{\mathbf{A}^T \text{diag}(\boldsymbol{\lambda}(\tilde{\mathbf{X}}_t))\mathbf{A}}d\mathbf{W}_t. \end{aligned} \quad (2.21)$$

In the infinite population limit, the stochastic contribution to this SDE is negligible and we obtain the deterministic ODE limit of the intensive process, $\tilde{\boldsymbol{\eta}}$, which solves the ODE,

$$\frac{d\tilde{\boldsymbol{\eta}}(t)}{dt} = \mathbf{A}^T \boldsymbol{\lambda}(\tilde{\boldsymbol{\eta}}(t)), \quad \tilde{\boldsymbol{\eta}}(t_0) = \tilde{\boldsymbol{\eta}}_0. \quad (2.22)$$

When stochasticity not negligible and the population size is large, i.e., when we think the leap conditions (2.18) and (2.19) are satisfied, we might reasonably expect \mathbf{X}_t to behave like its deterministic limit plus residual Poisson variation. Thus, we decompose

$$\mathbf{X}_t = N\tilde{\boldsymbol{\eta}}_t + \sqrt{N}\widetilde{\mathbf{M}}_t, \quad (2.23)$$

i.e., $\widetilde{\mathbf{M}}_t = \sqrt{N}(\widetilde{\mathbf{X}}_t - \widetilde{\boldsymbol{\eta}}_t)$, and assume $\|\widetilde{\mathbf{X}}_t - \widetilde{\boldsymbol{\eta}}_t\| = \mathcal{O}(N^{-1/2})$. Substituting (2.23) into (2.21),

$$d\widetilde{\boldsymbol{\eta}}_t + \frac{1}{\sqrt{N}}d\widetilde{\mathbf{M}}_t = \mathbf{A}^T \boldsymbol{\lambda} \left(\widetilde{\boldsymbol{\eta}}_t + \frac{1}{\sqrt{N}}\widetilde{\mathbf{M}}_t \right) dt + \frac{1}{\sqrt{N}}\sqrt{\mathbf{A}^T \text{diag} \left(\boldsymbol{\lambda} \left(\widetilde{\boldsymbol{\eta}}_t + \frac{1}{\sqrt{N}}\widetilde{\mathbf{M}}_t \right) \right) \mathbf{A} d\mathbf{W}_t}. \quad (2.24)$$

Taylor expanding $\boldsymbol{\lambda} \left(\widetilde{\boldsymbol{\eta}}_t + \frac{1}{\sqrt{N}}\widetilde{\mathbf{M}}_t \right)$ about $\widetilde{\boldsymbol{\eta}}_t$, and collecting terms of order $\mathcal{O}(N^{-1})$ and higher, gives

$$\boldsymbol{\lambda} \left(\widetilde{\boldsymbol{\eta}}_t + \frac{1}{\sqrt{N}}\widetilde{\mathbf{M}}_t \right) = \boldsymbol{\lambda}(\widetilde{\boldsymbol{\eta}}_t) + \frac{1}{\sqrt{N}}\widetilde{\mathbf{F}}_t \widetilde{\mathbf{M}}_t + \mathcal{O}(N^{-1}),$$

where $\widetilde{\mathbf{F}}$ is the Jacobian of $\boldsymbol{\lambda}(\cdot)$ evaluated at $\widetilde{\boldsymbol{\eta}}_t$. We substitute 2.24 and again discard terms of order $\mathcal{O}(N^{-1})$,

$$\begin{aligned} d\widetilde{\boldsymbol{\eta}}_t + \frac{1}{\sqrt{N}}d\widetilde{\mathbf{M}}_t &\approx \mathbf{A}^T \left(\boldsymbol{\lambda}(\widetilde{\boldsymbol{\eta}}_t) + \frac{1}{\sqrt{N}}\widetilde{\mathbf{F}}_t \widetilde{\mathbf{M}}_t \right) dt \\ &\quad + \frac{1}{\sqrt{N}}\sqrt{\mathbf{A}^T \text{diag} \left(\boldsymbol{\lambda}(\widetilde{\boldsymbol{\eta}}_t) + \frac{1}{\sqrt{N}}\widetilde{\mathbf{F}}_t \widetilde{\mathbf{M}}_t \right) \mathbf{A} d\mathbf{W}_t} \\ &= \mathbf{A}^T \left(\boldsymbol{\lambda}(\widetilde{\boldsymbol{\eta}}_t) + \frac{1}{\sqrt{N}}\widetilde{\mathbf{F}}_t \widetilde{\mathbf{M}}_t \right) dt + \frac{1}{\sqrt{N}}\sqrt{\mathbf{A}^T \text{diag}(\boldsymbol{\lambda}(\widetilde{\boldsymbol{\eta}}_t)) \mathbf{A} d\mathbf{W}_t} + \mathcal{O}(N^{-1}) \\ &= d\widetilde{\boldsymbol{\eta}}_t + \frac{1}{\sqrt{N}}\mathbf{A}^T \widetilde{\mathbf{F}}_t \widetilde{\mathbf{M}}_t dt + \frac{1}{\sqrt{N}}\sqrt{\mathbf{A}^T \text{diag}(\boldsymbol{\lambda}(\widetilde{\boldsymbol{\eta}}_t)) \mathbf{A} d\mathbf{W}_t}, \end{aligned}$$

where the last equation follows from (2.22). Hence,

$$d\widetilde{\mathbf{M}}_t = \mathbf{A}^T \widetilde{\mathbf{F}}_t \widetilde{\mathbf{M}}_t dt + \sqrt{\mathbf{A}^T \text{diag}(\boldsymbol{\lambda}(\widetilde{\boldsymbol{\eta}}_t)) \mathbf{A} d\mathbf{W}_t}. \quad (2.25)$$

This final equation is the LNA; it is a linear SDE for the residual process in the decomposition (2.23). Therefore, the CLE (2.21) is approximated by the sum of a non-linear deterministic ODE for its drift, and a linear SDE for the residual variability.

The SDE (2.25) for the residual process is linear in $\widetilde{\mathbf{M}}_t$ with time-inhomogeneous drift and diffusion. For fixed or Gaussian initial conditions, $\widetilde{\mathbf{M}}_0 = \widetilde{\mathbf{m}}_0$, the distribution of $\widetilde{\mathbf{M}}_t | \widetilde{\mathbf{m}}(t_0)$, $t \geq t_0$ is Gaussian,

$$\widetilde{\mathbf{M}}_t | \widetilde{\mathbf{m}}(t_0) \sim MVN(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \quad (2.26)$$

and the moments of (2.26) are obtained by solving the coupled ODEs,

$$\frac{d\tilde{\boldsymbol{\eta}}_t}{dt} = \mathbf{A}^T \boldsymbol{\lambda}(\tilde{\boldsymbol{\eta}}_t), \quad \tilde{\boldsymbol{\eta}}(t_0) = \tilde{\boldsymbol{\eta}}_0, \quad (2.27)$$

$$\frac{d\boldsymbol{\mu}_t}{dt} = \mathbf{A}^T \tilde{\mathbf{F}}_t \boldsymbol{\mu}_t, \quad \boldsymbol{\mu}(t_0) = \tilde{\mathbf{m}}_0, \quad (2.28)$$

$$\frac{d\boldsymbol{\Sigma}_t}{dt} = \mathbf{A} \tilde{\mathbf{F}}_t \boldsymbol{\Sigma}_t + \mathbf{A}^T \text{diag}(\boldsymbol{\lambda}(\tilde{\boldsymbol{\eta}}_t)) \mathbf{A} + \boldsymbol{\Sigma}_t \tilde{\mathbf{F}}_t^T \mathbf{A}^T, \quad \boldsymbol{\Sigma}(t_0) = \boldsymbol{\sigma}_0. \quad (2.29)$$

Hence,

$$\tilde{\mathbf{X}}_t | \tilde{\mathbf{x}}(t_0) \sim MVN(\tilde{\boldsymbol{\eta}}_t + \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t). \quad (2.30)$$

Note that we have suppressed the dependence of $\tilde{\mathbf{F}}$ and $\boldsymbol{\Sigma}$ on $\tilde{\boldsymbol{\eta}}$ for clarity. Derivations of the LNA solution may be found in [210, 216, 219]. In general, the LNA ODEs (2.27), (2.28), and (2.29), need to be solved numerically.

2.1.5 Inference and Computation for Stochastic Epidemic Models

Inference for SEMs based on CTMC representations of the epidemic process has historically relied on four, not mutually exclusive, classes of methods [167]: martingale methods, simulation-based methods, approximation of the CTMC, and data augmentation.

Martingale methods estimate the parameters of interest using estimating equations based on martingales for counting processes embedded within the SEM, e.g., for infections and recoveries [9, 17, 136, 141, 198]. However, these methods are not easily implemented for SEMs with complex dynamics fit to partially observed count data.

Simulation based methods use the SEM to generate latent epidemic paths that serve as the basis for inference. This class of methods includes approximate Bayesian computation (ABC) methods [147, 206, 149], pseudo-marginal methods [148, 194], and sequential Monte Carlo (or particle filter) methods [10, 64, 94, 112, 127, 206]. Within this class, the particle marginal Metropolis–Hastings (PMMH) algorithm of [10] stands out as a general method for Bayesian inference and is used as a benchmark method in Chapter 3. Although simulation-based methods have been used to fit complex models, the computational cost of simulating from CTMCs can become prohibitive for complex models. Furthermore, simulation-based

methods suffer from well known pitfalls. ABC methods are sensitive to the choice of summary statistic, rejection threshold, and prior [206]. Sequential Monte Carlo methods, on which pseudo-marginal methods often rely, are prone to “particle impoverishment” problems [41, 64]. Examples of particle degeneracy are presented in supplementary Chapter A, and an inability to fit models with complex dynamics with reasonable effort led us to abandon PMMH as a benchmark for the more complex models in Chapters 4 and 5.

Approximation methods replace the CTMC representation of a SEM with a model whose likelihood is more tractable. A common simplification is to discretize time and to construct a transition model for the population flow between model compartments over discrete time intervals. Discrete time SEMs are often derived from chain-binomial models where, most commonly, a binomially distributed subset of the susceptible population becomes infected over each time step. The most famous examples are the Greenwood model [96], where the number of newly infected individuals is directly modeled as a binomial sample of the susceptible population, and the Reed–Frost epidemic model [2], where binomially distributed incidence depends on the probability that each susceptible escapes infection. Typically, infected individuals are assumed to recover in a fixed number of time steps. It is also possible to construct models with more general dynamics, for example, by allowing for latent periods or random infectious period durations, [138]. When the probability of infection is low, it is possible to approximate binomial counts using Poisson or negative binomial distributions [103, 172]. A related class of discrete time models, referred to as time series susceptible–infected–recovered (TSIR) models, derives conditionally Poisson or negative binomial incidence distributions from birth–death processes [22, 71, 70, 90]. An overview of discrete time epidemic models is given in [215].

Discrete time SEMs are attractive due to their relatively low computational cost. This is particularly true when it is reasonable to assume that underreporting is negligible (and hence ignored) since the observed data likelihood decomposes into products of standard probability distributions and thus requires little in the way of specialized machinery for MCMC or likelihood based methods. Further approximation by Poissons or negative binomials provides

even more computational advantage, even when the dynamics are temporally or spatially complicated [16, 103, 75, 151]. In settings where incidence is under-reported, it may be reasonable to estimate an inflation factor for the observed count [71, 215]. Another approach is to integrate over the true incidence using sequential Monte Carlo or particle filter methods [64, 113, 112, 194]. However, as previously discussed, simulation-based methods can quickly become impractical for complex models.

Among continuous-time approximations, the ODE and SDE representations are, arguably, the most common. ODE models are typically quite easy to work with from a computational standpoint, and MCMC can proceed in a relatively straightforward manner without relying on specialized machinery. Inference based on the SDEs typically relies on either simulation-based computational tools, as discussed in the previous paragraph, or further simplification of the SEM. For example, [43, 44, 185] use diffusion processes that approximate the SEM dynamics, while [116] use a Gaussian process approximation of a related gravity model. As is the case for any approximation, the price of improved computational efficiency and tractability is that simplifying assumptions used in the various approximations must be justified. For instance, the diffusion approximation may not be valid in small populations where the system is far from its deterministic limit (and in which case we should be even more skeptical of the validity of the ODE approximation) [9].

The LNA has been a fixture in the literature on biochemical reaction networks since at least the 1970s when it appeared in a series of papers on approximations of density dependent Markov jump processes [129, 130]. The LNA has been broadly applied in the analysis of gene regulatory networks, e.g., [69, 85, 106, 128, 196, 201, 229]. A review of the LNA, along with related approximations that derive from various system size expansions of the MJP Kolmogorov forward equation, can be found in [190, 216]. The LNA has recently found applications in outbreak modeling in [23, 68, 93, 95, 187, 186, 178, 230]. To our knowledge, all applications of the LNA to outbreak modeling have assumed that the prevalence or cumulative incidence is normally distributed, or has used simulation based methods to incorporate non-Gaussian emission distributions into the model. We regard both of these

constraints on the emission distribution as limitations. Comparison of the LNA with other moment–closure approximations in [98] and [38], who benchmarked the LNA in the context of the SIR model. In this dissertation, we will use the a restarting version of the LNA where the ODEs of the LNA transition density are restarted as data accumulates. Resetting the LNA ODEs has been established to improve accuracy over long time intervals [68, 77, 85].

Finally, agent–based data augmentation (DA) methods for fitting SEMs, first presented in [86, 168], target the joint posterior distribution of the missing data and model parameters to obtain a tractable complete data likelihood. That the augmentation is agent–based refers to the introduction of subject–level disease histories, rather than population–level epidemic paths, as latent variables in the model. An advantage of the agent–based approach is that household structure and subject–level covariates may be incorporated into the model [14, 110, 43, 161, 166]. However, agent–based DA MCMC algorithms have traditionally relied on data–agnostic trans–dimensional proposals that can suffer from MCMC mixing and convergence problems as the fraction of missing information becomes large [185, 148, 174]. This is problematic in the context of epidemic count data, and so classical agent–based DA algorithms have been primarily used to analyze subject–level epidemic data. Development of DA methods for SEMs is of continuing interest, both in settings where the data consist of aggregate counts [174, 175, 193], and when the data reflect subject–level transition events [133, 225].

2.2 Bayesian Computation and Markov Chain Monte Carlo

The objective of Bayesian inference is to quantify uncertainty about parameters of interest, $\boldsymbol{\theta}$, given data, \mathbf{Y} . Uncertainty in the Bayesian paradigm is quantified through the posterior distribution,

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) = \frac{\pi(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{Y})},$$

where $\pi(\mathbf{Y}|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$ are the sampling distribution and the prior distribution. The marginal distribution of the data, $\pi(\mathbf{Y}) = \int \pi(\mathbf{Y}|\boldsymbol{\theta})d\boldsymbol{\theta}$, is constant with respect to $\boldsymbol{\theta}$, and so

we typically work with the unnormalized posterior,

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) \propto \pi(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

Having computed a posterior, we define a Bayes estimator as a decision rule that minimizes the expected posterior loss, or Bayes risk, $E_{\boldsymbol{\theta}|\mathbf{Y}}(L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}))$, i.e.,

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}'} \int L(\boldsymbol{\theta}, \boldsymbol{\theta}'|\mathbf{Y})\pi(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta}.$$

Some common Bayes estimators are the posterior mean, which minimizes the expected squared error loss, and the posterior median, which minimizes expected absolute error. More broadly, we are interested in the posterior distribution of a function, $f(\boldsymbol{\theta})$,

$$E_{\boldsymbol{\theta}|\mathbf{Y}}(f(\boldsymbol{\theta})) = \int f(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta}.$$

This integral is frequently intractable. However, we can approximate it numerically by Monte Carlo integration. Suppose $\boldsymbol{\theta}_i \stackrel{i.i.d.}{\sim} \pi(\boldsymbol{\theta}|\mathbf{Y})$ and $E_{\boldsymbol{\theta}|\mathbf{Y}}(f(\boldsymbol{\theta})) < \infty$. Then, by the strong law of large numbers (SLLN),

$$\begin{aligned}\hat{\mu}_n &= \frac{1}{n} \sum_{i=1}^n f(\boldsymbol{\theta}_i) \xrightarrow{a.s.} E_{\boldsymbol{\theta}|\mathbf{Y}}(f(\boldsymbol{\theta})), \quad n \rightarrow \infty \\ \hat{\sigma}_n^2 &= \frac{1}{n} \sum_{i=1}^n (f(\boldsymbol{\theta}_i) - \hat{\mu}_n)^2 \xrightarrow{a.s.} \operatorname{Var}_{\boldsymbol{\theta}|\mathbf{Y}}(f(\boldsymbol{\theta})).\end{aligned}$$

Hence, we approximate $E_{\boldsymbol{\theta}|\mathbf{Y}}(f(\boldsymbol{\theta}))$ by the Monte Carlo estimate, $\hat{\mu}_n$. The distribution of the ordinary Monte Carlo estimate is given to us by the central limit theorem (CLT),

$$\frac{\hat{\mu}_n - E_{\boldsymbol{\theta}|\mathbf{Y}}(f(\boldsymbol{\theta}))}{\hat{\sigma}_n/\sqrt{n}} \xrightarrow{\mathcal{L}} Z \sim N(0, 1),$$

as $\operatorname{Var}(\hat{\mu}_n) \approx \operatorname{Var}\left(\frac{1}{n} \sum_{i=1}^n f(\boldsymbol{\theta}_i)\right) = \frac{1}{n^2} \sum_{i=1}^n \operatorname{Var}(f(\boldsymbol{\theta}_i)) \implies \operatorname{Var}(\hat{\mu}_n) = \hat{\sigma}_n^2/n$. The quantity, $\hat{\sigma}_n/n$, is called the Monte Carlo standard error (MCSE).

2.2.1 Markov Chain Monte Carlo

MCMC is a framework for numerical integration that can be used to obtain an approximate sample from a posterior that we cannot evaluate analytically and from which it is not possible to sample directly via ordinary Monte Carlo. The strategy is to construct an ergodic

(aperiodic and irreducible) Markov chain, $\{\boldsymbol{\theta}_n\}$, whose stationary distribution is $\pi(\boldsymbol{\theta}|\mathbf{Y})$. The transition kernel of a Markov chain, $K(\boldsymbol{\theta}, \boldsymbol{\theta}')$, gives the probability of moving from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}'$, which are elements in the support of the stationary distribution. The Markov chain preserves the stationary distribution if it satisfies the global balance condition,

$$\pi(\boldsymbol{\theta}) = \int K(\boldsymbol{\theta}', \boldsymbol{\theta})\pi(\boldsymbol{\theta}')d\boldsymbol{\theta}'.$$

It is typically difficult to directly confirm that a transition kernel satisfies global balance. Fortunately, we verify global balance by proving that the kernel satisfies a stronger condition called detailed balance,

$$\pi(\boldsymbol{\theta})K(\boldsymbol{\theta}', \boldsymbol{\theta}) = \pi(\boldsymbol{\theta}')K(\boldsymbol{\theta}, \boldsymbol{\theta}'), \quad \forall \boldsymbol{\theta}, \boldsymbol{\theta}'.$$

Though samples drawn from the Markov chain are correlated, the ergodic theorem ensures that the ergodic mean of an integrable function will converge, almost surely, to its target,

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n f(\boldsymbol{\theta}_i) \xrightarrow{a.s.} \mathbb{E}_{\boldsymbol{\theta}|\mathbf{Y}}(f(\boldsymbol{\theta})).$$

Due to the autocorrelation of MCMC samples, the estimated Monte Carlo variance is now

$$\hat{\sigma}_n^2 = \frac{1}{n} \frac{1}{\hat{n}_{eff}} \sum_{i=1}^n (f(\boldsymbol{\theta}_i) - \hat{\mu}_n)^2,$$

where $\hat{n}_{eff} = n / (1 + 2 \sum_{\ell=1}^{\infty} \hat{\rho}(\ell))$, and $\hat{\rho}(\ell)$ is the estimated autocorrelation at lag ℓ .

It can be challenging to design an MCMC scheme that explores the posterior efficiently. The theory of MCMC merely guarantees that samples from a Markov chain will approximate the target posterior if we run the Markov chain long enough. In practice, the choice of MCMC transition kernel, along with the model parameterization, are critically important to the computational efficiency of the MCMC and, moreover, to the fidelity of the MCMC approximation to the target posterior [19]. We now present several MCMC algorithms that will be used throughout this dissertation.

Metropolis–Hastings sampler

One of the most important MCMC samplers, due to its simplicity and generality, is the Metropolis–Hastings (MH) algorithm [100, 150]. At each step in the Markov chain, a new state is proposed from the distribution $q(\boldsymbol{\theta}^{prop}|\boldsymbol{\theta}^{cur})$, and is accepted with probability

$$\alpha_{\boldsymbol{\theta}^{cur} \rightarrow \boldsymbol{\theta}^{prop}} = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}^{prop}|\mathbf{Y})}{\pi(\boldsymbol{\theta}^{cur}|\mathbf{Y})} \frac{q(\boldsymbol{\theta}^{cur}|\boldsymbol{\theta}^{prop})}{q(\boldsymbol{\theta}^{prop}|\boldsymbol{\theta}^{cur})} \right\} \quad (2.31)$$

$$= \min \left\{ 1, \frac{\pi(\mathbf{Y}|\boldsymbol{\theta}^{prop})\pi(\boldsymbol{\theta}^{prop})}{\pi(\mathbf{Y}|\boldsymbol{\theta}^{cur})\pi(\boldsymbol{\theta}^{cur})} \frac{q(\boldsymbol{\theta}^{cur}|\boldsymbol{\theta}^{prop})}{q(\boldsymbol{\theta}^{prop}|\boldsymbol{\theta}^{cur})} \right\}. \quad (2.32)$$

The Metropolis–Hastings acceptance ratio has a nice interpretation as the ratio of posteriors for the proposed and current state, multiplied by the ratio of proposal probabilities for the reverse and forward transitions. When the next value is proposed from a symmetric distribution centered at the current state, the proposal densities cancel out and the algorithm is referred to as the Metropolis algorithm (this was the original formulation given in [150]). The proposal distribution in the Metropolis algorithm is typically Gaussian.

The computational efficiency of the MH algorithm depends on having a proposal distribution that proposes jumps that are not so large as to never be accepted, but large enough that the sampler is able to explore the parameter space reasonably quickly. Naïvely implemented MH algorithms that iterate through single-site updates to parameters are known to be inefficient as the dimensionality of $\boldsymbol{\theta}$ increases, in settings where parameters are highly correlated, and when there are areas of high curvature in the posterior. All of these factors conspire to prevent the Markov chain from making big jumps as it explores the parameter space. A necessary first step for improving performance is to reparameterize the estimation scale on which the Markov chain explores the parameter space so that the parameters are more nearly multivariate Gaussian, and so that we eliminate problematic boundary conditions. Jointly updating blocks of correlated parameters is also known to improve performance, but relies on having a proposal kernel that captures the covariance structure of the posterior. One strategy obtaining such a kernel is to estimate the posterior covariance matrix from a preliminary run, which is then rescaled by a theoretically optimal factor [81, 182, 183]. Of

course, such a strategy is somewhat tautological as it relies on having a rough guess as to the scales of parameters to begin with. Adaptive algorithms, discussed in Section 2.2.2, attempt to learn the posterior covariance on the fly, and may aim to tune the sampler to achieve a target acceptance rate for proposals [11, 183, 140].

Algorithm 1 Metropolis–Hastings sampler.

- 1: **procedure** METROPOLISHASTINGS($\boldsymbol{\theta}^{cur}$)
- 2: Propose new value: $\boldsymbol{\theta}^{prop} \sim q(\boldsymbol{\theta}^{prop} | \boldsymbol{\theta}^{cur})$
- 3: Compute MH acceptance probability: $\alpha = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}^{prop} | y)}{\pi(\boldsymbol{\theta}^{cur} | y)} \frac{q(\boldsymbol{\theta}^{cur} | \boldsymbol{\theta}^{prop})}{q(\boldsymbol{\theta}^{prop} | \boldsymbol{\theta}^{cur})} \right\}$
- 4: Sample $u \sim Unif(0, 1)$
- 5: Accept/reject proposal:

$$\boldsymbol{\theta}^{new} = \begin{cases} \boldsymbol{\theta}^{prop}, & \text{if } \alpha \geq u, \\ \boldsymbol{\theta}^{cur}, & \text{if } \alpha < u. \end{cases}$$

Gibbs sampler

The Gibbs sampler [84] arises as a special case of MH when we update a parameter using its full conditional distribution, $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{Y})$, given the values of other parameters. MCMC samples obtained from single-site Gibbs updates can become strongly autocorrelated, resulting in poor MCMC mixing. One solution is to jointly update blocks of model parameters using their full conditionals, $\pi(\boldsymbol{\theta}_J | \boldsymbol{\theta}_{-J}, \mathbf{Y})$, ideally grouping together parameters that are strongly correlated and minimizing the correlation between blocks [126, 184, 188].

The MH acceptance probability for the Gibbs sampler is one, thus all Gibbs proposals are automatically accepted. To see why this is so, note that

$$q(\boldsymbol{\theta}_J | \boldsymbol{\theta}_{-J}, \mathbf{Y}) = \frac{\pi(\boldsymbol{\theta} | \mathbf{Y})}{\pi(\boldsymbol{\theta}_{-J} | \mathbf{Y})} \propto \pi(\boldsymbol{\theta} | \mathbf{Y}).$$

Therefore, the MH ratio simplifies as

$$\frac{\pi(\boldsymbol{\theta}_J^{prop}, \boldsymbol{\theta}_{-J}^{cur} | \mathbf{Y}) q(\boldsymbol{\theta}_J^{cur} | \boldsymbol{\theta}_{-J}^{prop}, \mathbf{Y})}{\pi(\boldsymbol{\theta}_J^{cur}, \boldsymbol{\theta}_{-J}^{cur} | \mathbf{Y}) q(\boldsymbol{\theta}_J^{prop} | \boldsymbol{\theta}_{-J}^{cur}, \mathbf{Y})} = \frac{\pi(\boldsymbol{\theta}_J^{prop}, \boldsymbol{\theta}_{-J}^{cur} | \mathbf{Y})}{\pi(\boldsymbol{\theta}_J^{cur}, \boldsymbol{\theta}_{-J}^{cur} | \mathbf{Y})} \frac{\pi(\boldsymbol{\theta}_J^{cur}, \boldsymbol{\theta}_{-J}^{cur} | \mathbf{Y})}{\pi(\boldsymbol{\theta}_J^{prop}, \boldsymbol{\theta}_{-J}^{cur} | \mathbf{Y})} \frac{\pi(\boldsymbol{\theta}_{-J} | \mathbf{Y})}{\pi(\boldsymbol{\theta}_{-J} | \mathbf{Y})} = 1.$$

An attractive property of the Gibbs sampler is that there are no tuning parameters that need to be optimized. This lends some robustness to the algorithm from a user standpoint.

Algorithm 2 Gibbs sampler.

```

1: procedure GIBBS( $\boldsymbol{\theta}^{cur}$ ,  $\mathcal{I} = \{J_1, \dots, J_p\}$  parameter block indices.)
2:   for  $k \in 1, \dots, p$  do
3:     Sample:  $\boldsymbol{\theta}_{J_1}^{new} | \mathbf{Y}, \boldsymbol{\theta}_{J_2, \dots, p}^{cur} \sim \pi(\boldsymbol{\theta}_{J_1} | \mathbf{Y}, \boldsymbol{\theta}_{J_2, \dots, p}^{cur})$ 
4:     Sample:  $\boldsymbol{\theta}_{J_2}^{new} | \mathbf{Y}, \boldsymbol{\theta}_{J_1}^{new}, \boldsymbol{\theta}_{J_3, \dots, p}^{cur} \sim \pi(\boldsymbol{\theta}_{J_2} | \mathbf{Y}, \boldsymbol{\theta}_{J_1}^{new}, \boldsymbol{\theta}_{J_3, \dots, p}^{cur})$ 
5:     :
6:     Sample:  $\boldsymbol{\theta}_{J_p}^{new} | \mathbf{Y}, \boldsymbol{\theta}_{J_1, \dots, p-1}^{new} \sim \pi(\boldsymbol{\theta}_{J_p} | \mathbf{Y}, \boldsymbol{\theta}_{J_1, \dots, p-1}^{new})$ 

```

Slice sampling

Slice sampling methods are a class of auxiliary variable MCMC methods predicated on the idea that sampling from a density proportional to a function, $\pi(\boldsymbol{\theta}) \propto f(\boldsymbol{\theta}) \in \mathbb{R}^n$, is equivalent to sampling uniformly from the $n + 1$ dimensional volume beneath $f(\boldsymbol{\theta})$. These methods occupy an attractive middle ground between simple algorithms that lazily explore the parameter space and may suffer from convergence and mixing issues (e.g., random walk Metropolis–Hastings), and complex gradient based methods that effectively suppress random walks but may be prohibitively expensive due to the need for repeated evaluation of likelihood gradients within each MCMC iteration (e.g., Hamiltonian Monte Carlo). Slice sampling methods have been shown to have attractive robustness properties and are able to suppress random walk behavior that can result in MCMC poor convergence and mixing for high dimensional models [154, 181, 163]. Moreover, slice samplers are flexible and easily

implemented for a broad class of computational problems. Hence, slice samplers will often be our method of choice for fitting complex SEMs.

We introduce slice sampling by considering the task of sampling a random variable, $\theta \in \mathbb{R}$ with density $\pi(\theta) \propto f(\theta)$. Rather than sample θ directly, we introduce an auxiliary variable, u , defined such that the joint density of (θ, u) is uniform over the area $\mathbf{A} = \{(\theta, u) : 0 < u < f(\theta)\}$. Slice sampling alternates Gibbs updates of $\theta|u$ and $u|\theta$ from their full conditionals,

$$\begin{aligned}\pi(u|\theta) &\sim \text{Unif}(0, f(\theta)) \\ \pi(\theta|u) &\sim \text{Unif}(S), \quad S = \{\theta : f(\theta) > u\}.\end{aligned}$$

In the case of slice sampling θ from its posterior, $\pi(\theta|\mathbf{Y}) \propto f(\theta) \equiv L(\mathbf{Y}|\theta)\pi(\theta)$. The algorithm is otherwise unchanged.

Algorithm 3 Univariate slice sampling intuition.

- 1: **procedure** SLICESAMPLER(θ^{cur}, u^{cur})
 - 2: Vertical update: $u^{new}|\theta^{cur} \sim \text{Unif}(0, f(\theta))$,
 - 3: Horizontal update: $\theta^{new}|u^{new} \sim \text{Unif}(S), \quad S = \{\theta : f(\theta) > u^{new}\}$.
-

Sampling $u^{new}|\theta^{cur}$ is trivial. However, we often lack an analytical expression for the slice, $S = \{\theta : f(\theta) > u^{new}\}$, that characterizes to the set of parameter values that have density greater u^{new} . In practice, we use a bracket, S' that, ideally, approximates the width of S , and which we position randomly around θ^{cur} . We then sample uniformly within the bracket until we propose a value θ^{new} that satisfies $f(\theta^{new}) > u^{new}$. Note that S' need not contain all of S . However, if S' is too small, the Markov chain will be unable to take large steps, resulting in highly autocorrelated posterior samples. In contrast, too large a bracket will result in many rejections, and hence many unnecessary likelihood evaluations. Several possibilities for approximating S are listed in [163]. The key condition that must be satisfied in any interval selection procedure in order for the Markov chain to satisfy detailed balance is that $\Pr(S'|\theta^{cur}) = \Pr(S'|\theta^{new})$. Hence, the set of valid proposals, P , is the set of points

for which the probability of selecting an approximate bracket, S' , is equal to the probability of selecting S' from current state, i.e.,

$$P = \{\theta' : \theta' \in S \cap S', \text{ and } \Pr(S'|\theta) = \Pr(S'|\theta')\}.$$

An attractive strategy for maximizing the expected jump distance of the Markov chain is to adapt the bracket by “stepping out” its endpoints until S' covers S . We are then assured that we will be sampling uniformly from as large a bracket as possible. This procedure preserves the stationary distribution as long as the approximate bracket is randomly positioned around θ^{cur} . If the resulting interval is too wide, we can shrink the corresponding bracket endpoint to the rejected value and propose a new value uniformly within the new bracket. We will typically tune the initial bracket width throughout the MCMC run to encourage an equal number of bracket expansions to shrinkages. This minimizes the expected number of likelihood evaluations per MCMC iteration [204].

Algorithm 4 Univariate slice sampling with stepping out.

```

1: procedure SLICESAMPLER1D( $\theta^{cur}$ ,  $S' = (L, U)$ ,  $\omega = U - L$ )
2:    $u \sim \text{Unif}(0, f(\theta^{cur}))$                                       $\triangleright$  Set threshold
3:    $p \sim \text{Unif}(0, 1)$ ;  $L \leftarrow \theta - \omega p$ ,  $U \leftarrow L + \omega$            $\triangleright$  Position  $S'$  around  $\theta$ 
4:    $S' \leftarrow \text{STEPOUT}(u, S')$                                           $\triangleright$  Step out bracket
5:   while  $u < f(L)$  do  $L \leftarrow L - \omega$ 
6:   while  $u < f(U)$  do  $U \leftarrow U + \omega$ 
7:    $\theta^{prop}|u \sim \text{Unif}(L, U)$                                           $\triangleright$  Propose new value
8:   if  $f(\theta^{prop}) > u$  then  $\theta^{new} \leftarrow \theta^{prop}$                        $\triangleright$  Accept proposal
9:   return  $\theta^{new}$ 
10:  else                                                                $\triangleright$  Shrink bracket
11:    if  $\theta^{prop} < \theta^{cur}$  then  $L \leftarrow \theta^{prop}$  else  $U \leftarrow \theta^{prop}$ 
12:    GoTo 7

```

Multivariate normal slice sampling

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. A variety of methods for slice sampling in multiple dimensions are explored in [163, 202, 204]. 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 [3]. Our approach differs in that we typically adapt the proposal covariance matrix using a Robbins–Monro recursion, discussed in the next section. This adaptation is helped by slight modifications to the algorithm during the adaptation phase of the MCMC [11, 140]. 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 [204], which helps to reduce the number of likelihood evaluations per iteration.

As in the previous section, we suppress dependence on the data for notational clarity since slice sampling $\boldsymbol{\theta} \in \mathbb{R}^d$ from its posterior $\pi(\boldsymbol{\theta}|\mathbf{Y})$ is, again, 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 decomposition of $\boldsymbol{\Sigma}$ (any other matrix square root would do). In practice, $\boldsymbol{\Sigma}$ is approximated by $\hat{\boldsymbol{\Sigma}}_n$, 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 $\boldsymbol{\theta}^{prop} = \boldsymbol{\theta}^{cur} + ch(w\xi_1 + (1 - w)\xi_2)$, where $\xi_1 = h(\mathbf{z}_1)$, $\mathbf{z}_1 \sim MVN(\mathbf{0}, \Sigma)$, $\xi_2 = h(\mathbf{z}_2)$, $\mathbf{z}_2 \sim MVN(\mathbf{0}, \mathbf{I}_d)$, $h(\mathbf{z}) = \mathbf{z}/\|\mathbf{z}\|$, 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 5 Multivariate normal slice sampling with stepping out.

```

1: procedure MVNSS( $\boldsymbol{\theta}^{cur}$ ,  $\mathbf{L}$ ,  $S' = (0, \omega)$ )
2:    $u \sim \text{Unif}(0, f(\boldsymbol{\theta}^{cur}))$                                       $\triangleright$  Set threshold
3:    $\mathbf{z} \sim MVN(\mathbf{0}, \mathbf{I}_d)$ ,  $\xi \leftarrow h(\mathbf{Lz})$                                  $\triangleright$  Propose direction
4:    $p \sim \text{Unif}(0, 1)$ ;  $L \leftarrow -\omega p$ ,  $U \leftarrow L + \omega$            $\triangleright$  Position  $S'$  around 0
5:    $S' \leftarrow \text{STEPOUT}(u, S')$                                           $\triangleright$  Step out bracket
6:   while  $u < f(\boldsymbol{\theta} - L\xi)$  do  $L \leftarrow L - \omega$ 
7:   while  $u < f(\boldsymbol{\theta} + U\xi)$  do  $U \leftarrow U + \omega$ 
8:    $c \sim \text{Unif}(L, U)$ ;  $\boldsymbol{\theta}^{prop} \leftarrow \boldsymbol{\theta}^{cur} + c\xi$             $\triangleright$  Propose new value
9:   if  $f(\boldsymbol{\theta}^{prop}) > u$  then  $\boldsymbol{\theta}^{new} \leftarrow \boldsymbol{\theta}^{prop}$             $\triangleright$  Accept proposal
10:  return  $\boldsymbol{\theta}^{new}$ 
11:  else                                                                $\triangleright$  Shrink bracket
12:    if  $c < 0$  then  $L \leftarrow c$  else  $U \leftarrow c$ 
13:    GoTo 8

```

Elliptical slice sampling

The elliptical slice sampler (ElliptSS) of [159] is an algorithm that can be used to efficiently sample from the posterior distribution of model parameters that have a Gaussian prior. The conditions we need for ElliptSS to be valid are that $\mathbf{x} \sim N(\mathbf{0}, \Sigma)$, and that the posterior factors as $\pi(\mathbf{x}|\mathbf{Y}) = \frac{1}{Z}N(\mathbf{x}; \mathbf{0}, \Sigma)L(\mathbf{Y}|\mathbf{x})$. The strategy behind ElliptSS is to reparameterize the model by introducing auxiliary variables in a way that preserves the marginal distribution of \mathbf{x} , and thus preserves the likelihood of the data conditional on \mathbf{x} .

We introduce auxiliary variables, $\boldsymbol{\eta}_0 \sim N(\mathbf{0}, \Sigma)$, $\boldsymbol{\eta}_1 \sim N(\mathbf{0}, \Sigma)$, $\theta \sim \text{Unif}(0, 2\pi)$, and define $\mathbf{x} = \boldsymbol{\eta}_0 \sin(\theta) + \boldsymbol{\eta}_1 \cos(\theta)$. Note that the marginal distribution of \mathbf{x} is still $\mathbf{x} \sim N(\mathbf{0}, \Sigma)$. ElliptSS alternates sampling $\boldsymbol{\eta}_0, \boldsymbol{\eta}_1, \theta$ under the constraint that \mathbf{x} is unchanged, and updating θ by univariate slice sampling. To this end, the model is reparameterized as,

$$\theta \sim \text{Unif}(0, 2\pi), \boldsymbol{\eta} \sim N(\mathbf{0}, \Sigma), \boldsymbol{\eta}_0 = \mathbf{x} \sin(\theta) + \boldsymbol{\eta} \cos(\theta), \boldsymbol{\eta}_1 = \mathbf{x} \cos(\theta) - \boldsymbol{\eta} \sin(\theta).$$

This leaves \mathbf{x} unchanged as $\mathbf{x}^{prop} = \boldsymbol{\eta}_0 \sin(\theta) + \boldsymbol{\eta}_1 \cos(\theta) = \mathbf{x}$. In practice, we need not compute $\boldsymbol{\eta}_1$ since it will just be discarded.

Algorithm 6 Elliptical slice sampler.

```

1: procedure ELLIPTSS( $\mathbf{x}^{cur}$ )
2:    $\boldsymbol{\eta} \sim N(\mathbf{0}, \Sigma)$                                       $\triangleright$  Sample ellipse
3:    $u|\mathbf{x} \sim \text{Unif}(0, L(\mathbf{x}))$                     $\triangleright$  Sample threshold
4:    $\theta \sim \text{Unif}(0, 2\pi); (L_\theta, R_\theta) \leftarrow (\theta - 2\pi, \theta)$      $\triangleright$  Initial angle, center bracket
5:    $\mathbf{x}^{prop} \leftarrow \mathbf{x} \cos(\theta) + \boldsymbol{\eta} \sin(\theta)$             $\triangleright$  Compute proposal
6:   if  $L(\mathbf{x}^{prop}) > u$  then  $\mathbf{x}^{new} \leftarrow \mathbf{x}^{prop}$            $\triangleright$  Accept proposal
7:   return  $\mathbf{x}^{new}$ 
8:   else                                                  $\triangleright$  Shrink bracket
9:     if  $\theta < 0$  then  $L_\theta \leftarrow \theta$  else  $U_\theta \leftarrow \theta$ 
10:     $\theta \sim \text{Unif}(L_\theta, U_\theta)$                           $\triangleright$  Propose new angle
11:    GoTo 5

```

The elliptical slice sampler is a rejection-free algorithm with no free tuning parameters that always perturbs the current state if it is not the only state that has non-zero probability. Its efficiency when \mathbf{x} is high dimensional can be motivated in a number of ways. In high dimensions, the typical set of \mathbf{x} is concentrated in near the surface of a Gaussian hyperellipsoid [19, 159]. Thus, proposals along an elliptical path determined by two points within the ellipse are less likely to escape the typical set than are line proposals (e.g., MVNSS, where slice sampling varies ϵ in the proposal $\mathbf{x}' = \mathbf{x} + \epsilon \boldsymbol{\eta}$, $\boldsymbol{\eta} \sim N(\mathbf{0}, \Sigma)$). Moreover, as the

dimensionality of \mathbf{x} grows large, the proposed ellipse, $\boldsymbol{\eta}$, and \mathbf{x} are likely to be at an angle of $\pi/2$ apart. Hence, if \mathbf{x} is a good explanation for the data, $\boldsymbol{\eta}$ is likely not to be and the line sampler will reject almost all values of $\epsilon > 0$.

As the authors in [159] point out, ElliptSS can also be seen as a slice sampling version of the pre-conditioned Crank-Nicolson (pCN) proposal that was proposed in [162], which was an autoregressive MH proposal of the form $\mathbf{x}' = \mathbf{x}\sqrt{1-\epsilon^2} + \boldsymbol{\eta}\epsilon$, $\epsilon \in [-1, 1]$. pCN proposals are known to have acceptance probabilities that are invariant to the dimension of the parameter space [55]. Finally, as shown in [25], the form of the elliptical slice sampling update corresponds to an analytical expression for the Hamiltonian flow where the potential is normally distributed. The auxiliary variable, $\boldsymbol{\eta}$, is analogous to the initial velocity.

2.2.2 Adaptive MCMC

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 [11].

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 θ^* . The function $f(\theta)$ is not directly observed. Instead, we use a noisy sequence of estimates, $h(\hat{\theta}_n)$, satisfying $E(h(\hat{\theta}_n)) = f(\theta)$ to recursively approximate θ^* . The recursion takes the form,

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

Hence, the recursion increments θ by an amount proportional to the difference between $h(\hat{\theta}_n)$

and its target. The gain factor sequence, $\{\gamma_{n+1}\}$, is a deterministic non-increasing, positive sequence such that

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

Note that since $\gamma_n \rightarrow 0$ and $E(h(\hat{\theta}_n)) = f(\theta)$, it follows that $|\theta_{n+1} - \theta_n| \rightarrow 0$ as $n \rightarrow \infty$, 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

$$\gamma_n = C(1 + pn)^{-\alpha}, \quad \alpha \in (0.5, 1], \quad p > 0 \quad (2.33)$$

will satisfy these conditions [11, 140].

We will typically 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. We now present an adaptive version of the random walk MH algorithm from Section 2.2.1, in which we adapt the proposal covariance and a global scaling parameter that is set to achieve a target acceptance rate of α^* . This is Algorithm 4 in [11]. The proposal covariance matrix for the MVNSS is tuned analogously.

Let μ_n and Σ_n denote the empirical mean and covariance of the posterior samples from the first n MCMC iterations, $\alpha(\boldsymbol{\theta}_{n-1}^{cur}, \boldsymbol{\theta}_n^{prop})$ be the MH acceptance probability for accepting the proposal in iteration n , α^* be the target acceptance rate (typically set to 0.234 based on [182]), and $\{\gamma_n\}$ be a sequence of gain factors. The adaptive MH algorithm with global adaptive scaling is given below.

Algorithm 7 Adaptive Metropolis–Hastings with global adaptive scaling.

1: **procedure** ADAPTIVEMH($\boldsymbol{\theta}_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0$)
2: **for** $n \in 1, \dots, N$ **do**
3: $\boldsymbol{\theta}_n^{prop} \sim MVN(\boldsymbol{\mu}_{n-1}, \lambda_{n-1} \boldsymbol{\Sigma}_{n-1})$ ▷ Propose new value
4: $\alpha(\boldsymbol{\theta}_{n-1}^{cur}, \boldsymbol{\theta}_n^{prop}) = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}_n^{prop}|y)}{\pi(\boldsymbol{\theta}_{n-1}^{cur}|y)} \right\}$ ▷ MH acceptance probability
5: $u \sim Unif(0, 1)$ ▷ Accept/reject proposal

6:
$$\boldsymbol{\theta}_n^{cur} = \begin{cases} \boldsymbol{\theta}_n^{prop}, & \text{if } \alpha \geq u, \\ \boldsymbol{\theta}_{n-1}^{cur}, & \text{if } \alpha < u. \end{cases}$$

6: **Adapt MH kernel:**

$$\begin{aligned} \log(\lambda_n) &= \log(\lambda_{n-1}) + \gamma_n (\alpha(\boldsymbol{\theta}_{n-1}^{cur}, \boldsymbol{\theta}_n^{prop}) - \alpha^*) \\ \boldsymbol{\mu}_n &= \boldsymbol{\mu}_{n-1} + \gamma_n (\boldsymbol{\theta}_n^{cur} - \boldsymbol{\mu}_{n-1}) \\ \boldsymbol{\Sigma}_n &= \boldsymbol{\Sigma}_{n-1} + \gamma_n ((\boldsymbol{\theta}_n^{cur} - \boldsymbol{\mu}_{n-1})(\boldsymbol{\theta}_n^{cur} - \boldsymbol{\mu}_{n-1})^T - \boldsymbol{\Sigma}_{n-1}) \end{aligned}$$

Chapter 3

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

3.1 Overview

In this chapter, we develop an agent-based DA Markov chain Monte Carlo (MCMC) framework for fitting stochastic epidemic models (SEMs) to time series count data of disease prevalence. We obtain a tractable complete data likelihood by augmenting the data with subject-level disease histories. Our MCMC targets the joint posterior distribution of the latent epidemic process and the model parameters as we alternate between updating subject-level paths and model parameters. We propose each new subject-path, conditionally on the data, using a time-inhomogeneous continuous-time Markov chain (CTMC) with rates determined by the disease histories of the other individuals. These data-driven path proposals result in highly efficient perturbations to the latent epidemic path, and enable us to analyze epidemic count data in the absence of any subject-level information. Thus, our MCMC algorithm enables exact Bayesian inference for SEMs fit to datasets that would have been impossible to study with existing agent-based DA methods. Our algorithm is not specific to any particular SEM dynamics or measurement process, and may be applied, with minimal modifications, to a broad class of SEMs.

Section 3.2 presents the DA algorithm in the context of fitting the stochastic Susceptible–Infected–Recovered (SIR) model to binomially distributed prevalence counts. The minimal adaptations required to fit Susceptible–Exposed–Infected–Recovered (SEIR) and Susceptible–Exposed–Infected–Susceptible (SIRS) models are given in Sections 3.2.5 and 3.2.6. Section 3.3 presents a series of simulations that demonstrate the utility of the algorithm for fitting a variety of SEMs, including models where the dynamics or population size are misspecified,

and that investigate the sensitivity of estimates to prior specification. In Section 3.4, we will use our algorithm to fit SIR and SEIR models to data from an outbreak of influenza in a British boarding school (described in Section 1.3.1).

3.2 The Data Augmentation Algorithm for an SIR Model

The SIR model describes the time evolution of an epidemic in terms of the disease histories of individuals as they transition through three states — susceptible (S), infected/infectious (I), and recovered (R). Under simple SIR dynamics, each individual becomes infectious immediately upon becoming infected, and acquires lifelong immunity upon recovery. For simplicity, we assume that the population is closed and mixes homogeneously, and that there is no external force of infection. Therefore, the epidemic ceases once the pool of infectious individuals is depleted.

3.2.1 Measurement process and data

Our data, $\mathbf{Y} = \{Y_1, \dots, Y_L\}$, are disease prevalence counts recorded at times $t_1, \dots, t_L \in [t_1, t_L]$. It should not beggar belief that the data could be subject to measurement error, for example underreporting in settings where asymptomatic individuals escape detection. Let S_τ , I_τ , and R_τ denote the total susceptible, infected, and recovered people at time τ . We model the observed prevalence as a binomial sample, with constant mean case detection probability ρ . Thus,

$$Y_\ell | I_{t_\ell}, \rho \sim \text{Binomial}(I_{t_\ell}, \rho). \quad (3.1)$$

3.2.2 Latent epidemic process

The data are sampled from a latent epidemic process, $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$, that evolves continuously in time as individuals become infected 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 the infection process for a single subject, \mathbf{X}_j , is $\mathcal{S}_j = \{S, I, R\}$, and a

realized subject-path is of the form

$$\mathbf{x}_j(\tau) = \begin{cases} S, & \tau < \tau_I^{(j)}, \\ I, & \tau_I^{(j)} \leq \tau < \tau_R^{(j)}, \\ R, & \tau_R^{(j)} \leq \tau, \end{cases} \quad (3.2)$$

where $\tau_I^{(j)}$ and $\tau_R^{(j)}$ are the infection and recovery times for subject j (though subject j may also never become infected or recover, or may become infected or recover outside of the observation period $[t_1, t_L]$). We write the configuration of \mathbf{X} at time τ as $\mathbf{X}(\tau) = (\mathbf{X}_1(\tau), \dots, \mathbf{X}_N(\tau))$, and adopt the convention that $\mathbf{X}(\tau)$ and derived quantities, e.g., I_τ , depend on the configuration just before τ . We use τ^+ for quantities evaluated just after a particular time. The waiting times between transition events are taken to be exponentially distributed, and we denote by β and μ the per-contact infectivity and recovery rates. Note that $1/\mu$ is interpreted as the mean infectious period duration, and $R_0 = \beta N/\mu$ is the basic reproduction number. The latent epidemic process evolves according to a time-homogeneous CTMC, with transition rate from configuration \mathbf{X} to \mathbf{X}' given by

$$\lambda_{\mathbf{X}, \mathbf{X}'} = \begin{cases} \beta I, & \text{if } \mathbf{X} \text{ and } \mathbf{X}' \text{ differ only in subject } j, \text{ with } \mathbf{X}_j = S, \text{ and } \mathbf{X}'_j = I, \\ \mu, & \text{if } \mathbf{X} \text{ and } \mathbf{X}' \text{ differ only in subject } j, \text{ with } \mathbf{X}_j = I, \text{ and } \mathbf{X}'_j = R, \\ 0, & \text{for all other configurations } \mathbf{X} \text{ and } \mathbf{X}'. \end{cases} \quad (3.3)$$

At the first observation time, we let $\mathbf{X}(t_1)|\mathbf{p}_{t_1} \sim \text{Categorical}(\{S, I, R\}, \mathbf{p}_{t_1})$, where $\mathbf{p}_{t_1} = (p_S, p_I, p_R)$ are the probabilities that an individual is susceptible, infected, or recovered. Let $\boldsymbol{\tau} = \{\tau_0, \dots, \tau_{K+1}\}$, where $t_1 \equiv \tau_0$ and $t_L \equiv \tau_{K+1}$, be the (ordered) set of K infection and recovery times of all individuals along with the endpoints of the observation period $[t_1, t_L]$. Let $\mathbb{1}_{\{\tau_k \trianglelefteq I\}}$ and $\mathbb{1}_{\{\tau_k \trianglelefteq R\}}$ indicate whether τ_k is an infection or recovery time, and let

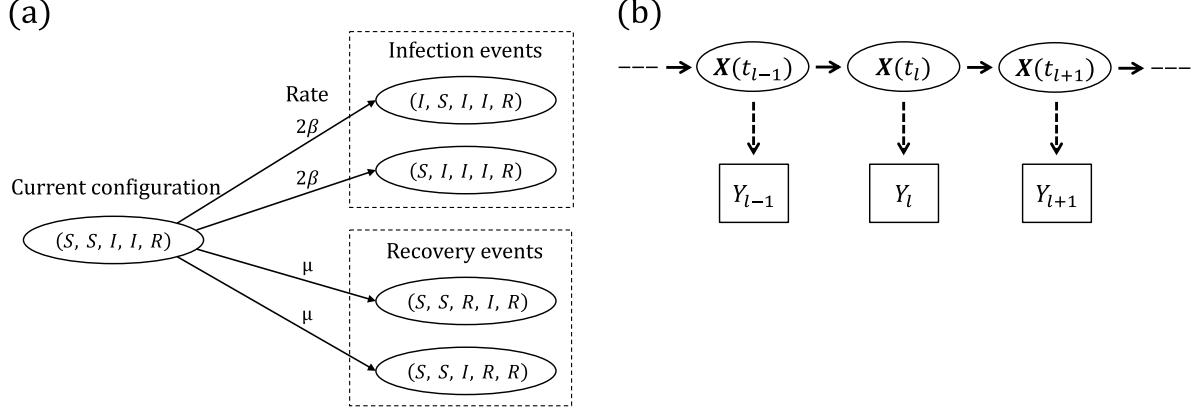


Figure 3.1: (a) SIR dynamics in a population of five subjects. The number of infecteds can increase from two to three via an infection of the first or second subject, reaching each of those configurations at rate 2β . The number of recovered individuals can increase from one to two via a recovery of the third or fourth subject, reaching each of those configurations at rate μ . (b) Hidden Markov model for the joint distribution of the latent epidemic process and the data. The observations, \mathbf{Y}_ℓ , $\ell = 1, \dots, L$, are conditionally independent given $\mathbf{X}(t)$, and $\mathbf{Y}_\ell | I_{t_\ell}, \rho \sim \text{Binomial}(I_{t_\ell}, \rho)$.

$\boldsymbol{\theta} = (\beta, \mu, \rho, \mathbf{p}_{t_1})$ denote the vector of unknown parameters. The complete data likelihood is

$$\begin{aligned}
 L(\mathbf{X}, \mathbf{Y} | \boldsymbol{\theta}) &= \Pr(\mathbf{Y} | \mathbf{X}, \rho) \times \Pr(\mathbf{X}(t_1) | \mathbf{p}_{t_1}) \times \pi(\mathbf{X} | \mathbf{X}(t_1), \beta, \mu) \\
 &= \left[\prod_{l=1}^L \binom{I_{t_\ell}}{Y_\ell} \rho^{Y_\ell} (1-\rho)^{I_{t_\ell}-Y_\ell} \right] \times \left[p_S^{S_{t_1}} p_I^{I_{t_1}} p_R^{R_{t_1}} \right] \\
 &\quad \times \prod_{k=1}^K \left\{ [\beta I_{\tau_k} \times \mathbb{1}_{\{\tau_k \leq I\}} + \mu \times \mathbb{1}_{\{\tau_k \leq R\}}] \exp [-(\tau_k - \tau_{k-1}) (\beta I_{\tau_k} S_{\tau_k} + \mu I_{\tau_k})] \right\} \\
 &\quad \times \exp \left[-(t_L - \tau_K) (\beta I_{\tau_K^+} S_{\tau_K^+} + \mu I_{\tau_K^+}) \right]. \tag{3.4}
 \end{aligned}$$

3.2.3 Subject-path proposal framework

The observed data likelihood in the posterior

$$\pi(\boldsymbol{\theta} | \mathbf{Y}) \propto \pi(\mathbf{Y} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) = \int L(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}) \pi(\mathbf{X} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\mathbf{X}$$

is analytically intractable for even moderately sized N as it involves a high dimensional integral over the collection of subject-paths, \mathbf{X} . DA methods introduce the subject-paths, \mathbf{X} ,

as latent variables in the model that are jointly estimated along with the model parameters, $\boldsymbol{\theta}$. Introducing \mathbf{X} in this way enables us to work with the tractable complete data likelihood, (3.4). The joint posterior distribution is

$$\pi(\boldsymbol{\theta}, \mathbf{X} | \mathbf{Y}) \propto \Pr(\mathbf{Y} | \mathbf{X}, \rho) \times \pi(\mathbf{X} | \mathbf{X}(t_1), \beta, \mu) \times \Pr(\mathbf{X}(t_1) | \mathbf{p}_{t_1}) \times \pi(\beta) \pi(\mu) \pi(\rho) \pi(\mathbf{p}_{t_1}), \quad (3.5)$$

where $\pi(\beta)$, $\pi(\mu)$, $\pi(\rho)$, and $\pi(\mathbf{p}_{t_1})$ are prior densities. Our MCMC targets the joint posterior distribution, given by (3.5), as we alternate between updating $\mathbf{X} | \boldsymbol{\theta}, \mathbf{Y}$ and $\boldsymbol{\theta} | \mathbf{X}, \mathbf{Y}$.

Given the current collection of subject–paths, \mathbf{x}^{cur} , we propose \mathbf{x}^{new} by sampling the path of a single subject \mathbf{X}_j , conditionally on the data, using a time–inhomogeneous CTMC with state space \mathcal{S}_j and rates conditioned on the collection of disease histories of other individuals, $\mathbf{x}_{(-j)} = \{\mathbf{x}_1, \dots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \dots, \mathbf{x}_N\}$. The proposed collection of paths is accepted or rejected in a Metropolis–Hastings step.

Let $\boldsymbol{\tau}^{(j)} = \{\tau_I^{(j)}, \tau_R^{(j)}\}$ be the (possibly empty) set of infection and recovery times for subject j , and define $\boldsymbol{\tau}^{(-j)} = \{\boldsymbol{\tau} \setminus \boldsymbol{\tau}^{(j)}\} = \{\tau_0^{(-j)}, \tau_1^{(-j)}, \dots, \tau_M^{(-j)}, \tau_{M+1}^{(-j)}\}$, where $t_1 \equiv \tau_0^{(-j)}$ and $t_L \equiv \tau_{M+1}^{(-j)}$, to be the set of $M \leq K$ (ordered) times at which other subjects become infected or recover, along with t_1 and t_L . Let $\mathcal{I} = \{\mathcal{I}_1, \dots, \mathcal{I}_{M+1}\}$ be the intervals that partition $[t_1, t_L]$, i.e. $\mathcal{I}_1 = [\tau_0^{(-j)}, \tau_1^{(-j)}], \mathcal{I}_2 = [\tau_1^{(-j)}, \tau_2^{(-j)}], \dots, \mathcal{I}_{M+1} = [\tau_M^{(-j)}, \tau_{M+1}^{(-j)}]$. Let $I_\tau^{(-j)} = \sum_{i \neq j} \mathbf{1}_{\{\mathbf{x}_i(\tau)=I\}}$ be the prevalence at time τ , excluding subject j . Let $\boldsymbol{\Lambda}^{(-j)}(\boldsymbol{\theta}) = \{\boldsymbol{\Lambda}_1^{(-j)}(\boldsymbol{\theta}), \dots, \boldsymbol{\Lambda}_{M+1}^{(-j)}(\boldsymbol{\theta})\}$ be the sequence of rate matrices corresponding to each interval in \mathcal{I} , where for $m = 1, \dots, M + 1$,

$$\boldsymbol{\Lambda}_m^{(-j)}(\boldsymbol{\theta}) = \begin{matrix} & S & I & R \\ & \begin{matrix} -\beta I_{\tau_m}^{(-j)} & \beta I_{\tau_m}^{(-j)} & 0 \end{matrix} \\ I & \begin{matrix} 0 & -\mu & \mu \end{matrix} \\ R & \begin{matrix} 0 & 0 & 0 \end{matrix} \end{matrix}. \quad (3.6)$$

We can construct the transition probability matrix for subject j over interval I_m ,

$$\mathbf{P}^{(j)}(\tau_{m-1}, \tau_m) = \left(p_{a,b}^{(j)}(\tau_{m-1}, \tau_m) \right)_{a,b \in \mathcal{S}_j},$$

where $p_{a,b}^{(j)}(\tau_{m-1}, \tau_m) = \Pr(\mathbf{X}_j(\tau_m) = b | \mathbf{X}_j(\tau_{m-1}) = a, \boldsymbol{\theta})$, using the matrix exponential

$$\mathbf{P}^{(j)}(\tau_{m-1}, \tau_m) = \exp \left[(\tau_m - \tau_{m-1}) \boldsymbol{\Lambda}_m^{(-j)}(\boldsymbol{\theta}) \right].$$

This computation requires an eigen-decomposition of each rate matrix. We may reduce the total computational burden by computing the eigen decompositions analytically, and by caching the decompositions to avoid duplicate computations. One additional point is that while the eigenvalues of any SIR rate matrix are always real valued, this is not generally true, e.g., it is possible for the rate matrix of an SIRS model to have complex eigenvalues. In this case, we obtain a real valued transition probability matrix by first applying a rotation to each rate matrix with complex eigenvalues to obtain its real canonical form [107]. This is discussed in Section A.1.

By the Markov property, the time-inhomogeneous CTMC density over the observation period $[t_1, t_L]$, denoted $\pi(\mathbf{X}_j | \mathbf{x}_{(-j)}, \boldsymbol{\theta}) \equiv \pi \left(\mathbf{X}_j | \boldsymbol{\Lambda}^{(-j)}(\boldsymbol{\theta}); \mathcal{I} \right)$, can be written as a product of time-homogeneous CTMC densities over the inter-event intervals $\mathcal{I}_1, \dots, \mathcal{I}_M$. Hence,

$$\pi \left(\mathbf{X}_j | \boldsymbol{\Lambda}^{(-j)}; \mathcal{I} \right) = \Pr(\mathbf{X}_j(t_1) | \mathbf{p}_{t_1}) \prod_{m=1}^M \pi \left(\mathbf{X}_j | \mathbf{x}_j(\tau_{m-1}), \boldsymbol{\Lambda}_m^{(-j)}(\boldsymbol{\theta}); \mathcal{I}_m \right). \quad (3.7)$$

Similarly, the transition probability matrix over an interval $\mathcal{I}_\ell = [t_{\ell-1}, t_\ell]$ can be written as the product of transition probability matrices over the sub-intervals in \mathcal{I}_ℓ , within which the subject-level CTMC is time-homogeneous. Thus, the transition probability matrix over an inter-observation interval, $\mathcal{I}_\ell = [t_{\ell-1}, t_\ell]$, partitioned by S transition events that define inter-event intervals with endpoints given by times $t_{\ell-1} \equiv \tau_{\ell,0}^{(-j)} < \tau_{\ell,1}^{(-j)} < \dots < \tau_{\ell,S-1}^{(-j)} < \tau_{\ell,S}^{(-j)} \equiv t_\ell$, is constructed as

$$\mathbf{P}^{(j)}(t_{\ell-1}, t_\ell) = \prod_{s=1}^S \mathbf{P}^{(j)} \left(\tau_{\ell,s-1}^{(-j)}, \tau_{\ell,s}^{(-j)} \right).$$

The algorithm for constructing a subject-path proposal proceeds in three steps, diagrammed in Figure 3.2:

1. *HMM step*: sample the disease state of the subject under consideration at the observation times, conditional on the data and disease histories of other subjects.

2. *Discrete time skeleton step*: sample the state at times when the time-inhomogeneous CTMC rates change, conditional on the states sampled in the HMM step.
3. *Event time step*: sample the exact times of transition events conditional on the sequence of states sampled in the previous steps.

HMM step

The key to sampling a sequence of disease states at the observation times is to rewrite the emission probability, given by (3.1), as

$$Y_\ell | X_j(t_\ell), I_{t_\ell}^{(-j)}, \rho \sim \text{Binomial} \left(\mathbf{1}_{\{X_j(t_\ell)=I\}} + I_{t_\ell}^{(-j)}, \rho \right). \quad (3.8)$$

If we treat the paths of all subjects except individual j as fixed, the emission probability in (3.8) only varies depending on whether subject j is infected at time t_ℓ . Furthermore, the data are conditionally independent of one another, given \mathbf{x} and $\boldsymbol{\theta}$, which induces a hidden Markov model (HMM) over the joint distribution \mathbf{X} and \mathbf{Y} (Figure 3.1b).

We sample the state of \mathbf{X}_j at times t_1, \dots, t_L from the conditional distribution of \mathbf{X}_j , denoted $\pi(\mathbf{X}_j | \mathbf{Y}, \mathbf{x}_{(-j)}, \boldsymbol{\theta}; t_1, \dots, t_L)$, using the stochastic forward-backward algorithm [191]. The algorithm enables us to efficiently sample from $\pi(\mathbf{X} | \mathbf{Y}, \mathbf{X}_{(-j)}, \boldsymbol{\theta})$ 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.

In the forward recursion, we construct a sequence of matrices $\mathbf{Q}_j^{(t_2)}, \dots, \mathbf{Q}_j^{(t_L)}$, where $\mathbf{Q}_j^{(t_\ell)} = \begin{pmatrix} q_{j,r,s}^{(t_\ell)} \end{pmatrix}$, and $q_{j,r,s}^{(t_\ell)} = \Pr(\mathbf{X}_j(t_\ell) = s, \mathbf{X}_j(t_{\ell-1}) = r | \mathbf{Y}_{t_1}^{t_\ell}, \mathbf{X}_{(-j)}, \boldsymbol{\theta})$. Let $\mathbf{P}_{r,s}^{(j)}(t_{\ell-1}, t_\ell) = \Pr(\mathbf{X}_j(t_\ell) = s | \mathbf{X}(t_{\ell-1}) = r, \boldsymbol{\theta}; \mathbf{X}_{(-j)})$. If there are changes in the numbers of infected individuals in interval \mathcal{I}_ℓ , we construct the transition probability matrix for that interval as in (3.2.3). Then,

$$q_{j,r,s}^{(t_\ell)} \propto \pi_j^{(t_\ell)}(r | \mathbf{X}_{(-j)}, \boldsymbol{\theta}) \times \mathbf{P}_{r,s}^{(j)}(t_{\ell-1}, t_\ell) \times f(Y_{t_\ell} | \mathbf{X}_j(t_\ell), \mathbf{X}_{(-j)}(t_\ell), \rho, \mathbf{p}_{t_1}), \quad (3.9)$$

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

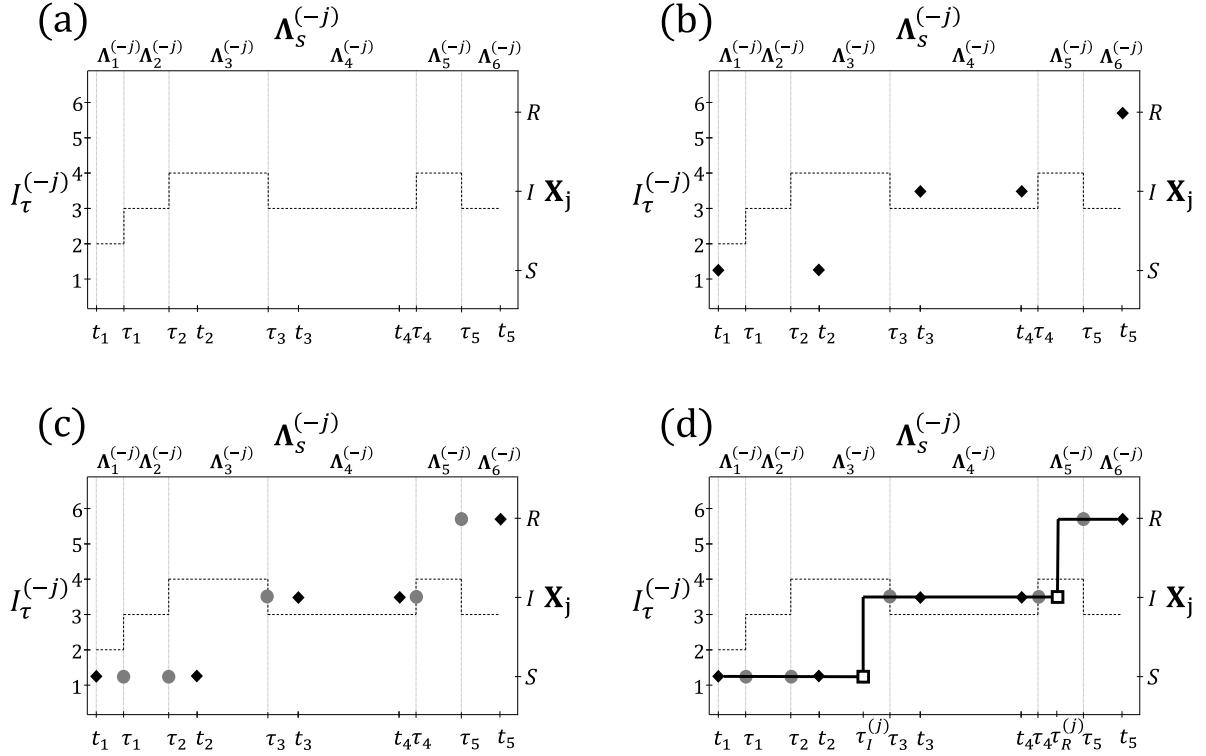


Figure 3.2: Procedure for constructing a subject-path proposal with SIR dynamics. (a) The dashed line depicts the number of infected individuals, excluding \mathbf{X}_j , the subject whose path is being sampled. The observation times, t_1, \dots, t_5 , and times at which other subjects change disease states, τ_1, \dots, τ_5 , are shown on the bottom axis. Rate matrices of the time-inhomogeneous CTMC (top axis) are constant within inter-event intervals (vertical lines). The state space of the subject-level process, \mathbf{X}_j , is shown on the right axis. (b) *HMM step:* Sample the state of \mathbf{X}_j at t_1, \dots, t_5 , conditional on the data and on the disease histories of other subjects. (c) *Discrete time skeleton step:* Sample the infection status at τ_1, \dots, τ_5 , conditional on the sequence of states sampled in the HMM step. (d) *Event time step:* Sample the infection and recovery times from endpoint-conditioned time-homogeneous CTMC distributions, conditional on the sequence of disease states sampled in the HMM and discrete time skeleton steps.

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

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

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

Discrete-time skeleton step

It would be straightforward to sample the exact infection and recovery times of subject j , conditional on the sequence of states at times t_1, \dots, t_L , if the subject-level CTMC rates did not possibly vary over each inter-observation interval. We may reduce our problem to the time-homogeneous case by first sampling the disease state at the intermediate event times when the CTMC rates change, and then sampling the full path within each inter-event interval. Consider an inter-observation interval, $\mathcal{I}_\ell = [t_{\ell-1}, t_\ell]$, containing inter-event intervals whose endpoints are given by times $t_{\ell-1} \equiv \tau_{\ell,0}^{(-j)} < \tau_{\ell,1}^{(-j)} < \dots < \tau_{\ell,n-1}^{(-j)} < \tau_{\ell,n}^{(-j)} \equiv t_\ell$. We recursively sample \mathbf{X}_j at each intermediate event time, beginning at $\tilde{\tau}_1$, from the discrete distribution with masses

$$\begin{aligned}\Pr(\mathbf{X}_j(\tilde{\tau}_i) = x_i | \mathbf{X}_j(\tilde{\tau}_{i-1}) = x_{i-1}, \mathbf{X}_j(\tilde{\tau}_n) = x_n) &= \frac{\Pr(\mathbf{X}_j(\tilde{\tau}_i) = x_i, \mathbf{X}_j(\tilde{\tau}_{i-1}) = x_{i-1}, \mathbf{X}_j(\tilde{\tau}_n) = x_n)}{\Pr(\mathbf{X}_j(\tilde{\tau}_{i-1}) = x_{i-1}, \mathbf{X}_j(\tilde{\tau}_n) = x_n)} \\ &= \frac{\Pr(\mathbf{X}_j(\tilde{\tau}_i) = x_i | \mathbf{X}_j(\tilde{\tau}_{i-1}) = x_{i-1}) \Pr(\mathbf{X}_j(\tilde{\tau}_n) = x_n | \mathbf{X}_j(\tilde{\tau}_i) = x_i)}{\Pr(\mathbf{X}_j(\tilde{\tau}_n) = x_n | \mathbf{X}_j(\tilde{\tau}_{i-1}) = x_{i-1})} \\ &= \frac{[\mathbf{P}^{(j)}(\tilde{\tau}_{i-1}, \tilde{\tau}_i)]_{x_{i-1}, x_i} [\prod_{k=i}^{n-1} \mathbf{P}^{(j)}(\tilde{\tau}_k, \tilde{\tau}_{k+1})]_{x_i, x_n}}{[\prod_{k=i-1}^{n-1} \mathbf{P}^{(j)}(\tilde{\tau}_k, \tilde{\tau}_{k+1})]_{x_{i-1}, x_n}}.\end{aligned}\tag{3.10}$$

Event time step

The final step in constructing a subject-path is to sample the exact infection and recovery times given the discrete sequence of states obtained in the previous two steps. This amounts to simulating the path of an endpoint-conditioned time-homogeneous CTMC, a task for which there exist a variety of efficient methods [109]. When fitting the SIR model, we chose to use modified rejection sampling, a modification of Gillespie's direct algorithm [87] that explicitly avoids simulating constant paths. This method is known to be efficient when the states differ at the endpoints of small time intervals. We used uniformization-based sampling [109] when fitting SEIR and SIRS models, which was more robust when sampling paths in intervals with multiple transitions. Fast implementations of these methods are available in the `ECctmc` package in R [72]. We briefly summarize the algorithms below, and refer to [109] for a more thorough discussion.

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 $\boldsymbol{\Lambda}$ be the rate matrix for the process. Let Λ_a denote the a, a diagonal element of $\boldsymbol{\Lambda}$, 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 .

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, using, for instance, Gillespie's direct algorithm. The proposed path is accepted if $\mathbf{X}(T) = b$. If a transition is not known to have 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

$$F^{-1}(u) = \frac{-\log[1 - u \times (1 - e^{-T\Lambda_a})]}{\Lambda_a}. \quad (3.11)$$

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 ($S \rightarrow R$, or $I \rightarrow S$) were almost never considered. Usually, only a single transition time was sampled in a given interval, so the inverse-CDF method was fast.

The uniformization algorithm samples the path for a time-homogeneous CTMC conditional on the state at the interval endpoints by coupling the original CTMC to the Markov chain for an auxilliary Poisson point process. State transitions, including virtual transition where the state does not change, are events of the point 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{Z} , as $R = I + \frac{1}{\mu} \boldsymbol{\Lambda}$, where $\mu = \max_a \boldsymbol{\Lambda}_a$. The probability mass function for the number of state transitions, N , conditional on $\mathbf{X}(0) = a$, $\mathbf{X}(T) = b$, is

$$P(N = n | \mathbf{X}(0) = a, \mathbf{X}(T) = b) = e^{-\mu T} \frac{(\mu T)^n}{n!} R_{ab}^n / P_{ab}(T). \quad (3.12)$$

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 draw 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, τ_i , $i = 1, \dots, n - 1$, are drawn from the discrete distribution with masses given by

$$P(\mathbf{X}(\tau_i) | \mathbf{X}(\tau_{i-1}), \mathbf{X}(T) = b) = \frac{R_{x_{i-1}, x_i} (R^{n-i})_{x_i, b}}{(R^{n-i+1})_{x_{i-1}, b}}. \quad (3.13)$$

Uniformization-based sampling was preferred in the case of the SEIR model since modified rejection sampling tended to get hung up when sampling paths in intervals where the endpoints suggested that at least two state transitions occurred (which though it seldom occurred, significantly slowed down the MCMC). We also note that the transition probability, $P_{ab}(T)$, is computed and cached in executing the HMM step of our algorithm. Therefore, there are no additional eigen-decompositions or matrix exponentiations required for the uniformization algorithm.

Metropolis–Hastings step

Having constructed a complete subject–path proposal, we decide whether to accept or reject it via a Metropolis–Hastings step. We emphasize that the true distribution of $\mathbf{X}_j|\mathbf{x}_{(-j)}, \boldsymbol{\theta}$ does not match the time–inhomogeneous CTMC in our proposal. Suppressing the dependence on $\boldsymbol{\theta}$, the target distribution of the subject–path proposal is $\pi(\mathbf{X}|\mathbf{Y}) \propto \pi(\mathbf{Y}|\mathbf{X})\pi(\mathbf{X})$. Note that \mathbf{x}^{new} and \mathbf{x}^{cur} differ only in the path of the j^{th} subject, so $\Lambda^{(-j)}(\mathbf{x}^{\text{cur}}) = \Lambda^{(-j)}(\mathbf{x}^{\text{new}}) = \boldsymbol{\Lambda}^{(-j)}$.

Suppressing the dependence on $\boldsymbol{\theta}$ for clarity, the acceptance ratio is

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

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 (3.4). Let $\pi(\mathbf{x}_j^{\text{new}}|\boldsymbol{\Lambda}^{(-j)}; \mathcal{I})$ and $\pi(\mathbf{x}_j^{\text{cur}}|\boldsymbol{\Lambda}^{(-j)}; \mathcal{I})$ denote the time–inhomogeneous subject–level CTMC proposal densities given by (3.7). Then,

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

and similarly,

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

Therefore,

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

Hence, the Metropolis–Hastings acceptance probability is

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

which depends on the ratio of the population-level time-homogeneous CTMC densities, multiplied by the ratio of time-inhomogeneous CTMC proposal densities.

Initializing the collection of subject-paths

We initialize the collection of subject paths at the start of our MCMC by simulating paths using Gillespie’s direct algorithm [87] until we have found a path under which the data have non-zero probability. A sufficient condition for this under the binomial sampling model is that the number of infected individuals is greater than the observed prevalence at each observation time.

3.2.4 Parameter Updates and MCMC Scan Order

One MCMC iteration includes a number of subject-path updates, followed by a set of parameter updates. Conjugate priors were available for all parameters of the models in this chapter. Therefore, we chose to use Gibbs sampling to update parameter values from their univariate full conditional distributions (below).

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 with other DA methods [185], and incur a relatively high computational cost. Frequently updating model parameters may help to break this

correlation, despite the strong autocorrelations in posterior samples of model parameters. Often, the effective sample size (ESS) per CPU time is optimized by sampling only a handful of subject–paths per MCMC iteration. However, 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 all possible factors affecting the optimal number of subject–path updates per iteration. Therefore, we set the number of subject–paths per iteration on the basis of log–posterior ESS per CPU time in an initial run of 5,000–10,000 iterations.

3.2.5 Data augmentation for SEIR dynamics

The SEIR model adds a 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 CTMC 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, $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$, 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

$$\mathbf{x}_j(\tau) = \left(S, \tau < \tau_E^{(j)}; E, \tau_E^{(j)} \leq \tau < \tau_I^{(j)}; I, \tau_I^{(j)} \leq \tau < \tau_R^{(j)}; R, \tau_R^{(j)} \leq \tau \right)$$

where $\tau_E^{(j)}$, $\tau_I^{(j)}$, and $\tau_R^{(j)}$ 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 β be the per–contact infectivity rate, γ be the rate at which an exposed individual becomes infectious, and μ 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 $\boldsymbol{\Lambda} = \beta I$ if $\mathbf{X}_j = S$ and $\mathbf{X}'_j = E$, γ if $\mathbf{X}_j = E$ and $\mathbf{X}'_j = I$, and μ 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

$$\boldsymbol{\Lambda}_m^{(-j)}(\boldsymbol{\theta}) = \begin{pmatrix} & S & E & I & R \\ S & -\beta I_{\tau_m}^{(-j)} & \beta I_{\tau_m}^{(-j)} & 0 & 0 \\ E & 0 & -\gamma & \gamma & 0 \\ I & 0 & 0 & -\mu & \mu \\ R & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (3.14)$$

As with the SIR model, the eigenvalues 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.

3.2.6 Data augmentation for SIRS dynamics

The 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 increases the computational burden since the disease state at the interval endpoints does not absolve us of sampling the full 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, $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$, 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

$$\mathbf{x}_j(\tau) = \left(S, \tau < \tau_{I_1}^{(j)}; I, \tau_{I_1}^{(j)} \leq \tau < \tau_{R_1}^{(j)}; R, \tau_{R_1}^{(j)} \leq \tau < \tau_{L_1}^{(j)}; S, \tau_{L_1}^{(j)} \leq \tau < \tau_{I_2}^{(j)}; \dots \right),$$

where $\tau_{I_k}^{(j)}$, $\tau_{R_k}^{(j)}$, and $\tau_{L_k}^{(j)}$ are times at which subject J becomes infected, recovers, and loses immunity, and are enumerated by the subscript k as the process may revisit each state multiple time. As with the SIR and SEIR models, it is possible that some or all of these events may not come about within the observation period $[t_1, t_L]$. We let β be the per-contact infectivity rate, μ be the rate at which an infectious individual recovers, and γ be the rate at which immunity is lost. 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 $\Lambda = \beta I$ if $\mathbf{X}_j = S$ and $\mathbf{X}'_j = E$, μ if $\mathbf{X}_j = I$ and $\mathbf{X}'_j = R$, and γ 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

$$\Lambda_m^{(-j)}(\boldsymbol{\theta}) = \begin{pmatrix} S & I & R \\ -\beta I_{\tau_m}^{(-j)} & \beta I_{\tau_m}^{(-j)} & 0 \\ 0 & -\mu & \mu \\ \gamma & 0 & -\gamma \end{pmatrix}. \quad (3.15)$$

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 eigenvalues of the rate matrix are complex, we must rotate that rate matrix to obtain its real canonical form. This is further discussed in Section A.1.

3.3 Simulation results

3.3.1 Inference Under a Variety of Epidemic Dynamics

We fit SIR, SEIR, and SIRS dynamics to binomially distributed prevalence counts sampled from epidemics simulated under corresponding dynamics in populations of 750, 500, and 200

individuals (details provided in Section A.3). Priors for the rate parameters and binomial sampling probability were chosen so that the priors spanned reasonable ranges of values (e.g. recovery durations ranging from days to weeks/months rather than seconds to eons under extremely diffuse priors), but were otherwise only mildly informative, while the initial distribution parameters were assigned informative priors (see tables A.4, A.6, and A.8). The three datasets, depicted in Figure 3.3 along with the estimated pointwise posterior prevalence, presented a range of challenges. The SIR example was arguably the most “standard” example as the observation period captured the exponential growth and decline of the epidemic. Thus, much of the curvature in the latent path was reflected in the data. In contrast, data from the outbreak simulated under near-endemic SEIR dynamics contained very little information about the shape of the epidemic curve. The task of disentangling whether the data were sampled with low probability from a high-prevalence outbreak, or visa-versa, was further complicated by the inclusion of an additional disease state — the exposed state — that was not directly observed. Finally, the SIRS model was more computationally challenging for two reasons. First, the recurrent nature of the disease process demanded that the disease state at each event time, and the path within each inter-event interval, be sampled in the subject-path proposal. Second, it was possible for CTMC rate matrices to have complex eigen-decompositions, which made computing transition probability matrices more expensive. This affected the optimal number of subject-path updates per MCMC iteration.

The true epidemic paths and parameter values fell well within the 95% Bayesian credible intervals in all three simulations (Figure 3.3 presents the estimated latent posterior prevalence; Figure 3.4 presents posterior estimates of model parameters; Figure A.11 presents estimated latent posterior distributions and true epidemic paths for all model compartments). The acceptance rates for subject-path proposals were roughly 92% for the SIR model, 91% for the SEIR model, and 77% for the SIRS model. Our posterior estimates of the model parameters also closely match estimates obtained using the particle marginal Metropolis–Hastings (PMMH) algorithm of [10], implemented using the `pomp` package in R [124]. We simulated particle paths in the PMMH algorithm in two ways; exactly using Gillespie’s direct algo-

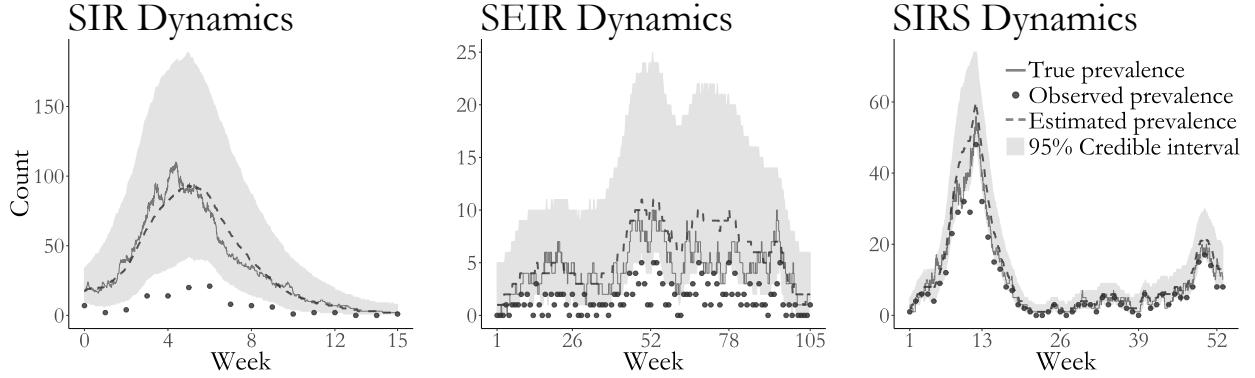


Figure 3.3: Estimated latent posterior distributions of disease prevalence in outbreaks simulated under SIR (left), SEIR (middle), and SIRS (right) dynamics. Depicted are the true unobserved prevalence (solid line), observed data (dots), pointwise posterior median prevalence (dashed line), and pointwise 95% credible intervals (shaded region). Latent posterior estimates are based on a thinned sample, with every 250th sample retained.

rithm [87], and approximately using a multinomial modification of τ -leaping [31]. In these small population examples, the exact algorithm is arguably more appropriate, as the leap conditions for τ -leaping may not be met in small populations, but it is also substantially slower. In these simple settings, PMMH tended to outperform our algorithm in terms of log-posterior effective sample size (ESS) per CPU time. When PMMH particle paths were simulated by τ -leaping, the average ESS per CPU compared to BDA was roughly $350\times$ greater for the SIR model, $4.4\times$ greater for the SEIR model, and $13\times$ greater for the SIRS model. Exact simulation of PMMH particle paths reduced the computational advantage of PMMH substantially. In this case, the average log-posterior ESS per CPU time was $10.5\times$ greater for PMMH in fitting the SIR model, $2\times$ for the SEIR model, and $0.7\times$ for the SIRS model. These comparisons did not include the time required to tune the MCMC for PMMH, which was nontrivial. In contrast, our algorithm required no tuning beyond selecting the number of subject-paths to update per MCMC iteration. We also note that in fitting the models using PMMH, we were required to make several implementation decisions to prevent particle degeneracy and to balance speed with precision. These included selecting the num-

ber of particles and the time-step in the approximate τ -leaping algorithm. For example, when using τ -leaping to simulate particle paths, the number of particles required to obtain good mixing for the SIRS model fit with PMMH was higher than for the other two models. Details of the PMMH implementations and further results are presented in Section A.3.

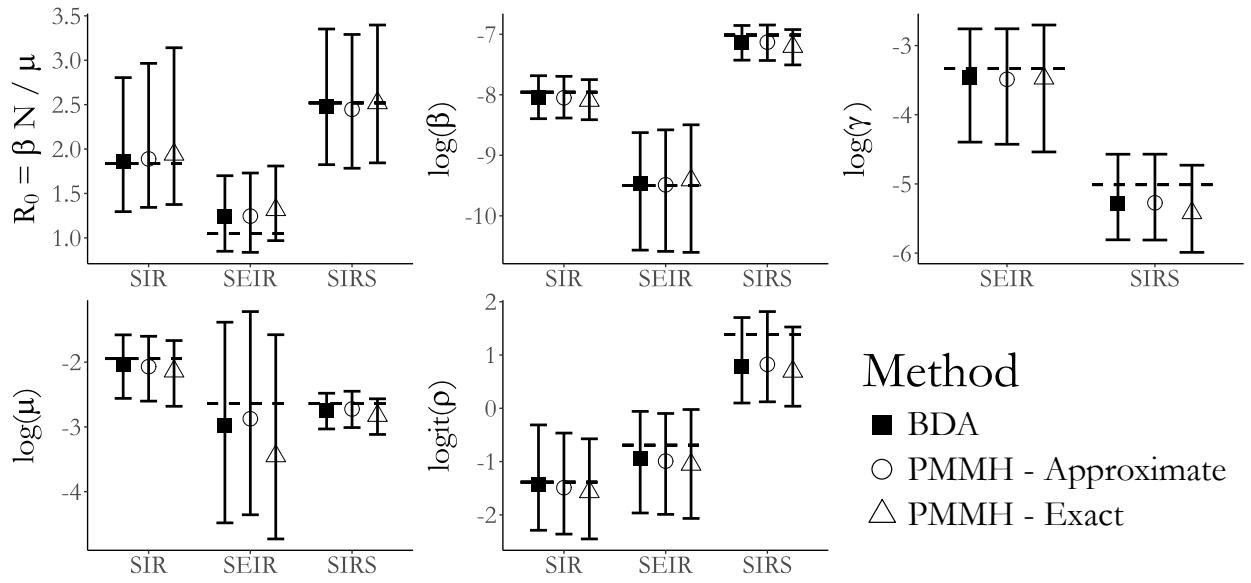


Figure 3.4: Posterior medians and 95% credible intervals of parameters in the SIR, SEIR, and SIRS models fit with Bayesian data augmentation (BDA) and particle marginal Metropolis–Hastings (PMMH) with particle paths simulated approximately (using τ -leaping) and exactly (using Gillespie’s direct algorithm). Displayed are estimates of the basic reproductive number, R_0 , the rate parameters, and the binomial sampling probability. In all models, β is the per-contact infectivity rate, μ is the recovery rate, and ρ is the binomial sampling probability. In the SEIR model, γ denotes the rate at which an exposed individual becomes infectious, while in the SIRS model γ denotes the rate at which immunity is lost.

3.3.2 Inference under model misspecification

In practice, every stochastic epidemic model is misspecified with respect to the real world epidemic process from which the data arise, and the malignancy of the model misspecification is often impossible to fully diagnose. We can build up an understanding of the

epidemic dynamics by fitting SEMS under a range of dynamics, beginning with simple, easily interpretable models. Given the iterative nature of epidemic modeling and the inherent misspecification of simple models that form the building blocks for more realistic models, a minimal criteria for the usefulness of a computational algorithm is that be computationally robust to model misspecification. In other words, we had better, at the very least, be able to use the algorithm to sample from the posterior distribution of simple models.

It is precisely the inherent misspecification of SEMs that leads simulation-based methods to struggle in many instances, and it is here that we highlight a critical advantage of our DA algorithm. Our subject-path proposals are driven, not just by the SEM dynamics, but also by the data. This buys us computational robustness to model misspecification in situations where simulation-based methods degenerate due to their reliance on having a model that is a good approximation of the true data generating process from which to simulate epidemic paths. We demonstrate this in a simple example in which we fit SIR and SEIR models to four years of weekly prevalence data sampled from an epidemic simulated under time-varying SEIR dynamics, where the latent period, infectious period, and per-contact infectivity rate were modulated over four discrete epochs (depicted in Figure 3.5, details presented in Section A.4).

We fit SIR and SEIR models using our DA algorithm, and using PMMH with 2,500 particles, the paths for which were simulated approximately via τ -leaping with a time-step of 1 day. We assigned weakly informative priors for the rate parameters governing the epidemic dynamics in both models, and informative priors for the binomial sampling probability and the initial state probabilities (Table A.11). MCMC chains for models fit via PMMH were plagued by particle degeneracy and did not converge (Figures A.12 and A.14).

Both models fit via DA yield reasonable estimates for the within-subject disease dynamics (i.e., the infectious period, as well as the latent period in the case of the SEIR model). The posterior median average infectious period duration was estimated to be 292 days (95% BCI: 263 days, 323 days) under SIR dynamics, and 287 days (95% BCI: 260 days, 318 days) under SEIR dynamics. The posterior median average latent period under SEIR dynamics was

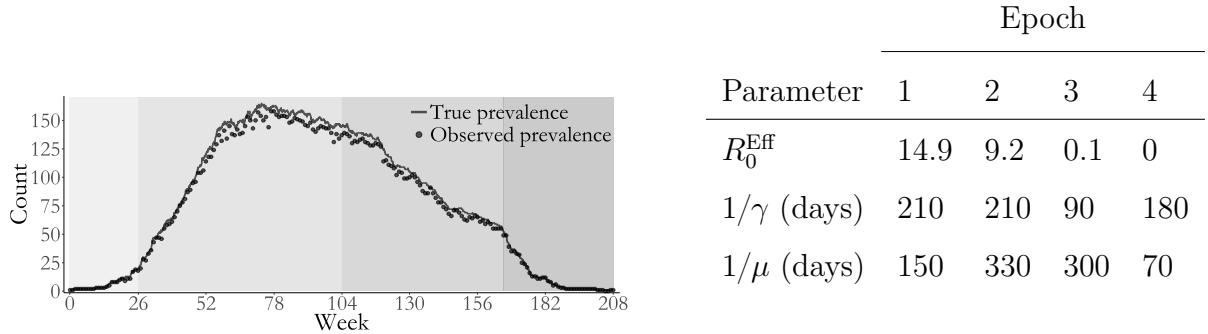


Figure 3.5: Simulated outbreak with SEIR dynamics that varied over four epochs (shaded regions). Weekly prevalence counts (points) were binomially sampled with sampling probability $\rho = 0.95$ from the true unobserved prevalence (solid line). The table presents the effective reproductive number computed based on the number of susceptibles at the beginning of each epoch, $R_0^{\text{Eff}} = \beta(\tau)S(\tau)/\mu(\tau)$, the mean latent period, $1/\gamma$, and the mean infectious period, $1/\mu$.

211 days (95% BCI: 165 days, 260 days). The posterior median estimate of R_0 under SIR dynamics was 4.05 (95% BCI: 3.40, 4.81), while under SEIR dynamics, the posterior median estimate of R_0 was 23.8 (95% BCI: 15.1, 37.0). While the true prevalence fell well within the pointwise 95% credible interval for both models (Figure 3.6), we notice that the degree of model misspecification drastically affected our ability to estimate the history of the numbers of noninfectious people over the course of the epidemic. Under SIR dynamics, we drastically overestimate the number of susceptible individuals. The SEIR model much more closely resembles the time-varying SEIR model used to simulate the epidemic. Although the true path for the number of susceptible still falls outsize the 95% credible interval at times, we are still able to reconstruct a reasonable range of paths for the number of exposed individuals. This contrasts with the models fit in Section 3.3.1, which were not misspecified with respect to the true epidemic dynamics. In that case, the complete path of the epidemic fell well within the estimated credible intervals for all disease states for all three models (Figure A.11). Therefore, we advise caution in reconstructing the epidemic history for disease states that were not measured, particularly when severe model misspecification is suspected.

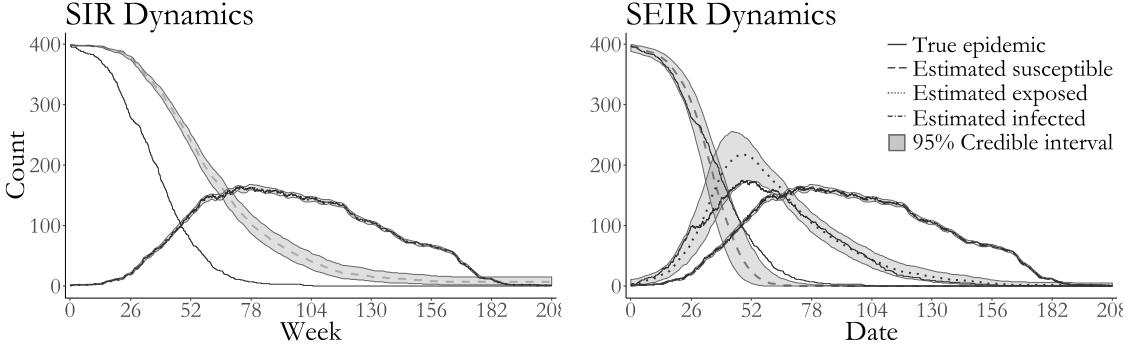


Figure 3.6: True epidemic path (solid lines), pointwise posterior median estimate of the numbers of susceptibles (dashed line), exposed (dotted line), and infected individuals (dash-dotted line) and pointwise 95% credible intervals (shaded regions) under SIR and SEIR dynamics.

3.3.3 Effect of prior specification on posterior inference

Given the limited extent of aggregated prevalence counts, we must consider how our choices of prior distributions influence our posterior inferences. We simulated an outbreak with SIR dynamics in a population of 750 individuals for which $R_0 = \beta \times 763/\mu \approx 1.84$ and the mean infectious period was $1/\mu = 7$ days. We fit SIR models to binomially distributed weekly prevalence data, sampled with detection probability $\rho = 0.2$, under the following four prior regimes: Regime 1 — informative priors for all model parameters; Regime 2 — vague priors for the rate parameters and an informative prior for the sampling probability; Regime 3 — informative priors for the rate parameters and a flat prior for the sampling probability; Regime 4 — vague priors for the rate parameters and a flat prior for the sampling probability. The same prior for the initial state probabilities was used in all four regimes. Complete simulation details and convergence diagnostics are supplied in Section A.6.

The true values for all model parameters fell within the 95% credible intervals under all four prior regimes. Unsurprisingly, informative priors tended to result in narrower credible intervals for the parameters (Figure 3.7) as well as for the latent process (Figure 3.8). The strength of prior information about the sampling probability affected the widths of credible

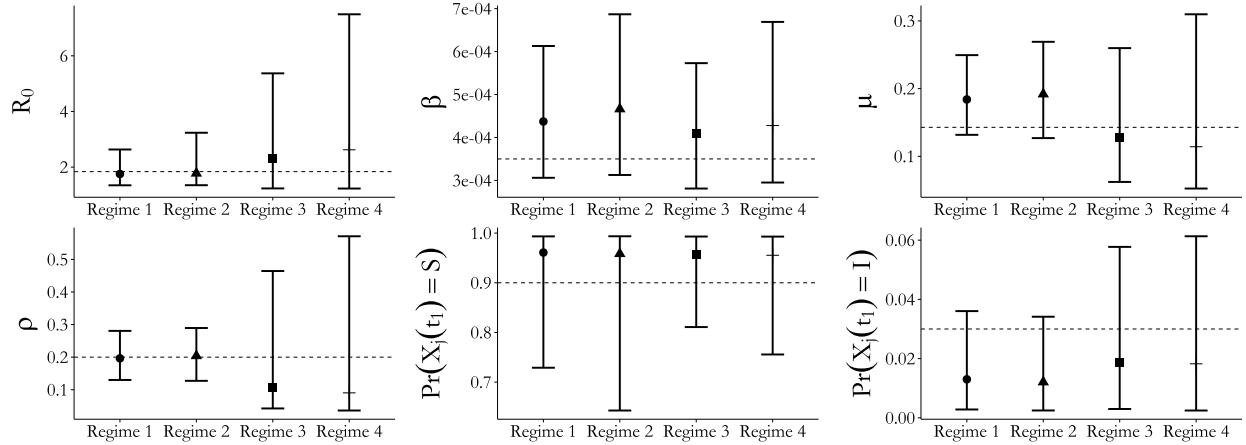


Figure 3.7: Posterior median estimates and 95% credible intervals for all SIR model parameters under four different prior regimes (Table A.14). Regimes 1 and 3 set informative priors for the per-contact infectivity and recovery rates. Regimes 1 and 2 set informative priors for the binomial sampling probability. The same mildly informative prior for the initial state probabilities was used in all four regimes.

intervals to a much greater extent than the priors for the rate parameters. Strong prior information about the sampling probability also resulted in substantially narrower credible intervals for disease prevalence under each of the prior regimes for the rate parameters. In contrast, informative priors for the rate parameters yielded only slightly narrower credible intervals for disease prevalence when holding constant the strength of the sampling probability prior. The effects on the initial state probability parameters seem to reverse this pattern, although we caution against overinterpretation given the paucity of data available for estimating those parameters. MCMC chains with strong priors for the binomial sampling probability also appeared to mix somewhat better than chains with diffuse priors for the sampling probability (see traceplots in Section A.6).

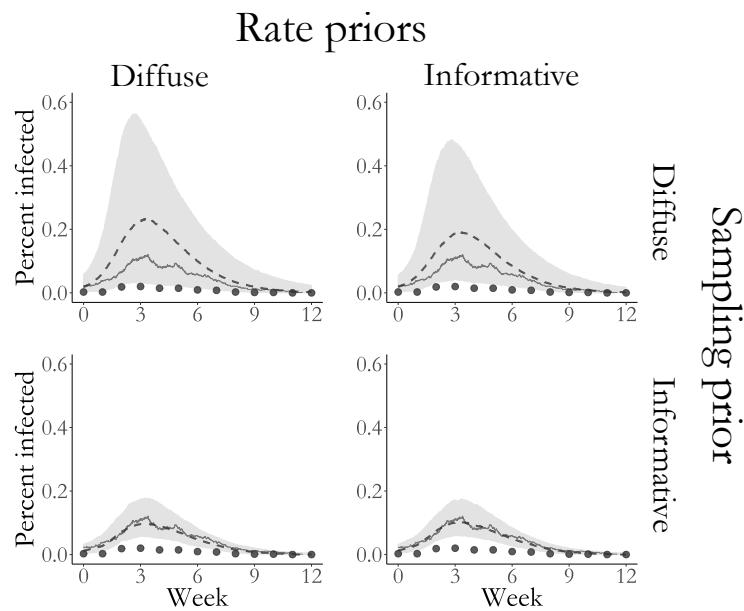


Figure 3.8: Estimated latent posterior distributions of disease prevalence in outbreaks simulated under four prior regimes for SIR model rate parameters and the binomial sampling probability. Depicted are the true unobserved prevalence (solid line), observed data (dots), pointwise posterior median prevalence (dashed line), and pointwise 95% credible intervals (shaded region). Latent posterior estimates are based on a thinned sample, with every 250th sample retained.

3.4 Example: Influenza in a British boarding school

As an example, we apply the methods developed in this chapter data from the boarding school outbreak of influenza that was described in Section 1.3.1. We used our DA algorithm and PMMH to fit SIR and SEIR models with a binomial emission distribution to the data (see Section A.7 of the supplement for complete details). All of the parameters were assigned diffuse priors, which are plotted over the posterior ranges in Figure 3.10. The PMMH algorithm failed to converge for both models, which we suspect was due to a combination of model misspecification and the constrained state space of the binomial measurement process. We also fit a set of supplementary SIR and SEIR models in Section A.7.2, in which we assumed a negative-binomial emission distribution. This was done in order to facilitate comparison with PMMH, although we feel that a negative binomial emission distribution is not appropriate in such a closely monitored outbreak setting since it does not rule out over-reporting of cases.

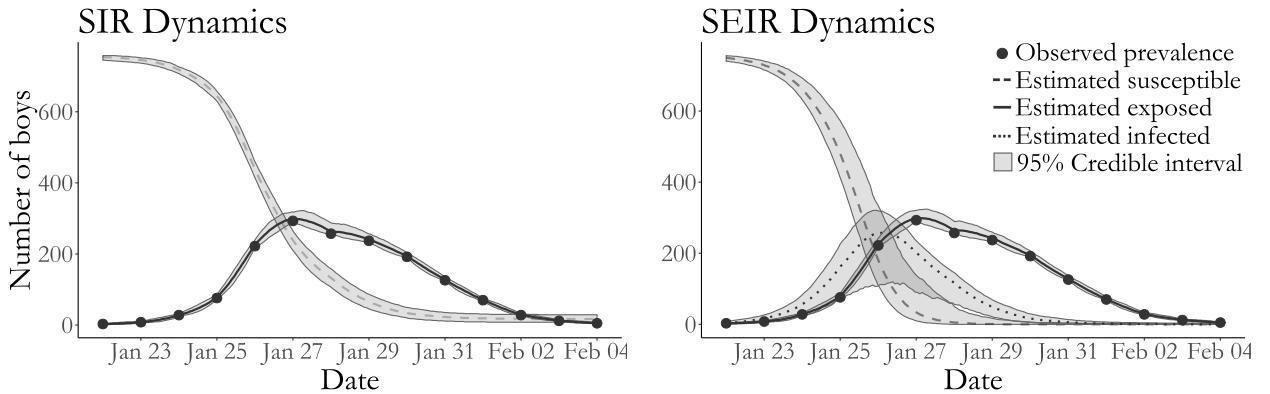


Figure 3.9: Boarding school data, pointwise posterior median estimates and pointwise 95% credible intervals (grey shaded areas) under SIR and SEIR dynamics of the numbers of susceptible boys (dashed line), exposed boys (dotted line), and infected boys (solid line). Posterior estimates based on a thinned sample, with every 250th configuration retained.

Together, the SIR and SEIR models suggest that cases were detected with high probability and that the outbreak, though aggressive, was not atypical given the closed environment

in which it occurred. The posterior median estimates of the detection probability, roughly 0.98 for both models (SIR 95% BCI: 0.92, 1.00; SEIR 95% BCI: 0.91, 1.00), suggested that while almost all of the infectious boys were detected, a handful of cases went unnoticed. The posterior median recovery rate under SIR dynamics corresponds to an average period of 2.16 days (95% BCI: 1.99, 2.37) during which an infectious boy could transmit an infection to other boys before being confined to the infirmary. Under SEIR dynamics, the posterior median average infectious period was 2.12 days (95% BCI: 1.95, 2.33), and the posterior median average latent period was 1.19 days (95% BCI: 0.84, 1.51). These results are consistent with the typical progression of influenza, in which individuals typically incubate for between one to four days before symptoms manifest, and are typically infectious for one day before, and up to a week after, symptom onset [47]. The posterior median estimates of R_0 were 3.89 (95% BCI: 3.40, 4.47) under SIR dynamics, and 10.38 (95% BCI: 7.40, 14.11) under SEIR dynamics. Previous analyses of this dataset with trajectory matching estimate R_0 to be roughly 3.7 for the SIR model and 35.9 for the SEIR model [218, 119], though we note that these estimates are based on deterministic models that do not properly account for distributional properties of the data. Another analysis using the linear noise approximation produced point estimates of R_0 in the range of 3.4–3.6, depending on the model specification [187]. Our results for both models are also in agreement with estimates of SIR and SEIR model dynamics under a negative binomial emission distribution (see Section A.7.2).

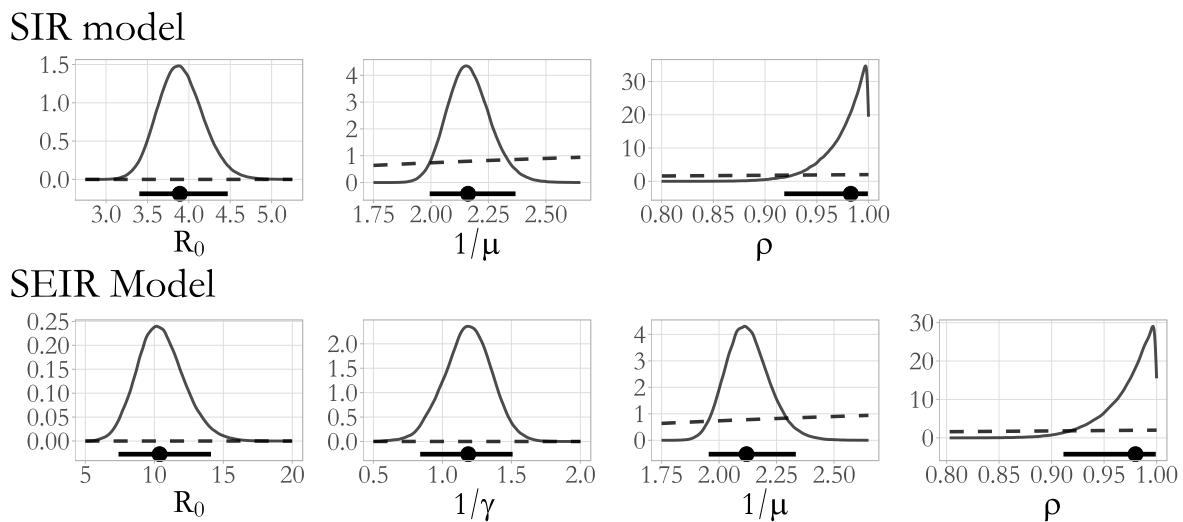


Figure 3.10: Posterior density estimates for $R_0 = \beta N / \mu$, the mean latent and infectious periods, $1/\gamma$ and $1/\mu$, and the binomial sampling probability, ρ , from SIR and SEIR model parameters fit to the British boarding school data (solid lines). The posterior median and 95% Bayesian credible intervals are drawn below the density plots (solid lines with circles). The implied prior densities (dashed lines) for R_0 and the latent and infectious periods, and the prior density for the binomial sampling probability, are plotted over the posterior ranges.

3.5 Discussion

We have presented an agent-based Bayesian DA algorithm for fitting SEMs to disease prevalence time series counts. This was previously difficult, if not computationally infeasible, to carry out using traditional agent-based DA methods in the absence of subject-level data. Although we outlined the algorithm in the context of fitting an SIR model to binomially distributed prevalence data, our algorithm represents a general solution for fitting SEMs to prevalence counts. In simulations and the applied example, we fit SEIR and SIRS models to prevalence data, and in the supplement also fit SIR and SEIR models with a negative binomial emission distribution to the British boarding school data. We have demonstrated that our algorithm yields approximately valid inference when the population size is misspecified. Moreover, our algorithm is usable in settings where simulation-based methods, such as PMMH, break down due to misspecification of the SEM. Finally, our DA algorithm is carried out entirely at the subject level, making it possible to also incorporate subject-level covariates and household structure, or to fit models to subject-level data.

There are two fundamental limitations of agent-based DA methods from which our algorithm is not excepted. First, the bookkeeping required to track the collection of subject-paths increases in size and complexity as the number of events grows large. Attempts to fit stochastic epidemic models in large populations using agent-based DA may be thwarted by prohibitive computational overhead. MCMC run times using our implementation, which was coded for reliability rather than speed, substantially degraded once the assumed population size was greater than a few thousand people. Second, we suspect that MCMC mixing in large populations could eventually become too slow for agent-based DA to be of practical use, even if solutions could be found for the computational bottlenecks. As the population size gets large, perturbations to the likelihood from re-sampling one subject at a time become relatively less significant. For this reason, we view extensions for jointly sampling multiple subject-paths as a critical step in mitigating slow MCMC mixing in large populations.

To conclude, we would like to comment on directions for future work that could be pur-

sued. The DA algorithm in this paper addresses the problem of fitting SEMs to prevalence data. This type of data summarizes total number of infections in the population at a particular time. However, outbreak data often consist of incidence counts, which are the number of new cases accumulated in each inter-observation interval. Extending our DA algorithm to accommodate incidence data is an important next step and should be straightforward in situations where the state space for the subject level process is finite — for instance, if a subject cannot become reinfected more than once or twice in a given inter-observation interval. Another line of inquiry involves improving the computational efficiency of the algorithm. One possibility would be to use a coarse grid for the time-varying force of infection used in each subject-path proposal, and to approximate the transition probabilities in the HMM step of the algorithm. For example, we could propose each subject-path conditionally on the disease states of all other subjects but assume that the force of infection in each inter-observation interval is fixed. This might lead to more subject-paths being rejected, but the loss could well be offset by not having to as many matrix exponentials. Finally, we note that the algorithm could also be used to fit semi-Markov models where transition probabilities depend on the duration of state occupancy. This would require a modification of the complete data likelihood and the Metropolis–Hastings ratio, though we could retain the Markovian structure of the proposal and possibly avoid costly rejected proposals by using a phase-type distribution to approximate the target semi-Markov process.

Chapter 4

APPROXIMATE INFERENCE FOR STOCHASTIC EPIDEMIC MODELS OF OUTBREAKS IN LARGE POPULATIONS

4.1 Overview

Surveillance and outbreak response systems often report incidence counts of new cases detected in each inter-observation time interval. Analyzing this type of time series data is challenging since we must overcome many of the same challenges that we face in modeling the transmission dynamics of infectious diseases in small population settings with prevalence data — discrete snapshots of a continuously evolving epidemic process, detecting a fraction of the new cases, and often directly observing only one aspect of the disease process. Furthermore, our task is made more difficult by the additional computational burden that results from repeated evaluation of continuous time Markov chain (CTMC) likelihoods; the products of exponential waiting time distributions consist of polynomially increasing numbers of terms, and agent-based data augmentation (DA) Markov chain Monte Carlo (MCMC) algorithms become unwieldy as the numbers of subject-path proposals required to meaningfully perturb the CTMC likelihood get large [74].

In this chapter, we show how the LNA of Section 2.1.4 can be adapted to obtain approximate inference for SEMs fit to epidemic count data in large populations. Our contributions are threefold: First, we demonstrate how the SEM dynamics should be reparameterized so that the LNA can be used to approximate transition densities of the counting processes for disease state transition events. Second, we fold the LNA into a Bayesian DA framework in which latent LNA paths are sampled using the elliptical slice sampling (EliptSS) algorithm of [159]. This provides us with general machinery for jointly updating the latent paths while absolving us of the *de facto* requirement that the emission probability distribution to be

Gaussian in order to preserve computational efficiency as in [68, 128, 186, 187] to avoid the need for intensive particle filter methods as in [93]. Finally, we introduce a non-centered parameterization (NCP) for the LNA that massively improves the efficiency of our DA MCMC framework and makes it tractable for fitting complex models.

Section 4.2 presents the methods developed in this chapter. Sections 4.2.1, 4.2.5, and 4.2.3 introduce the data and models in the context of fitting a susceptible–infected–recovered (SIR) model to negative binomially distributed incidence counts. Sections 4.2.4, 4.2.5 derive an LNA that is appropriate for fitting SEMs to incidence data by reparameterizing the stochastic differential equation (SDE) representation of a SEM to represent the latent incidence process, and then obtaining the LNA for the reparameterized system of SDEs. Section 4.2.6 further develops the LNA framework to facilitate efficient MCMC. Simulations and model diagnostics are presented in Section 4.3, and an analysis of incidence data from the 2014–2015 Ebola outbreak in West Africa is presented in Section 4.4.

4.2 Fitting Stochastic Epidemic Models via the Linear Noise Approximation

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

pathogen to be consistent across populations.

4.2.1 Measurement Process and Data

Incidence data, $\mathbf{Y} = \{Y_1, \dots, Y_L\}$, arise as increments of the numbers of new cases accumulated in a set of time intervals, $\mathcal{I} = \{\mathcal{I}_1, \dots, \mathcal{I}_L : \mathcal{I}_\ell = (t_{\ell-1}, t_\ell]\}$. In outbreak or surveillance settings, we do not typically believe that every case is detected since individuals may be asymptomatic or may escape detection. Let $\mathbf{N}^c = (N_{SI}^c, N_{IR}^c)$ denote the counting process for the cumulative numbers of infections ($S \rightarrow I$ transitions) and recoveries ($I \rightarrow R$ transitions), and let $\Delta \mathbf{N}^c(t_\ell) = \mathbf{N}^c(t_\ell) - \mathbf{N}^c(t_{\ell-1})$ denote the change in cumulative numbers of transitions over \mathcal{I}_ℓ ; so, $\Delta N_{SI}^c(t_\ell)$ is the incidence over $(t_{\ell-1}, t_\ell]$. We might choose to model the number of observed cases as a negative binomial sample of the true incidence with detection rate ρ and over-dispersion parameter ϕ . Thus,

$$Y_\ell | \Delta N_{SI}^c(t_\ell), \rho \sim \text{Neg.Binom.}(\mu_\ell = \rho \Delta N_{SI}^c(t_\ell), \sigma_\ell^2 = \mu_\ell + \mu_\ell^2/\phi). \quad (4.1)$$

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

4.2.2 Latent Epidemic Process

The SIR model is often expressed in terms of compartment counts, $\mathbf{X}^c = \{S^c, I^c, R^c\}$, that evolve in continuous time on state space

$$\mathcal{S}_X^c = \{(l, m, n) : l, m, n \in \{0, \dots, N\}, l + m + n = P\}.$$

We will make the (not particularly limiting) modeling choice to express the waiting times between disease state transitions as being exponentially distributed. Thus, \mathbf{X} evolves according to a Markov jump process (MJP). If our data had consisted of prevalence counts, which arise as partial observations of infected individuals, we might have chosen to approximate transition densities of the MJP for \mathbf{X} in the usual way that appears in [128, 68].

However, incidence data are discretely observed, partial realizations of the increments of counting processes that evolve continuously in time as individuals transition among disease states. The emission probabilities for incidence data, e.g., (4.1), depend on the change in N_{SI}^c over the time interval $(t_{\ell-1}, t_\ell]$, not on the change in I over the interval. It would be incorrect to treat incidence as simply the difference in prevalence. We could easily conjure up a scenario where there are positive numbers of infections, but where the prevalence, discretely observed, does not appear to change due to an equal number of recoveries. We need to construct the LNA that approximates transition densities of \mathbf{N} if we are to write down correctly specified emission probabilities.

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

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

which is the set of non-decreasing cumulative incidence 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 denote the rate at which each infected individual recovers. The rate at

which \mathbf{N}^c transitions from state \mathbf{n} to \mathbf{n}' is

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

4.2.3 Tractable Approximations for Intractable Likelihoods

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

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

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

4.2.4 Diffusion Approximation

There are a variety of methods for arriving at the diffusion approximation for a MJP [78]). In the interest of clarity, we follow [68, 92, 93, 222] and appeal to an intuitive, though somewhat informal, construction of a stochastic differential equation (SDE) whose drift and diffusion are obtained match the approximate moments of increments of the MJP path in infinitesimal time intervals. This SDE is referred to as the chemical Langevin equation (CLE). We refer to [78, 88, 216] for more detailed presentations of the following discussion.

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

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

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

2. dt is sufficiently *large* that we can expect many disease state transitions of each type:

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

Condition (4.4), which can be trivially satisfied by choosing dt to be infinitesimally small, implies that the numbers of infections and recoveries in $(t, t + dt]$ are essentially independent of one another since the rates at which they occur are approximately constant within the interval [88]. This condition also implies that the numbers of infections and recoveries in the interval are independent Poisson random variables with rates $\boldsymbol{\lambda}(\mathbf{x}^c(t)dt)$, i.e., $N_{SI}^c(dt) \sim \text{Poisson}(\beta S(t)I(t)dt)$ and $N_{IR}^c(dt) \sim \text{Poisson}(\mu I(t)dt)$. Condition (4.5), which we can reasonably expect to be satisfied in large populations where transmission dynamics are near their deterministic ODE limits [216], implies that the Poisson distributed increments can be well approximated by independent Gaussian random variables.

When (4.4) and (4.5) are satisfied, we approximate the integer-valued processes, \mathbf{X}^c and \mathbf{N}^c , with the real-valued processes, \mathbf{X} and \mathbf{N} . For the SIR model, the state space of \mathbf{X} is

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

and the state space of \mathbf{N} is

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

In words, the state space of \mathbf{X} will be the set of compartment volumes that are non-negative and that sum to the population size, while the state space of \mathbf{N} is the set of non-decreasing and non-negative incidence paths, constrained so that they do not lead to invalid prevalence paths (e.g., if at some point there are more recoveries than infections, which would lead to a negative number of infected individuals). For now, we will ignore the constraints on \mathcal{S}_N^R and \mathcal{S}_X^R , and approximate the changes in cumulative incidence of infections and recoveries in an infinitesimal time step as

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

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

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

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

Reparameterizing the CLE in terms of incidence

We borrow from [31, 108] a reparameterization for $\mathbf{X}(t)$ in terms of $\mathbf{N}(t)$, conditional on the initial conditions $\mathbf{X}(t) = \mathbf{X}_0$ and $\mathbf{N}(t) = \mathbf{0}$. Let \mathbf{A} denote the matrix whose rows specify changes in counts of susceptible, infected, and recovered individuals corresponding to one infection or recovery event:

$$\mathbf{A} = \begin{array}{c} S \quad I \quad R \\ \xrightarrow{S \rightarrow I} \left(\begin{array}{ccc} -1 & 1 & 0 \\ 0 & -1 & 1 \end{array} \right) \\ \xrightarrow{I \rightarrow R} \end{array}. \quad (4.8)$$

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

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

For the SIR model,

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

so we rewrite (4.7) as

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

Log transforming the CLE

Changes in compartment volumes affect the rates, and hence incidence increments, multiplicatively. Therefore, we would like for perturbations about the drift in (4.11) to be symmetric on a multiplicative, not an additive scale, which leads us to log transform (4.11). Let $\tilde{\mathbf{N}} = \log(\mathbf{N} + \mathbf{1}) \implies \mathbf{N} = \exp(\tilde{\mathbf{N}}) - \mathbf{1}$. By Itô's lemma [164], the corresponding SDE for $\tilde{\mathbf{N}}$ is

$$\begin{aligned} d\tilde{\mathbf{N}}(t) &= \text{diag}\left(\exp(-\tilde{\mathbf{N}}(t)) - 0.5\exp(-2\tilde{\mathbf{N}}(t))\right)\boldsymbol{\lambda}\left(\exp(\tilde{\mathbf{N}}(t)) - \mathbf{1}\right)dt + \\ &\quad \text{diag}\left(\exp(-\tilde{\mathbf{N}}(t))\right)\boldsymbol{\Lambda}\left(\exp(\tilde{\mathbf{N}}(t)) - \mathbf{1}\right)^{1/2}d\mathbf{W}_t \end{aligned} \quad (4.12)$$

$$= \boldsymbol{\eta}(\tilde{\mathbf{N}}(t))dt + \boldsymbol{\Phi}(\tilde{\mathbf{N}}(t))^{1/2}d\mathbf{W}_t \quad (4.13)$$

4.2.5 Linear Noise Approximation

In Section 2.1.4, we followed [68, 92] in obtaining the LNA for SDEs of the same form as (4.13). Briefly, the derivation proceeded as follows: we first decomposed $\tilde{\mathbf{N}}$ into its deterministic ODE limit and a stochastic residual. The SDE corresponding to (4.12) was

then Taylor expanded around its deterministic limit, discarding higher order terms, to obtain a linear SDE for the residual. This linear SDE had an explicit solution as a Gaussian random variable. As noted in [216], the LNA can reasonably approximate the stochastic aspects of a density dependent MJP when conditions (4.4) and (4.5) are satisfied, at least over short time horizons. Over longer time periods the approximation may deteriorate as departures from the deterministic behavior of the system, which is determined by its initial conditions, accumulate. One solution, proposed in [68, 153] and that we will adopt here, is to restart the LNA approximation at the beginning of each inter-observation interval.

When we restart the LNA (4.13 at the beginning of each time interval, $(t_{\ell-1}, t_\ell]$, we obtain a Gaussian approximation for the transition density of $\tilde{\mathbf{N}}$, which is of the form

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

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

$$\frac{d\boldsymbol{\mu}(t)}{dt} = \boldsymbol{\eta}(\boldsymbol{\mu}(t)), \quad (4.15)$$

$$\frac{d\mathbf{m}(t)}{dt} = \mathbf{F}(t)\mathbf{m}(t), \quad (4.16)$$

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

with respect to initial conditions $\mathbf{N}(t_{\ell-1}) = \mathbf{0}$, $\mathbf{X}(t_{\ell-1}) = \mathbf{x}(t_{\ell-1})$, $\mathbf{m}(t_{\ell-1}) = \mathbf{0}$, and $\boldsymbol{\Sigma}(t_{\ell-1}) = \mathbf{0}$, and where $\mathbf{F}(t)$ is the Jacobian $\left(\frac{\partial \boldsymbol{\eta}_i(\boldsymbol{\mu}(t))}{\partial \boldsymbol{\mu}_j(t)}\right)_{i,j \in 1, \dots, |\tilde{\mathbf{N}}|}$ evaluated along the solution to (4.15). Note that we never need to solve (4.16) since $\mathbf{m}(t_0) = \mathbf{0}$ implies that $\mathbf{m}(t_\ell) = \mathbf{0} \forall l = 1, \dots, L$.

Approximating the transition densities of \mathbf{N} using the LNA, (4.14), yields an approximation of the observed data likelihood in (4.3) in the form of a Gaussian state space model. The augmented approximate posterior is

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

The emission probabilities in (4.18) depend on the incidence, not the log-incidence. Hence, we exponentiate LNA paths when computing emission probabilities. 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} become negative and that can cause routines for numerically integrating the LNA ODEs to fail.

4.2.6 Inference via the Linear Noise Approximation

To this point, we have discussed how to approximate transition densities of a MJP via the LNA. However, this is only half the battle since we must also address the computational aspects of sampling from the augmented approximate posterior, (4.18). A central computation challenge that plagues DA MCMC in weak data regimes is that MCMC chains may suffer from severe autocorrelation when the algorithm alternately updates the latent variables given the parameters, and parameters given the latent variables, see e.g., [18, 170, 171, 228]. As we can see in Figure 4.1, a DA MCMC algorithm that alternates between updates to LNA paths and model parameters is no exception.

This phenomenon of poorly mixing MCMC chains can be traced to the use of a centered parameterization (CP) for the LNA in (4.18). The top plot in Figure 4.3, which depicts the CP representation of an LNA path, provides some insight into why the CP leads to poor MCMC mixing. Under the CP, updates to $\boldsymbol{\theta}|\tilde{\mathbf{N}}, \mathbf{Y}$ are made conditionally on a *fixed* LNA path. Therefore, proposed parameter values are accepted depending on whether they are concordant with the data *and* the current path. Small perturbations to model parameters can result in shifts of the LNA transition densities (grey densities) that would render the current path (red points) unlikely under the proposal.

Non-centered Parameterization

We can improve MCMC mixing by decoupling the latent process from the model parameters in the model hierarchy. We do this by using a non-centered parameterization (NCP) of

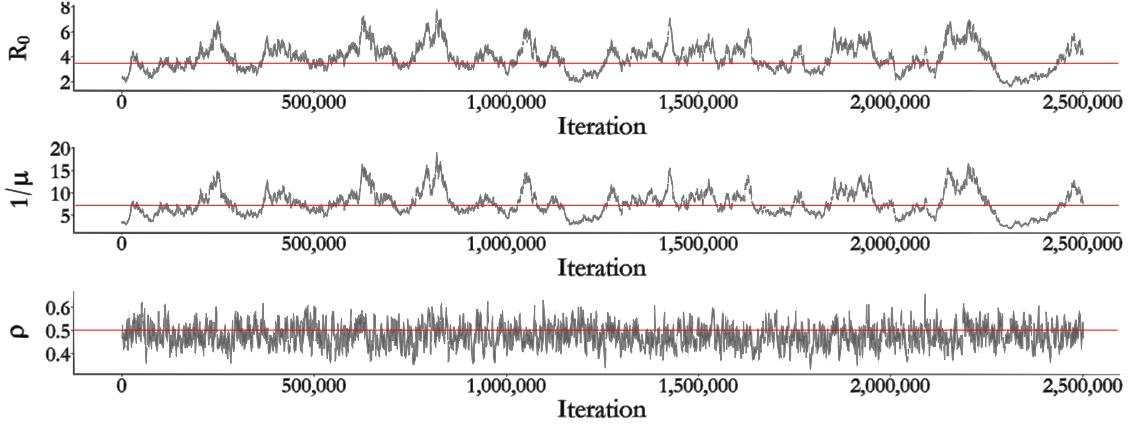


Figure 4.1: Posterior traceplots for parameters of interest for a single MCMC chain of an SIR model fit to negative binomial distributed incidence data. MCMC targeted the centered parameterization of the posterior, (4.18), alternately updating the non-restarting LNA for $\tilde{\mathbf{N}}|\boldsymbol{\theta},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

the log-incidence process, which is a deterministic mapping of standard normal random variables, $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$, onto an LNA path (Figure 4.3; bottom plot). The variables, \mathbf{Z} , are *a priori* independent of the model parameters. Let $\tilde{\mathbf{N}}(t_\ell) \sim MVN(\boldsymbol{\mu}(t_\ell), \Sigma(t_\ell))$ and $\mathbf{Z}(t_\ell) \sim MVN(\mathbf{0}, \mathbf{I})$. The NCP is given by $(\boldsymbol{\theta}, \mathbf{Z})$, and we map \mathbf{Z} onto $\tilde{\mathbf{N}}$ via

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

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

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

We will denote by $\mathbf{N}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I})$ and $\mathbf{X}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I})$ the incidence and prevalence sample paths that are output by the doLNA procedure, which is presented in Algorithm 8.

In each iteration of our DA MCMC algorithm, we alternate between updates to $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ and $\boldsymbol{\theta}|\mathbf{Z}, \mathbf{Y}$. In contrast with the CP, updates to model parameters depend only on the likelihood of the data, not on the joint likelihood of the data and LNA draws. Hence, an

Algorithm 8 Mapping standard normal draws onto LNA sample paths.

```

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


---



```

update to the model parameters need only concord with the current path insofar as the data are likely under the current path. Furthermore, updates to model parameters under the NCP also perturb the LNA path, even as the LNA draws are clamped to their current values (i.e., parameter updates shift the grey densities in the top plot of Figure 4.3 while the red dots identify the LNA path on the new, shifted LNA densities). The NCP also stabilizes updates of $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ since we can make large perturbations to \mathbf{Z} and be assured that the proposed LNA path will be concordant with the data since the LNA draws do not affect the sequence of LNA transition densities that is induced by the model parameters (i.e., updates to LNA draws only change the placement of the red dots in the bottom row of Figure 4.3).

Figure 4.1 shows traceplots of model parameters for one of MCMC chains for an SIR model fit to Poisson distributed incidence data using the CP. Each MCMC chain was run for 2.5 million iterations, following a tuning run of equal length, but each chain 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 that followed a short tuning run of equal length. Figure 4.2 shows the traceplot for one of the NCP MCMC

chains, which mixes quite well.

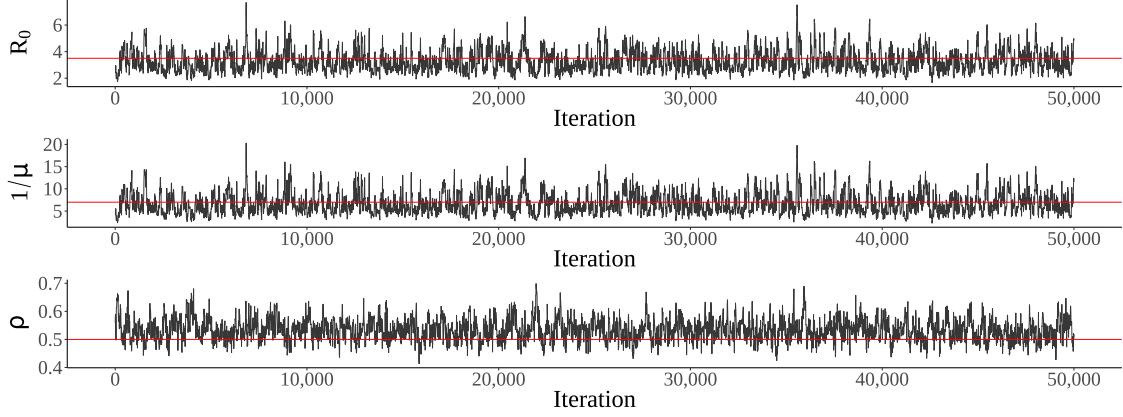


Figure 4.2: Posterior traceplots of parameters of interest sampled by a single MCMC chain of an SIR model fit to Poisson distributed incidence data. MCMC targeted the non-centered parameterization of the posterior, 4.20, alternately updating $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ via elliptical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

The NCP of the LNA also plays an important role in enabling efficient updates of $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ via the elliptical slice sampling (ElliptSS) algorithm of [159], which was detailed in Section 2.2.1 and is presented in Algorithm 9. ElliptSS is an efficient and easy to implement MCMC algorithm for sampling latent Gaussian random variables, \mathbf{Z} , in models where the posterior of can be decomposed as the Gaussian prior for \mathbf{Z} and likelihood, $L(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta})$, i.e.,

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

The target posterior under the LNA NCP, (4.20), is of this form, regardless of whether the LNA is restarted at the beginning of each inter-observation interval, as in [68], or the non-restarting version is used as in [128]. Note that the CP cannot be expressed as a jointly Gaussian collection of random variables with complete data likelihood of the form (4.21) when we use the restarting version of the LNA. Although each transition density, (4.14), is itself Gaussian, the joint LNA path, \mathbf{N} , is not *a priori* Gaussian when the LNA ODEs are

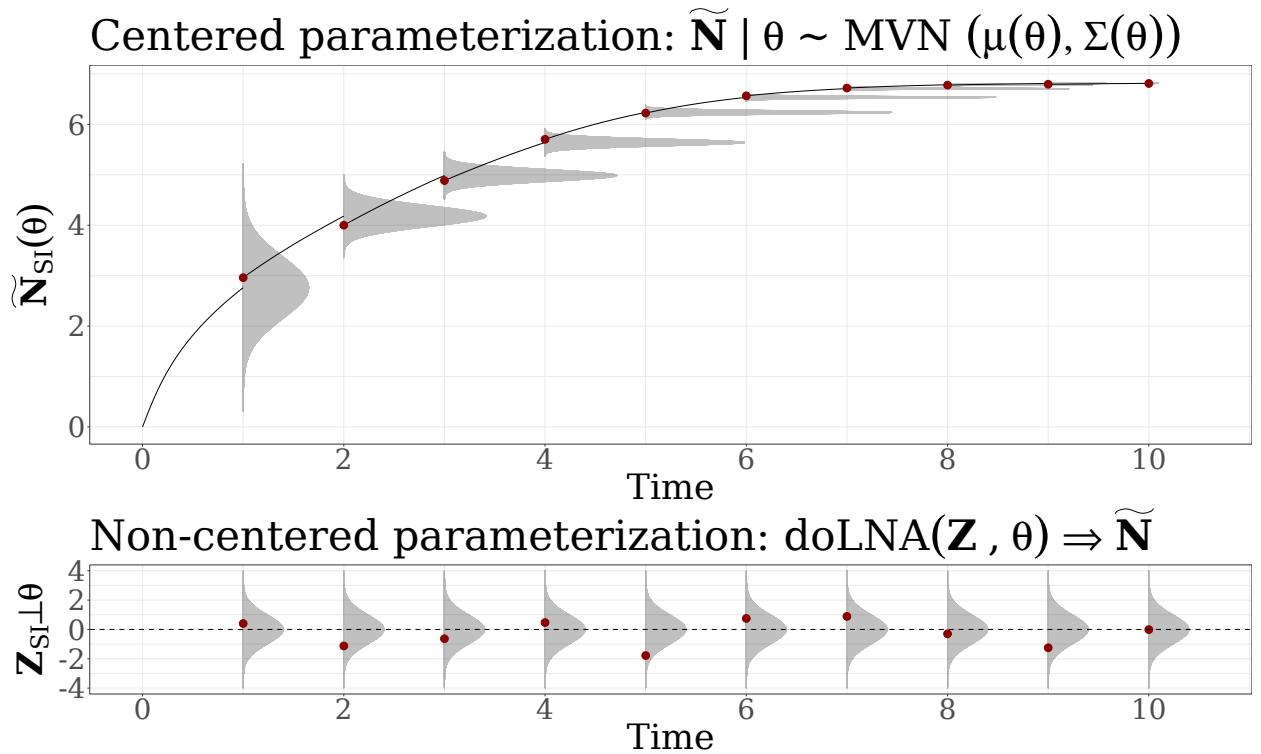


Figure 4.3: Centered (top) and non-centered (bottom) parameterizations of an LNA incidence path. In the CP, the log-incidence is normally distributed with mean and covariance obtained by solving the LNA ODEs, (4.15) and (4.17). In the NCP, the log-incidence is a draw from a standard normal distribution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

restarted since the mean of $\mathbf{N}(t_\ell)$ depends non-linearly on the value of $\mathbf{N}(t_{\ell-1})$. The quality of the LNA approximation is known to degenerate over long time intervals. Restarting the LNA ODEs has been established to improve the approximation when analyzing time series data of non-negligible length [68, 85, 77, 153]. Hence, use of the NCP is critical to enabling the use of ElliptSS for jointly updating of $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ when using the restarting version of the LNA.

We note that the elliptical slice sampling Algorithm 9 differs slightly from, but is equivalent to, the algorithm in [159] regarding how the initial proposal is made. The original algorithm was modified so that the distribution of angles for accepted proposals would be symmetric about zero to facilitate tuning of the initial bracket width. We have found that shrinking the initial bracket width often improves computational efficiency when fitting more complex models. This is discussed further in Section B.1.

Initializing the LNA Draws

In simple models, reasonable parameter values will generally lead to valid LNA paths for initial $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$, i.e., paths that satisfy the monotonicity and positivity conditions, and thus have non-zero likelihood. However, this is not necessarily the case for complex models with many types of transition events, or when the time-series of incidence counts is long. One option is to include a re-sampling step after line 6 in Algorithm 8, in which $\mathbf{Z}(t_\ell)$ is redrawn in place until the conditions for a valid path over the interval are met. It is important to note that such a procedure does not sample from the correct distribution since \mathbf{Z} is not actually a truncated multivariate Gaussian. To correct for this, we will “warm-up” the LNA path with an initial run of 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, updates to model parameters conditional on LNA draws will also preserve the validity of LNA paths.

Algorithm 9 Sampling LNA draws via elliptical slice sampling.

```

1: procedure doELLIPTSS( $\mathbf{Z}_{cur}, \boldsymbol{\theta}, \mathbf{Y}, \mathcal{I}, \omega = 2\pi$ )
2:   Sample ellipse:  $\mathbf{Z}_{prop} \sim N(\mathbf{0}, \mathbf{I})$ 
3:   Sample threshold:  $u | \mathbf{x} \sim \text{Unif}(0, L(\mathbf{Y} | \text{doLNA}(\mathbf{Z}_{cur}, \boldsymbol{\theta}, \mathcal{I})))$ 
4:   Position the bracket and make initial proposal:
   
$$\psi \sim \text{Unif}(0, \omega)$$

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

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

5:   Set  $\mathbf{Z}' \leftarrow \mathbf{Z}_{cur} \cos(\phi) + \mathbf{Z}_{prop} \sin(\phi)$ .
6:   if  $L(\mathbf{Y} | \text{doLNA}(\mathbf{Z}', \boldsymbol{\theta}, \mathcal{I})) > u$  then accept  $\mathbf{Z}'$ 
7:     return  $\mathbf{Z}'$ 
8:   else
9:     Shrink bracket and try a new angle:
10:    If:  $\phi < 0$  then:  $L_\phi \leftarrow \phi$  else:  $R_\phi \leftarrow \phi$ 
11:     $\phi \sim \text{Unif}(L_\phi, R_\phi)$ 
12:    GoTo: 5

```

Parameter Updates

Each MCMC iteration will consist of a number of ElliptSS updates, typically one but possibly 2–3 for complex models, followed by a set of parameter updates. We will generally use either a global adaptive random walk Metropolis algorithm (Algorithm 4 in [11]) or the adaptive multivariate normal slice sampler (MVNSS) presented in Section 2.2.1. In models where the initial state, \mathbf{X}_0 , is not fixed, we will assign a strongly informative prior $\mathbf{X}_0 \sim TMVN_{\mathcal{S}_X^R}(N\mathbf{p}, N(\text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^T))$, which is a truncated multivariate normal approximation to a multinomial with initial state probabilities, \mathbf{p} , constrained to the state space of compartment volumes. We update \mathbf{X}_0 jointly with the LNA path via ElliptSS. Additional details are presented in Section B.2.

We have found it helpful, for the purpose of assigning sensible priors to model parameters and for improving MCMC mixing and convergence, to parameterize the estimation scale on which the MCMC explores the parameter space in terms of how the parameters directly affect the model dynamics. For the SIR model, this might mean re-expressing the model parameters in terms of the basic reproductive number of an outbreak, $R_0 = \beta N/\mu$, and the recovery rate, μ . Additionally, we would like our estimation scale to be unconstrained and therefore sample (and either accept or reject) values for $\log(R_0)$ and $\log(\mu)$. The importance of appropriately parameterizing the estimation scale is discussed at length in Section B.3.

4.2.7 Implementation

The algorithms for approximate inference via the LNA and ODE models are implemented in the `stemr` R package, which is available from the following stable GitHub repository: <https://github.com/fintzij/stemr>. The implementation is flexible and provides facilities for specification of arbitrary SEM dynamics, a variety of emission probability distributions, and capabilities for accommodating time-varying covariates, time-varying parameters, and deterministic forcings. Computationally intensive operations are implemented in C++ via `Rcpp` and `RcppArmadillo` [65, 66]. ODE integration functions are dynamically compiled in

C++ with the help of the `odeintr` R package [120] and ODEs can be integrated using a variety of methods available in the `Odeint C++` library [4].

4.3 Simulations

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

The LNA is by no means the only approximation of the transition density of the MJP representation of a SEM. In the following subsection, we will illustrate why the LNA is an attractive choice, balancing computational cost with fidelity of the MJP approximation. We benchmark the LNA against two commonly used approximations of the MJP: the deterministic approximation given by a system of deterministic ODEs that are the functional infinite population limit of the MJP [78], and a discrete-time approximation of the MJP using a multinomial modification of the τ -leaping algorithm (MMTL) [31] to simulate epidemic paths within a particle marginal Metropolis–Hastings (PMMH) framework [10]. The ODE approximation was chosen because of its ubiquity in the study of epidemic modeling, while the MMTL approximation in combination with PMMH was chosen because of a straightforward and general implementation in the popular `pomp` package in R [125]. Arguably, the MMTL approximation is closer to the original MJP than the LNA since it preserves the discreteness of the latent state space, while the ODE approximation, being deterministic, is further removed from the MJP.

The fidelity of each approximation to the original MJP depends on the population size and the epidemic dynamics. In relative terms, outbreaks with explosive dynamics in large populations will tend to deviate less from their infinite population deterministic limits than outbreaks that occur in small populations, that are less contagious, or that are characterized by uncertainty in the probability and timing of a major outbreak. We fit SIR models to 500 datasets simulated under a range of SIR dynamics. Each dataset is simulated by drawing the model parameters from a set of prior distributions, simulating an outbreak via Gillespie’s direct algorithm [87], and finally simulating the dataset as a negative binomial sample of the

true incidence. Datasets arising from outbreaks that died off immediately were discarded and re-simulated, while datasets arising from outbreaks lasting longer than 50 epochs were truncated at 50 observations. SIR models were then fit via the LNA, ODE, and MMTL approximations under the priors from which the parameters were drawn. The simulation was repeated under three different regimes for the population size and the initial number of infected individuals, reflecting different levels of initial stochasticity in the epidemic trajectory. The priors and population sizes (Table 4.1) were chosen because they were typical of settings in which the methods might reasonably be applied, e.g., the population sizes are not so big that the outbreaks would evolve deterministically, nor so small that the approximations would be unreasonable. All individuals who were not initially infected were susceptible at the start of each outbreak (i.e., no individuals with pre-existing immunity). The population sizes and initial conditions were fixed at their true values. Hence, the only model misspecification was in the approximation used for the latent epidemic process. Additional results and details about the simulation setup are provided in Section B.5. We also performed four analogous supplementary simulations, with similar results, where we generated datasets under fixed parameter regimes (presented in Section B.6).

Results

This simulation was designed to be generous to the approximations that were used in fitting SEMs to the simulated data. The initial compartment counts and true population sizes were known, and there was no misspecification with respect to either the sampling model or the epidemic dynamics. Despite this, the ODE models struggle to reliably recover the true parameters, particularly those governing the sampling process. As shown in Figure 4.4, coverage of credible intervals for ODE models was low for all model parameters, and this was only somewhat mitigated as the population size increased. Coverage of credible intervals for models fit via the LNA and via MMTL were close to the nominal 95% levels for all model parameters in all three population size regimes. The distributions LNA point estimates and credible intervals were also similar to MMTL point estimates and credible

Table 4.1: Population sizes, initial conditions, and priors under which datasets were simulated. Five hundred datasets were simulated for each of the population size 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)	$\implies R_0 = 2.00 \text{ (1.38, 3.66)}$
$1/\mu$	Mean infectious period	LogNormal(-0.7, 0.35)	1.43 (0.72, 2.84)
$\rho/(1 - \rho)$	Odds of case detection	LogNormal(0, 1)	$\implies \rho = 0.5 \text{ (0.12, 0.88)}$
ϕ	Neg.Binom. over-dispersion	Exponential(0.1)	6.93 (0.25, 36.89)

intervals for all parameters at all three population sizes. Further inspection of the relative posterior median deviations indicate that the ODE estimates are less precise, and credible interval widths tend to be narrower than credible intervals obtained with the two stochastic approximations. This is in agreement with findings by other authors who have found that ODE models tend to underestimate uncertainty in epidemic dynamics (see e.g., [123]). Taken together, these results suggest that the LNA is, at least in this simple example, about as good at approximating the original MJP as is the more exact MMTL.

We note that these results are not intended to suggest that there is no place for ODE models in the computational toolbox of disease modelers. To the contrary, when time is of the essence, as in an outbreak setting, crude estimates via the ODE may be obtained quickly. We also have found ODE models to be extremely useful in exploring complex models and iterating through SEM parameterizations in order to identify a parameterization that allows for efficient MCMC mixing. Average ODE run times were substantially shorter than LNA and MMTL run times and required far less CPU time per effective sample (see Table 4.2). This is helped by the fact that ODE models they lend themselves to analytic characterizations

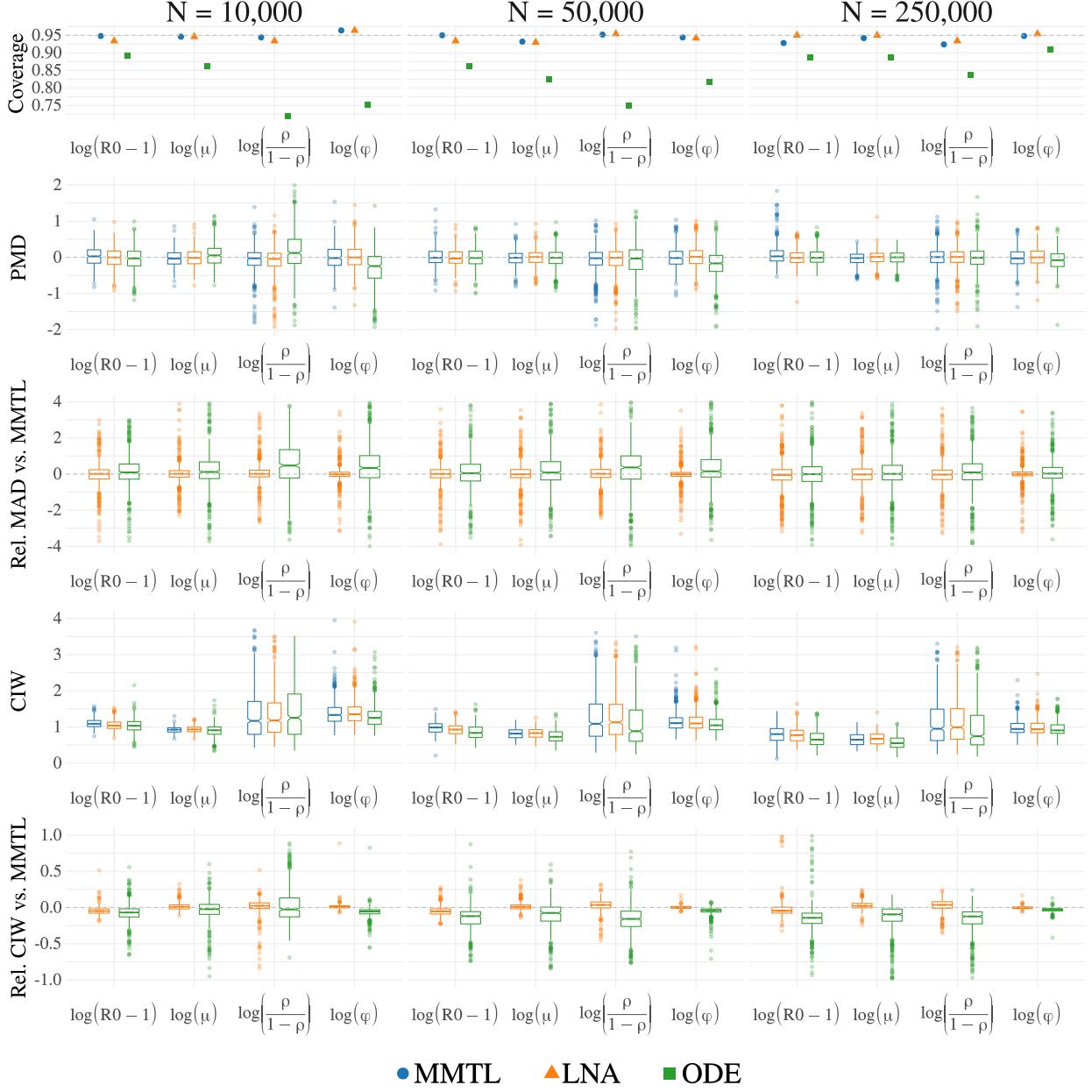


Figure 4.4: Comparison of posterior results for SIR models fit to 500 datasets via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of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

of various aspects of the outbreak dynamics, e.g., relating the final outbreak size to the basic reproductive number (see, e.g., [9, 33, 119]).

In this simple simulation, the LNA and MMTL approximations had comparable computational performance, with the LNA perhaps being somewhat faster, but also with the caveat that comparing the ODE/LNA approximations with the MMTL approximation on the basis of computational performance is a bit misleading since the comparison would have turned out differently had we made other choices for the LNA and MCMC settings (e.g., timestep of MMTL, number of particles in PMMH, or tuning the initial EllipSS bracket width for the LNA). The important point to make regarding computational performance is that as model dynamics get more complex and the time series get longer, approximations, such as MMTL, that are used within a particle filter framework, such as PMMH, will become computationally infeasible. In many cases, the lack of an adequate model from which to simulate particle paths will lead to issues of particle degeneracy and an inability to fit even simple models (see [74] for an example). Indeed, PMMH was abandoned as a computational strategy for analyzing Ebola data in later sections because of difficulty fitting SEMs with reasonable effort. However, as we shall see in the following sections, the LNA remains performant even as the model dynamics become more complex.

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

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

4.3.2 Assessing Model Fit

Although model assessment is not the focus of this work, we emphasize the importance of diagnostics and briefly present several in-sample diagnostic tools that can be used to interrogate SEMs fit via the LNA. We highlight the following diagnostics because they are easily implemented post-hoc and require little effort beyond caching random quantities throughout the MCMC run. We do not delve into out-of sample diagnostics that could be used to assess the utility of a model for forecasting because proper predictive diagnostics involve recursively re-fitting the model and forecasting new data (see [172, 104]). This is infeasible in our setting due to the computational burden of fitting each model.

One of the central objectives in infectious disease modeling is to estimate the severity and duration of an outbreak [142]. However, the true epidemic path is not observed so it is impossible to check whether it is captured by the latent posterior. This, of course, assumes that the latent process is itself identifiable, which may not be the case when estimating the effective population size (this is discussed in Section B.4). Similarly, we cannot typically check that the distribution of expected observed incidence actually captures the true mean incidence we would expect to observe if we knew the true incidence (note that $\rho \mathbf{N}_{\text{true}}$ is always identifiable in the models we consider). A standard alternative is to check whether the observed incidence is captured in the posterior predictive distribution.

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

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

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

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

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

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

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

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

In addition to validating a particular model, we can also use posterior predictive distributions to compare models. Figure 4.6 compares posterior predictive p-values (PPPs) and relative posterior predictive interval (PPIs) widths for two SIRS models (discussed further in Section 5.2) that were fit to data generated from an SIRS model with time-varying dynamics. The posterior predictive distribution of the more flexible model, which assigned a Gaussian Markov random field shrinkage prior to the time-varying basic reproduction numbers in the model, produces more accurate predictive intervals (most PPPs between 0.25 and 0.75, and fewer PPPs near 0 or 1) has substantially narrower PPI widths (almost all dots are red, many are dark red).

We are also interested in understanding the extent to which the data (and boundary

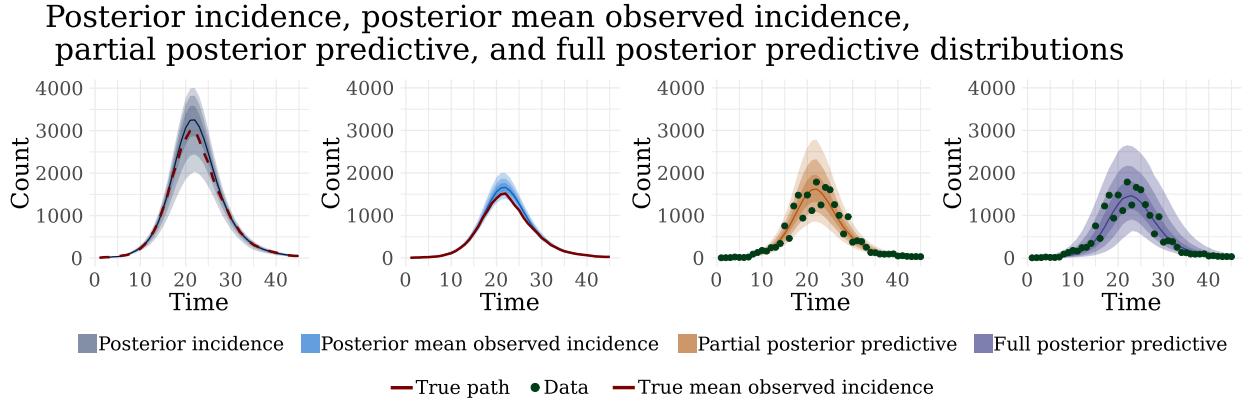


Figure 4.5: From left to right, the posterior incidence, posterior mean observed incidence, partial posterior predictive, and full posterior predictive distributions. The true incidence is the dashed red line in the leftmost plot and the observed incidence counts are the solid green dots in the right two plots. The solid red line is the true mean observed incidence, given by the true incidence multiplied by the case detection probability. The shaded bands, in order of lightest to darkest, correspond to pointwise posterior 95%, 80%, and 50% credible intervals, with the pointwise posterior median given by the corresponding colored solid line.

conditions on the latent process) suggest that there should be an excess or dearth of infections and recoveries beyond what we would expect under the LNA prior. Figure 4.7 shows the posterior distributions of LNA draws for an SIR model, along with theoretical quantiles of the posterior predictive distribution and empirical quantiles of the posterior predictive distribution, accounting for boundary conditions. The distribution in each interval, \mathcal{I}_ℓ , is shaded by the posterior shrinkage, defined as $s_\ell = 1 - \sigma_\ell$, where σ_ℓ is the posterior standard deviation of the LNA draws over \mathcal{I}_ℓ . Note that the posterior predictive distribution is the same as the LNA prior.

To the extent that the emission distribution is correctly specified, discrepancies between the prior and posterior distributions of LNA draws can point to intervals where the data are informative, intervals over which the prior and posterior are in conflict, or possibly a combination of influential observations and model misspecification. An important feature of the NCP of the LNA transition density is that it standardizes the latent posterior so that we

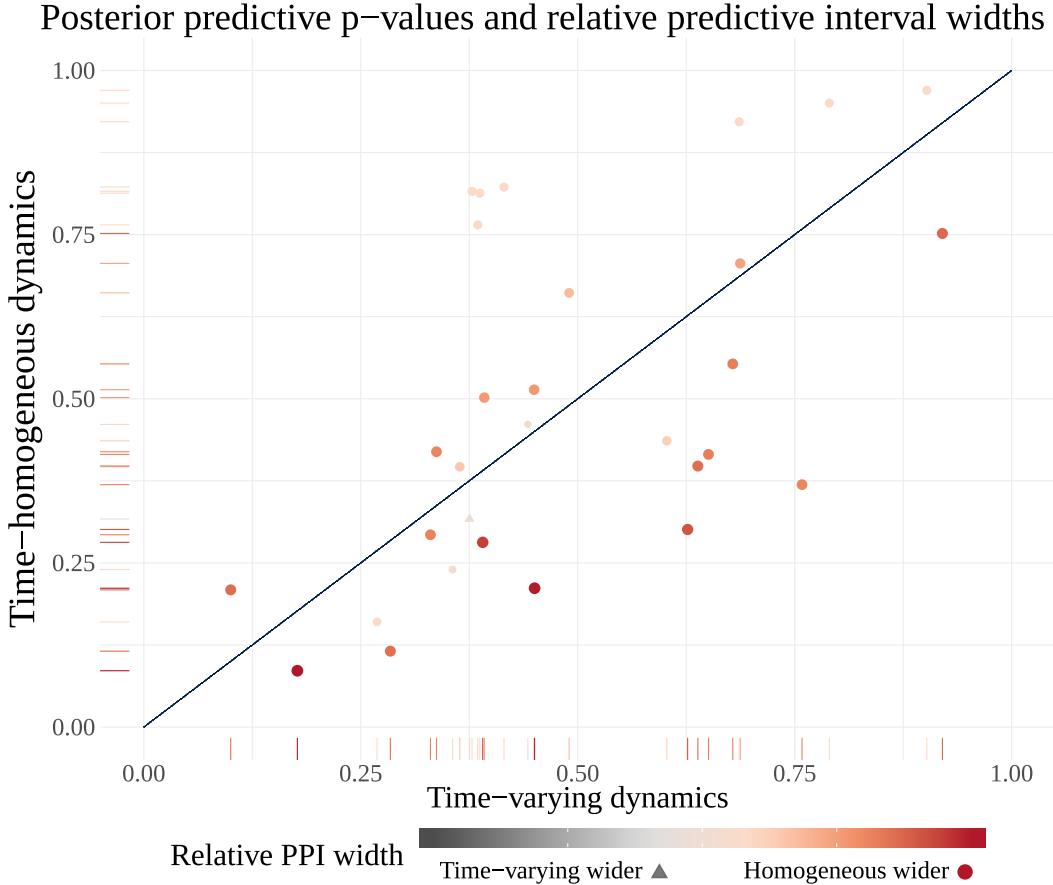


Figure 4.6: Comparison of models with time–varying and constant force of infection using posterior predictive p-values (PPPs) and relative posterior predictive interval (PPI) widths. Each point corresponds to the observed incidence in a given week. The X–Y coordinates give the PPPs under a model with a time–varying force of infection (FOI), where R0 was modeled as a Gaussian Markov random field of order one, the PPP under a model with constant per–contact infection rate. The size and color of each point corresponds to the relative PPI width, computed as $(\hat{\sigma}_{post,\ell}^{constant} - \hat{\sigma}_{post,\ell}^{RW1})/\hat{\sigma}_{post,\ell}^{RW1}$, and the sign of the relative width is further emphasized by the shape of the point. Dots indicate that PPIs with constant FOI are wider, the lone triangle corresponds to the one data point where the PPI for the model with time–varying FOI was wider.

can interpret contractions of the posterior, vis-a-vis the LNA prior, as pointing to intervals where the data add information. Shifts of the posterior, absent contraction, point to intervals over which the model dynamics do not explain the observed incidence. If we were to examine the CP of the LNA posterior, it would be difficult to identify shifts and contractions due to the mean-variance relationship of LNA transition densities.

Figure 4.7 shows pronounced shifts and contractions at the beginning and end of the outbreak. During those periods, the MJP used to simulate the outbreak is both far from its deterministic limit and is not yet well approximated by the LNA. Data at the start of the outbreak are also highly informative about the timing and intensity of the exponential growth phase of the epidemic, and hence its dynamics. In contrast, during the exponential growth and decay of the outbreak, the LNA posterior is concordant with the prior. At times when we observe a contraction with little shift in the posterior, we would conclude that the data are informative but do not suggest model misspecification. Finally we note that the boundary conditions on the latent state space have little effect the posterior predictive distribution of the latent process, except at the beginning of the outbreak (though if we would expect to see similar behavior if we extended the right tail of the observation window).

Finally, we can assess whether the model captures the temporal dependencies in the data by computing summary statistics for the observed incidence and posterior predictive incidence. A variety of summary statistics are presented in the supplement of [123], which are computed and compared for the observed and predicted data. These include the lag 1 autocorrelation (ACF) and the detrended autocorrelations at lags 1, 2, and 3. We will also consider the partial autocorrelation function (PACF) of the log-transformed data. The PACF is a measure of the residual autocorrelation at a particular lag, adjusting for the linear dependencies on the intermediate variables, and is useful in identifying the order of an autoregressive process [34]. We compute the PACF of $\tilde{\mathbf{Y}} = \log(\mathbf{Y} + 0.5)$ at lag k as

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

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

Posterior distributions of LNA draws

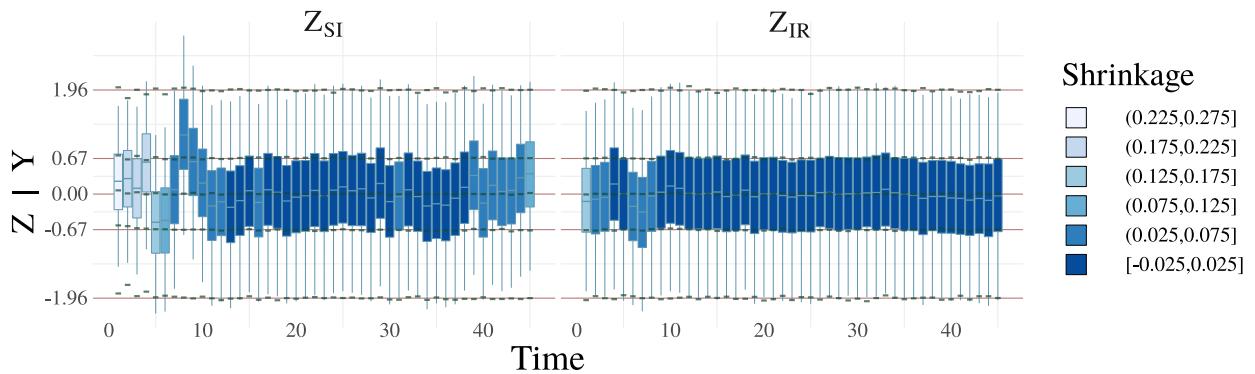


Figure 4.7: Posterior distributions of the LNA draws for infections and recoveries in an SIR model (blue boxplots). The lower and upper whisker tips correspond to the 2.5th and 97.5th posterior quantiles, the lower and upper hinges to the 25th and 75th quantiles, and the middle hash mark to the posterior median. The solid red lines are the theoretical quantiles of the posterior predictive distribution (or equivalently, the prior distribution) of the LNA draws, drawn at the quantiles of a standard normal distribution corresponding to the boxplot quantiles. The green ticks are the estimated quantiles of the posterior predictive distributions of the LNA draws, accounting for boundary conditions on the state space of the latent process and obtained by simulating LNA paths from the posterior predictive distribution. The posterior distributions of LNA draws are shaded according to the level of posterior shrinkage, computed as one minus the ratio of standard deviations of LNA draws in the posterior and prior.

where the constant 0.5 is added to stabilize the logarithm for zero counts and $\text{Proj}_{(t,k)}(\tilde{\mathbf{Y}}_{(j)})$ is the projection of $\tilde{\mathbf{Y}}_{(j)}$ onto the span of $\tilde{\mathbf{Y}}_{(t+1)}, \dots, \tilde{\mathbf{Y}}_{(t+k-1)}$ (N.B. this is equivalent to regressing $\tilde{\mathbf{Y}}_{(j)}$ on $\tilde{\mathbf{Y}}_{(t+1)}, \dots, \tilde{\mathbf{Y}}_{(t+k-1)}$).

Posterior predictive partial autocorrelations

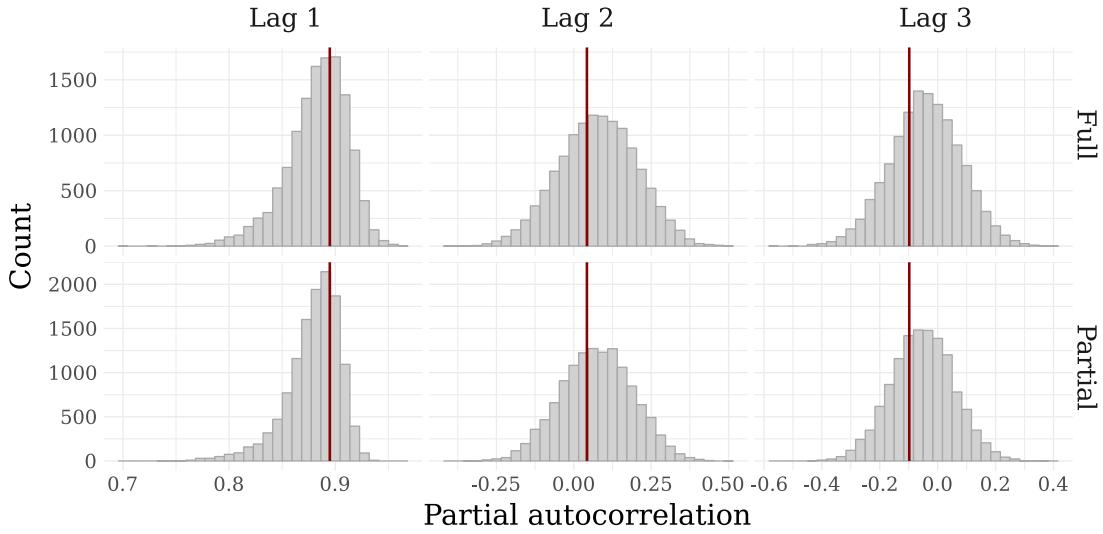


Figure 4.8: Distributions of partial autocorrelations at lags 1, 2, and 3 for datasets generated under the full and partial posterior predictive distributions. Vertical red lines are the partial autocorrelations for the observed incidence data at the respective lags.

The distributions of PACFs for the full and partial posterior predictive distributions inform us about whether the temporal structure of the data is reflected in the posterior. A reasonable data generating mechanism should produce data replications that have similar autoregressive structure as the observed data. We might worry about whether the model fails to explain residual dependence in the data if we find that the partial autocorrelation of the data at a particular lag is an outlier with respect to the predictive distribution. This would be particularly concerning if the partial predictive PACF distribution which is conditioned on the data did not reflect the autoregressive structure of the data, especially if it occurs at lag 1, since this would suggest that the emission distribution is misspecified. Figure 4.8 shows

the PACFs generated from the full and partial posterior predictive distributions along with the partial autocorrelations for the observed data. As we expect, since there was no model misspecification, the residual autocorrelation structure in the data appears to be captured by the model.

4.3.3 A Stratified SEIR Model for a Simulated Outbreak

In the next section, we will jointly model the spread of Ebola in Guinea, Liberia, and Sierra Leone during the 2014–2015 outbreak. To do so, we will fit a stratified SEIR model that allows for country specific epidemic dynamics, and transmission between the countries. It should go without saying that this model (or any model), despite being fully stochastic in all aspects of the measurement and epidemic process, is still an overly simplistic abstraction of the outbreak with respect to some obvious and potentially important complexities, e.g., spatial heterogeneity and time varying dynamics. However, it is a good starting point and is non-trivial to fit due to the number of parameters and size of the latent space. Moreover, this simulation will allow us to work with the diagnostics discussed in the previous subsection to develop better intuition for the strengths and weaknesses of the model.

We simulated data from a three country SEIR model, diagrammed in Figure B.5, and described briefly below, with additional details in Section B.7.1. The simulation was designed to crudely reflect the Ebola outbreak in West Africa and allowed for cross-border transmission via the migration of infected individuals between countries. The three countries differed in their true and effective population sizes, times at which transmission commenced, epidemic dynamics, and case detection probabilities. We included a single change-point for the basic reproductive number for Guinea at 33 weeks since the outbreak duration was longest in that country. The change in transmission dynamics coincided roughly with the WHO declaration of a state of emergency in early August 2014 and the closures of country borders with Liberia and Sierra Leone [52]. The outbreak was simulated from a MJP via Gillespie’s direct algorithm [87] using the parameters in Table B.17 and the rates in Table B.19. The observed incidence was a negative binomial sample of the true incidence with the case

detection probability varying across countries, but with the same value of over-dispersion.

We fit a stratified SEIR model to the simulated data that was largely the same as the model used in simulating the data, except with respect to migration of infected individuals from one country to another. In this example, as we would expect for the real-world Ebola outbreak, transmission to susceptibles was largely driven by contacts with natively infected individuals within the same country. Here, we approximate the contribution of cross-border transmission to the local force of infection by replacing migrations of infected individuals with virtual migrations that modify the effective number of infected individuals in a country. Hence, we replace the rates at which individuals become infected in country A and import infections from countries B and C,

$$\lambda_{SE}^A(t) = \beta_A S_A I_A \mathbb{1}_{\{t \geq t_0^A\}}, \quad (4.27)$$

$$\lambda_I^{BA}(t) = \alpha_B I_B \mathbb{1}_{\{t \geq t_0^A\}} \mathbb{1}_{\{t \geq t_0^B\}}, \quad \lambda_I^{CA}(t) = \alpha_C I_C \mathbb{1}_{\{t \geq t_0^A\}} \mathbb{1}_{\{t \geq t_0^C\}}, \quad (4.28)$$

with the approximate rates,

$$\lambda_{SE}^A = \beta_A S_A \left(I_A + \alpha_{BA} I_B \mathbb{1}_{\{t \geq t_0^B\}} + \alpha_{CA} I_C \mathbb{1}_{\{t \geq t_0^C\}} \right) \mathbb{1}_{\{t \geq t_0^A\}} \quad (4.29)$$

$$\lambda_I^{BA} = \lambda_I^{CA} = 0, \quad (4.30)$$

where β_A is the per-contact rate of infection in country A, α are parameters for rates of real or virtual migration of infected individuals from countries B and C, t_0 is the time at which transmission was assumed to commence in each country. We note that the models are approximately equivalent to the extent that the infectious period durations in each country are similar and that the contribution to the local force of infection from other countries is relatively minimal. The upside is that the quantities, $\alpha_{BA} I_B$ and $\alpha_{CA} I_C$, are interpretable as the effective number of infected individuals from countries B and C with whom susceptibles in country A come into contact. The approximate model is also more computationally tractable since we have fewer ODEs to solve in the LNA approximation and fewer boundary conditions on the LNA state space. Finally, the approximate model is arguably more appropriate from the standpoint of the LNA approximation since the rates of infectious migration events are

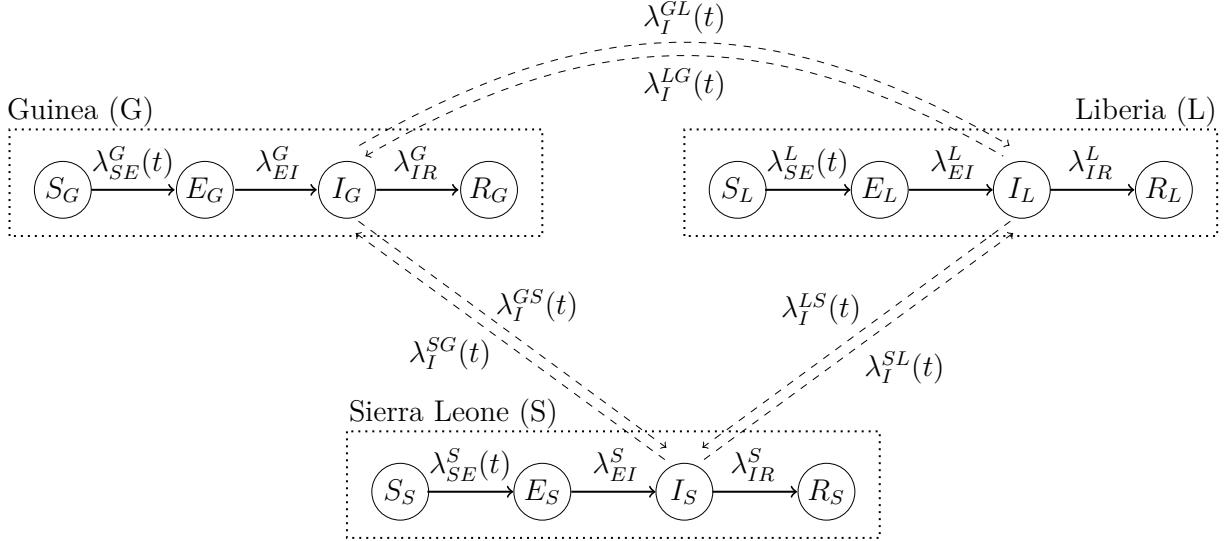


Figure 4.9: Diagram of state transitions for a stratified SEIR model fit to a simulated Ebola outbreak in Guinea, Liberia, and Sierra Leone. Dashed boxes denote countries, nodes in circles denote the model compartments: susceptible (S), exposed (E), infectious (I), recovered (R). Compartments are subscripted with country indicators. Solid lines with arrows indicate stochastic transitions between model compartments, which occur continuously in time. Dashed lines indicate virtual migrations, where infected individuals in one country contribute to the force of infection in another country but do not migrate. Rates at which individuals transition between compartments are denoted by λ and are subscripted by compartments and superscripted by countries, e.g., λ_{SE}^L is the rate at which susceptible individuals become exposed in Liberia, λ_I^{SG} is the rate at which an infected individual in Sierra Leone migrates to Guinea. The rate at which susceptibles in Guinea become infected is time varying with one change-point at 33 weeks. Transmission in Liberia and Sierra Leone commenced at 10 and 19 weeks, respectively. Full expressions for the rates are given in Table B.20.

low. The full model with explicit migration was fit as a supplementary analysis and yielded generally similar results (see Table B.22), other than cross–border transmission parameters, which were poorly estimated.

Before proceeding, we pause to note that the effective population size and case detection probability and effective population size are only weakly identifiable in this model. Rather, the mean case detection rate, ρN_{adj} , is estimable because of its relationship to the true and observed outbreak sizes. In practice, we require prior information about either ρ or N_{adj} in order to estimate the parameters separately. This is discussed further in Section B.4.

We fit the model using a combination of weakly informative priors for the adjusted reproductive numbers, informative priors for the latent period durations, infectious period durations, and cross–border transmission parameters, and diffuse priors for the measurement process parameters and effective population sizes (Table B.17). The 95% credible intervals contained the true values (Figure 4.10) for all parameters, or identifiable functions as for the case detection rate. We essentially recover the priors for parameters that were assigned informative priors.

Although the latent epidemic process is not identifiable in this model, we recover, for the most part, the mean incidence we would have expected to observe if the true incidence were known (Figure 4.11). The partial posterior predictive distributions capture the observed incidence curves, suggesting that the emission distribution adequately described the sampling process for the observed incidence, conditional on the latent incidence being distributed according to its posterior. The full posterior predictive distributions for Guinea and Sierra Leone suggest that the model may be a reasonable data generating mechanism for the outbreaks in those countries. However, the full posterior predictive distribution for Liberia predicts slower dynamics than were actually observed, which is reflective of an estimated adjusted reproductive number for Liberia that was somewhat lower than the true value. There is also some residual autocorrelation in the log transformed data for Liberia that was not well captured in the posterior predictive distributions (Figure 4.12), unlike Guinea and Sierra Leone. Finally, the posterior distributions of the LNA draws suggest that

the LNA approximation might not fully account for the uncertainty in the latent epidemic process, particularly in Guinea and Liberia. This is expected since the effective population sizes were not particularly large and the data included incidence from the beginning and end of the outbreak, when the relative stochasticity was large. The posterior LNA draws for Guinea, which had the smallest effective population size, are more variable than we would expect under the LNA prior. The posterior and posterior predictive draws for Liberia exhibit systematic deviations from normality during the periods when transmission commenced and tailed off.

Posterior distributions of stratified SEIR model parameters

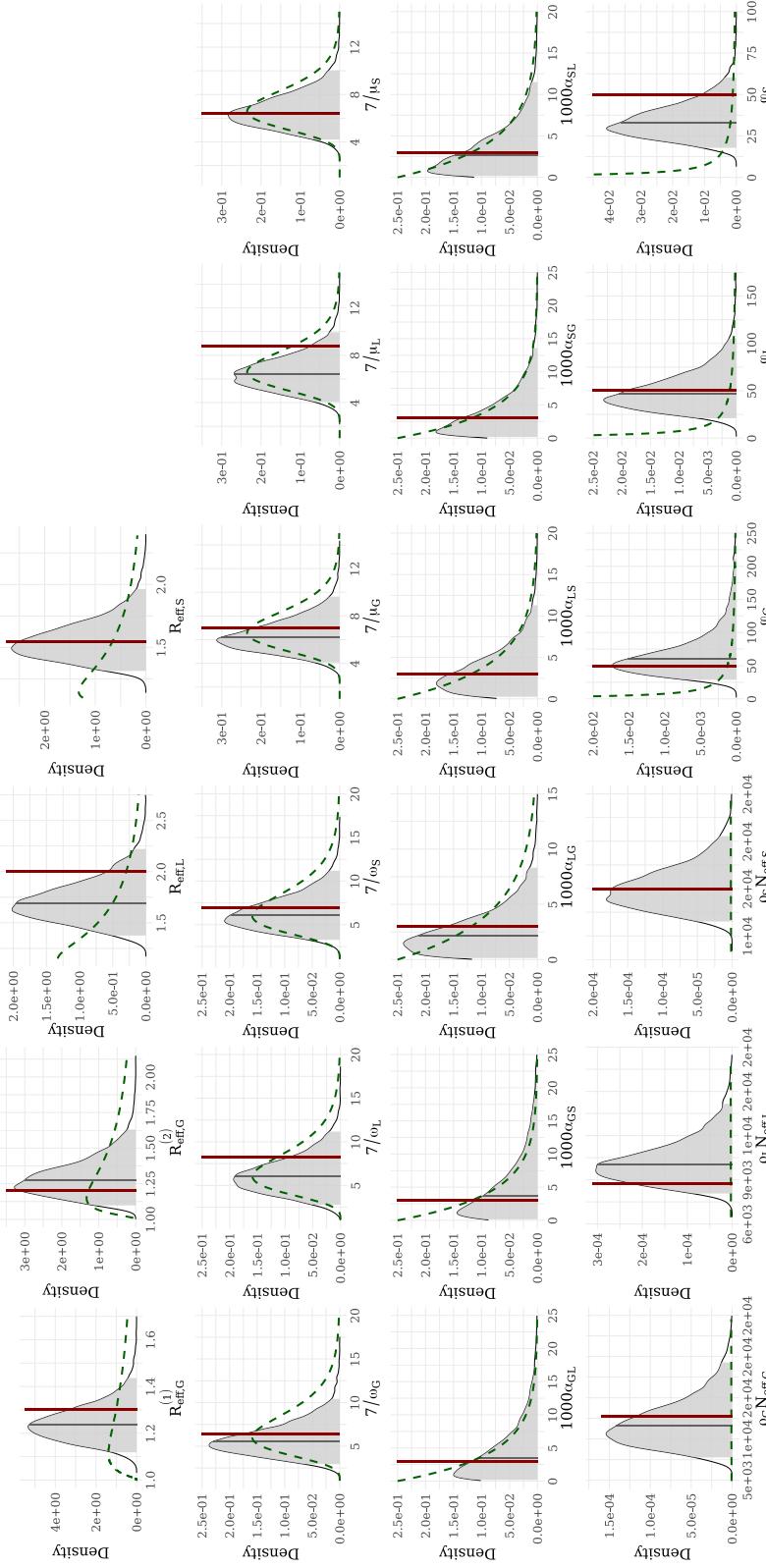


Figure 4.10: Posterior distributions of parameters of a stratified SEIR model for a simulated Ebola outbreak in three countries. We show posterior medians (solid gray lines), 95% Bayesian credible intervals (light gray areas under the posterior densities), prior densities (induced priors for the reporting rate and latent period durations) over the posterior ranges (dashed green curves), and true parameter values (solid red lines). $R_{\text{adj}} = \beta N_{\text{adj}} / \mu$ is the adjusted reproductive number, where β is the per-contact infection rate, N_{adj} is the effective population size, and μ is the recovery rate. The latent and infectious period durations, $7/\omega$ and $7/\mu$, respectively, are given in days. The effective number of cross-border infectious contacts in country B per 1000 infected individuals in country A is $1000\alpha_{AB}$. The effective reporting rate is ρN_{adj} , and ϕ is the negative binomial over-dispersion parameter. Subscripts indicate countries.

Posterior mean observed incidence, partial posterior predictive, and full posterior predictive distributions

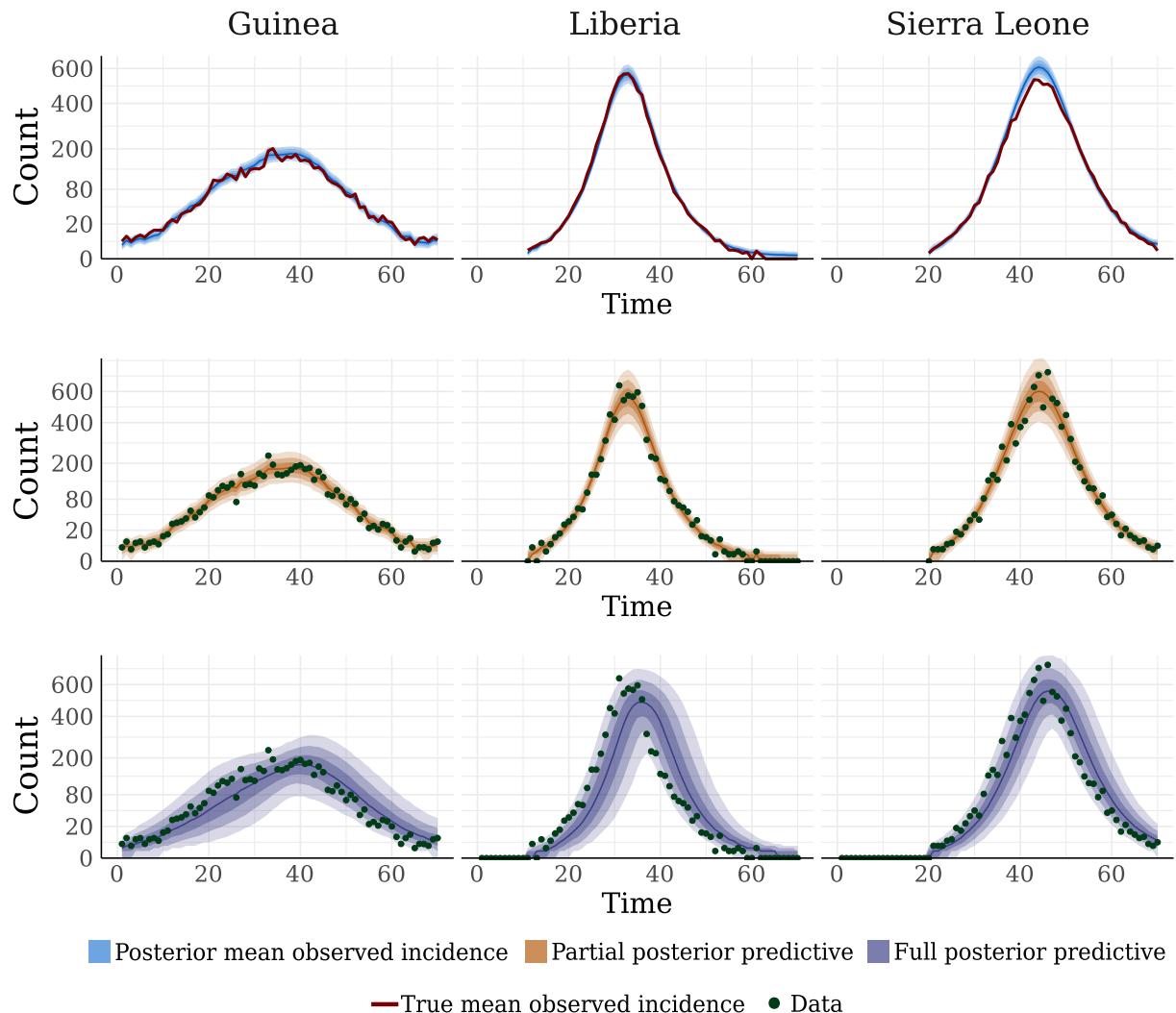


Figure 4.11: From top to bottom, the posterior mean observed incidence, partial posterior predictive, and full posterior predictive distributions. The solid red line is the incidence we would expect to observe if we knew the true incidence. Green points correspond to the observed incidence. The shaded bands, in order of lightest to darkest, correspond to pointwise posterior 95%, 80%, and 50% credible intervals, with the pointwise posterior median given by the corresponding colored solid line.

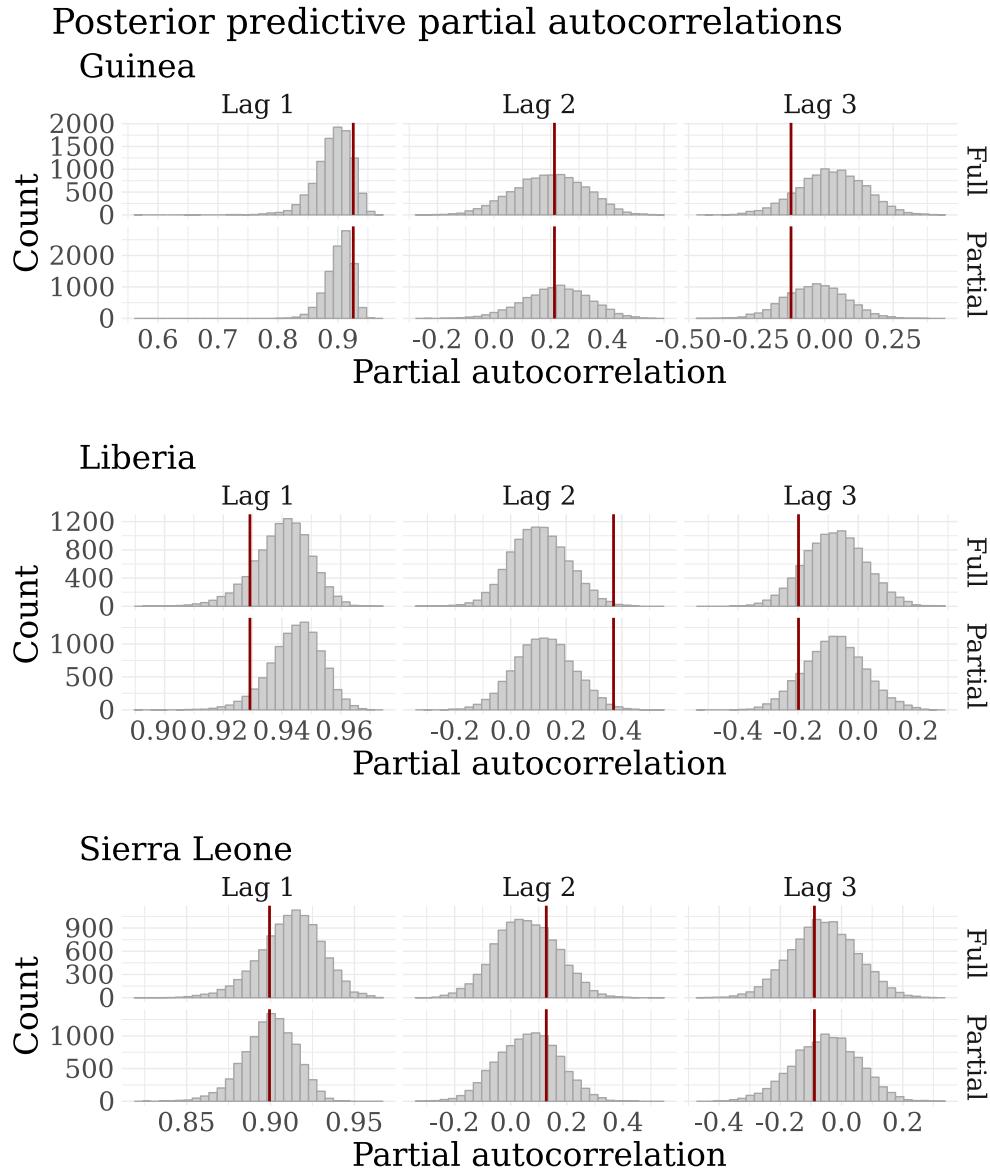


Figure 4.12: Distributions of partial autocorrelations at lags 1, 2, and 3 for datasets generated under the full and partial posterior predictive distributions. Vertical red lines are the partial autocorrelations for the observed incidence data at the respective lags.

Posterior distributions of LNA draws for a joint Ebola model

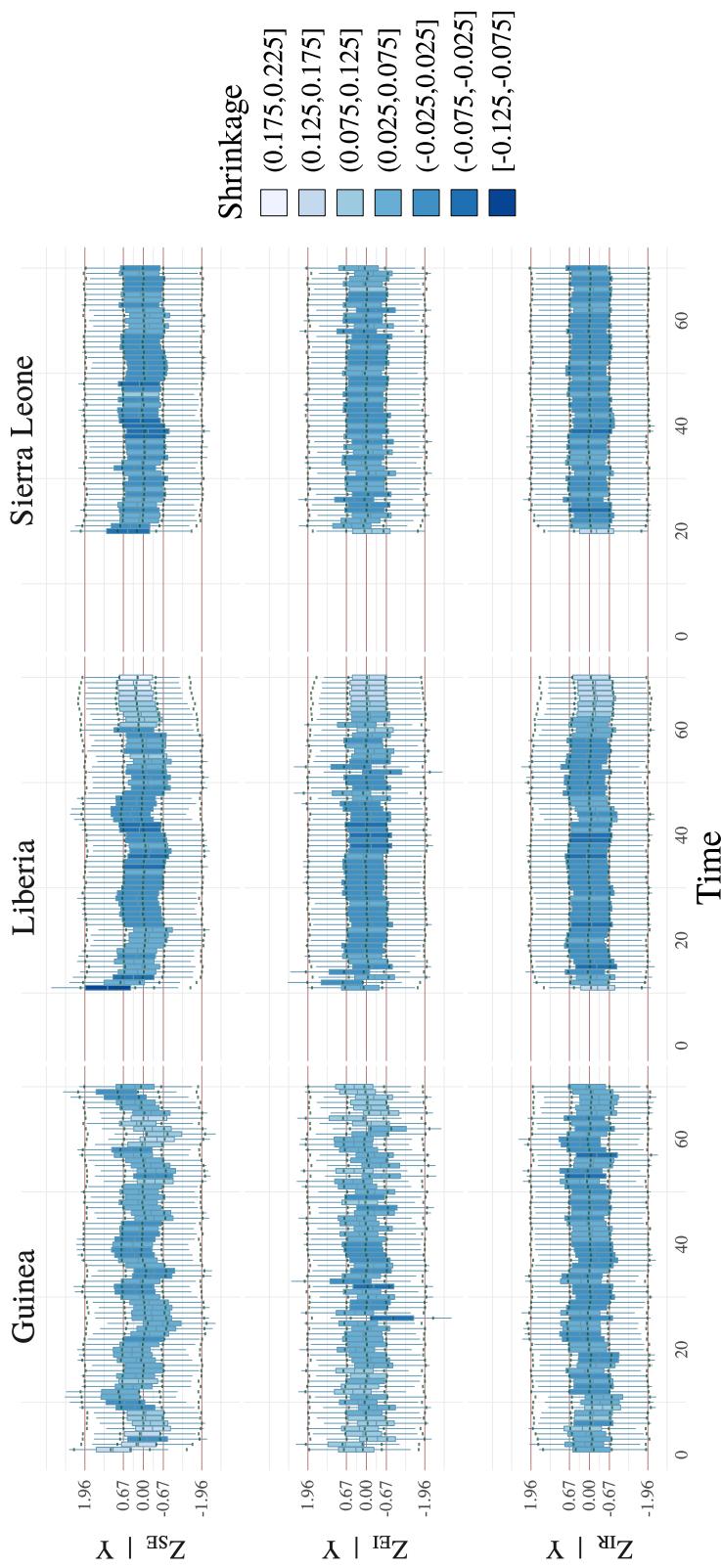


Figure 4.13: Posterior distributions of the LNA draws for exposures, infections, and recoveries (blue boxplots) in a stratified SEIR model fit to a simulated Ebola outbreak in Guinea, Liberia, and Sierra Leone. The lower and upper whisker tips correspond to the 2.5th and 97.5th posterior quantiles, the lower and upper hinges to the 25th and 75th quantiles, and the middle hash mark to the posterior median. The solid red lines are the theoretical quantiles of the posterior predictive distribution (or equivalently, the prior distribution) of the LNA draws, drawn at the quantiles of a standard normal distribution corresponding to the boxplot quantiles. The green ticks are the estimated quantiles of the posterior predictive distributions of the LNA draws, accounting for boundary conditions on the state space of the latent process and obtained by simulating LNA paths from the posterior predictive distribution. The posterior distributions of LNA draws are shaded according to the level of posterior shrinkage, computed as one minus the ratio of standard deviations of LNA draws in the posterior and prior.

4.4 Application: Modeling the Spread of Ebola

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 number. The data, shown in Figure 4.14 and described in Section 1.3.2, consist of national case counts from the World Health Organization patient database consisting of weekly confirmed and probable Ebola cases [224]. 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 consisted of 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 [52]. For the purpose of this analysis, we will lump together confirmed and probable cases.

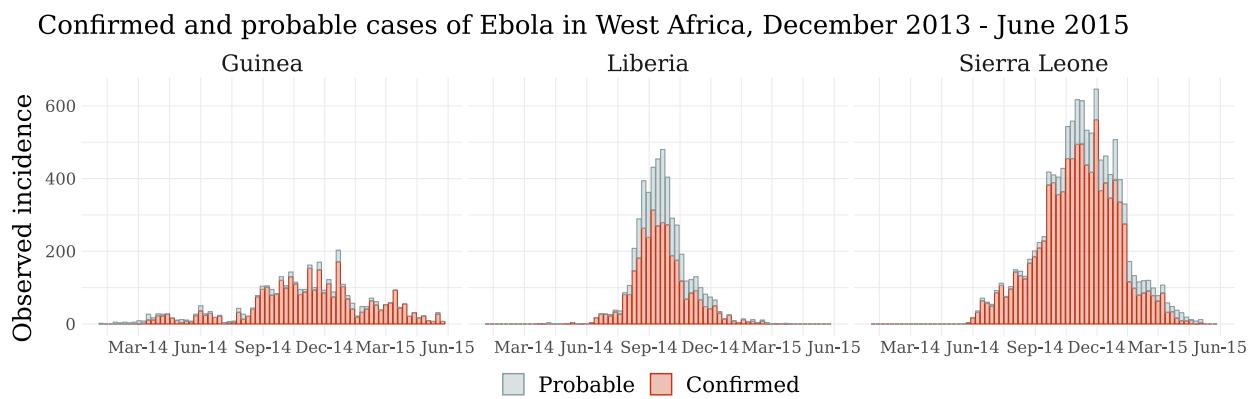


Figure 4.14: Weekly incidence of confirmed and probable cases of Ebola in Guinea, Liberia, and Sierra Leone.

4.4.1 Country-Specific and Joint Models for Ebola in West Africa

Country-Specific Models

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 joint 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 and that will enable us to fit a more complicated joint model discussed in the next section (see Section B.3 for further discussion MCMC parameterization).

We fit separate SEIR models, diagrammed in Figure 4.15, 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. Phylogeographic models fit to viral sequence data collected during the outbreak suggested that Guinea experienced re-importation of Ebola throughout the outbreak from Liberia and Sierra Leone [63]. Therefore, we included a single change-point for the adjusted reproductive number for Guinea to account for possible changes in the outbreak dynamics. The change-point was set to August 10, 2014, two days after the WHO declared an international state of emergency, and one day after borders were closed with Sierra Leone and Liberia [52]. The force of infection in each country included a constant term for infectious contact from outside the population. To account for the small scale of each outbreak relative to the population size in the country, we estimate the effective population size as a parameter in the model. The number of susceptible individuals is then equal to the effective population size, minus the numbers of exposed, infected, and recovered individuals. The implications of this choice for identifiability of parameters and the latent epidemic process are discussed in Section B.4. 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 (4.1).

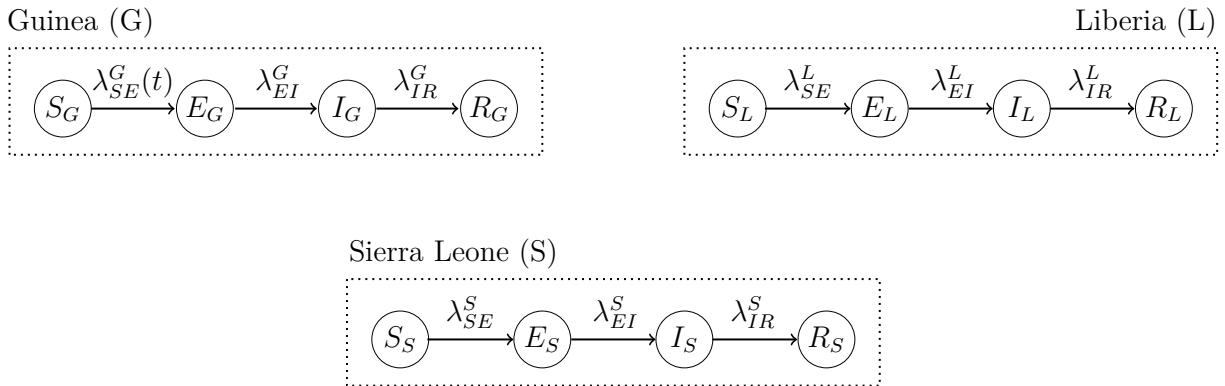


Figure 4.15: Diagram of state transitions for SEIR models fit to Ebola incidence data from Guinea, Liberia, and Sierra Leone. Dashed boxes denote countries, nodes in circles denote the model compartments: susceptible (S), exposed (E), infectious (I), recovered (R). Compartments are subscripted with country 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_{SE}^L is the rate at which susceptible individuals become exposed in Liberia. The rate at which susceptibles in Guinea become infected is time varying with one change-point at 33 weeks. Transmission in Liberia and Sierra Leone commenced at 10 and 19 weeks, respectively. Full expressions for the rates are given in Table B.24.

Table 4.3: Population sizes for Guinea, Liberia, and Sierra Leone [207], and times at which within-country transmission was assumed to begin. Week zero was the time at which transmission was assumed to commence in each country, and corresponded to the first reported case in Guinea, and to three weeks prior to the first reported cases in Liberia and Sierra Leone.

	Guinea	Liberia	Sierra Leone
True population size (N)	11.8 million	4.4 million	7.1 million
Week zero (epiweek, 2014)	2	12	21

We fit the models under identical sets of informative priors, listed along with references in Table B.26, for parameters governing the outbreak dynamics, and diffuse priors for parameters governing the emission distribution, including the effective population sizes. The priors for adjusted reproduction numbers, and sojourn time distributions in the exposed and infectious compartments were based on previously published estimates. The prior for the baseline rate of infectious contact from outside the population reflected 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 [63], who estimated between one-half to two dozen reintroduction events, depending on the country, between April 2014 and May 2015. We were agnostic in our priors for the mean case detection probability, effective population size, and the over-dispersion parameter in the emission distribution. The prior for the detection probability reflected our belief that a case was unlikely to be detected with probability nearly equal to zero or one, but also our uncertainty about the reliability of published estimates. Our priors for the effective population sizes were uniformly distributed between the total observed incidence and one quarter of the population in each country. Results for a supplementary model, fit under more diffuse priors for the outbreak dynamics, are presented in Table B.29.

A Joint Model for the Spread of Ebola in West Africa

An inherent limitation of the country-specific SEIR models is that they cannot capture the time-varying interactions between countries resulting in cross-border transmission. This would still be the case, even if we had allowed the force of infection from outside each to be time-varying, since the exogenous rates of infection were divorced from the outbreak states in the neighboring countries. Hence, we also fit a country-stratified SEIR model, diagrammed in Figure 4.9, with cross-border transmission modeled via virtual migration of infectious individuals using the LNA and ODE approximations. The model allowed for country-specific outbreak dynamics and emission distributions. As with the country-specific models, transmission was assumed to commence in Liberia at the beginning of March 2014, and in Sierra Leone at the beginning of May 2014, and a change-point in the basic reproductive number for Guinea was set at the beginning of August 2014. The observed incidence in each country was again modeled as a negative binomial sample of the true incidence, with country-specific mean detection probabilities and over-dispersion parameters. We fit the model using the same informative priors for the rate and cross-border transmission parameters as the country-specific models (summarized and sourced in Table B.35). MCMC details are presented in Section B.8.2. Supplementary results for a model using more diffuse priors for the outbreak dynamics are presented in B.38.

4.4.2 Results

Country-specific and joint models fit via the LNA

Estimates of Ebola transmission dynamics obtained using the country-specific and joint SEIR models (Table 4.4) are consistent with previously published results obtained with stochastic models fit to aggregate incidence data [51]. We find that there is significant variability in the outbreak dynamics among the three countries. The adjusted reproduction number was highest in Liberia, followed by Sierra Leone, and then Guinea. The mean sojourn times in the exposed and infectious compartments were shorter in Sierra Leone than in Guinea and

Liberia, which were more similar to estimates in the literature [212]. The contribution to the force of infection in Guinea from infected individuals in Liberia and Sierra Leone in the joint model, or generically in the country-specific model, was higher than the exogenous contributions to the force of infection in the other countries. This is consistent with findings that Ebola was re-introduced to Guinea at various times throughout the outbreak [63].

The rate of infectious contact from cross-border transmission in the joint model was tightly constrained in the prior, with 95% of the mass below twelve effective infected individuals per 1000 foreign infecteds. However, even this modest level of cross-border transmission led to discernible differences in estimates of the outbreak dynamics. Estimates of the adjusted reproduction numbers were lower, and mean sojourn periods in the exposed and infected states were shorter, under the joint model than estimates obtained from country-specific models. The starker differences were in posterior estimates of the baseline rate of infectious contact (for the country-specific models) and the rates of cross-border infectious contact (for the joint model). In the country-specific models, the baseline rate of infectious contact barely contributed to the force of infection, effectively equivalent to thousandths of an additional infected at each point in time. This is an artifact of the external rate of infection being constant throughout the outbreak, and decoupled from the prevalence in other countries. This rate was forced to be small in order to time the exponential growth phase of the outbreak, which meant that persistence of the outbreak was driven by within-country infectious contacts. One consequence of this was that estimates of the adjusted reproductive number in Guinea after August 10, 2014, was higher than before that date. This might be surprising given that various control and response measures were implemented in the summer of 2014 [52]. In contrast, the joint model allows the force of infection arising outside each country to vary over time along with the prevalence in neighboring countries. This enables the joint model to explain the major outbreak in Guinea, beginning in the summer of 2014, as the results of cross-border transmission rather than, with high probability, an up-tick in the rate of infectious contact.

The country-specific models and the joint model exhibit similar, reasonably good per-

Table 4.4: Posterior medians (95% Bayesian credible intervals) for parameters controlling the outbreak dynamics in country-specific SEIR models (Figure 4.15) and a joint country-stratified SEIR model fit to Ebola outbreak data. Subscripts, G, L, S , indicate specific countries, or generic countries A, B . Adjusted reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{adj}/\mu$. Priors are given in Tables B.26 and B.35.

Parameter	Interpretation	Posterior median (95% BCI)			
		Country-Specific Models		Joint Models	
		LNA	ODE	LNA	ODE
$R_{adj,G}^{(1)}$	Adjusted reproduction #, $t \leq 8/10/2014$	1.2 (1.1, 1.4)	1.2 (1.1, 1.4)	1.1 (1.0, 1.3)	1.2 (1.1, 1.4)
$R_{adj,G}^{(2)}$	Adjusted reproduction #, $t > 8/10/2014$	1.4 (1.2, 1.8)	1.5 (1.3, 1.9)	1.2 (1.1, 1.4)	1.3 (1.1, 1.5)
$7/\omega_G$	Mean latent period, days	8.4 (4.0, 17.2)	10.4 (5.1, 19.9)	7.5 (3.9, 13.6)	7.8 (4.2, 13.9)
$7/\mu_G$	Mean infectious period, days	9.0 (5.9, 13.4)	10 (6.8, 14.4)	8.4 (5.9, 11.8)	8.7 (6.2, 12.3)
$1000\alpha_G$	Effective # of baseline infecteds $\times 1,000$	0.2 (0.0, 0.6)	0.3 (0.1, 0.6)	—	—
$1000\alpha_{LG}$	Effective # of infecteds in Guinea per 1,000 infected in Liberia/Sierra Leone	—	—	3.9 (0.5, 14.2)	1.7 (0.1, 7)
$1000\alpha_{SG}$	1,000 infected in Guinea/Sierra Leone	—	—	2.6 (0.1, 13.5)	2.6 (0.1, 13.5)
$R_{adj,L}$	Adjusted reproduction #	2.2 (1.7, 3.3)	2.4 (1.7, 3.5)	2.1 (1.6, 3.3)	2.2 (1.7, 3.2)
$7/\omega_L$	Mean latent period, days	10.5 (4.8, 19.6)	11.9 (5.6, 20.5)	9.0 (4.1, 19.9)	10.5 (4.9, 19.1)
$7/\mu_L$	Mean infectious period, days	9.7 (6.6, 14.1)	10.2 (7.1, 14.6)	9.0 (6.0, 13.2)	9.6 (6.6, 13.8)
$1000\alpha_L$	Effective # of baseline infecteds $\times 1,000$	0.0 (0.0, 0.1)	0.0 (0.0, 0.00)	—	—
$1000\alpha_{GL}$	Effective # of infecteds in Liberia per 1,000 infected in Guinea/Sierra Leone	—	—	3.0 (0.1, 15.5)	2.3 (0.1, 12.5)
$1000\alpha_{SL}$	1,000 infected in Guinea/Sierra Leone	—	—	3.0 (0.1, 16.7)	2.9 (0.1, 15.4)
$R_{adj,S}$	Adjusted reproduction #	1.3 (1.2, 1.5)	1.8 (1.5, 2.3)	1.3 (1.1, 1.4)	1.3 (1.2, 1.5)
$7/\omega_S$	Mean latent period, days	4.5 (2.6, 7.4)	10.8 (5.9, 18.6)	4.1 (2.4, 6.6)	3.9 (2.2, 6.6)
$7/\mu_S$	Mean infectious period, days	6.2 (4.5, 8.5)	10.1 (7, 14.4)	6.1 (4.4, 8.2)	5.9 (4.3, 8.2)
$1000\alpha_S$	Effective # of infecteds per 1,000	0.0 (0.0, 0.1)	0.4 (0.2, 0.8)	—	—
$1000\alpha_{GS}$	Effective # of infecteds in Sierra Leone	—	—	2.6 (0.1, 13.4)	4.7 (0.2, 24.3)
$1000\alpha_{LS}$	per 1,000 infected in Guinea/Liberia	—	—	1.4 (0.1, 7.2)	0.7 (0, 5.2)

formance in reconstructing the observed outbreak peaks in Guinea and Liberia, as well as the exponential growth and decay in the observed incidence curves. However, both models perform poorly for Sierra Leone. The joint model, in particular, exhibits several indications that point to severe model misspecification. First, the full posterior predictive distribution for Sierra Leone (Figure 4.17) suggests that the model underestimates the incidence during the first six months of the outbreak, and overestimates the observed incidence in the last six months. The posterior predicted incidence appears to not only miss the mark when considered point-wise, but also fails to match the curvature of the data. This is confirmed by the distribution of posterior predictive partial autocorrelations at lag 1, which indicates that the model is producing much stronger first order autocorrelations from week-to-week than were observed in the data (Figure B.31). When we compare the posterior predictive p-values for the two models (Figure 4.20), we see that the observed data are not as unlikely under the country-specific model for Sierra Leone vis-a-vis the joint model. However, this appears to largely be the results of wider PPI widths as the country-specific model exhibits the same failure to reconstruct the autocorrelation structure of the data as does the joint model (Figure B.17).

The posterior distributions of the LNA draws for the joint and country-specific model identify the early weeks of the outbreak in Sierra Leone as exhibiting the most severe misspecification (Figures B.18 and B.32). Both models indicate that there should be substantially more exposures *a posteriori* than we would expect under either model. The posterior distributions of the LNA draws for the weeks just after the peak of the outbreak in Sierra Leone (around January, 2015) also indicate fewer exposures than are expected under either the joint or country-specific model (i.e., that the outbreak is dying off sooner than expected). There are two explanations for this behavior, possibly both correct. First, it is likely that the distribution for compartment counts in Sierra Leone at the start of the modeling period in that country was misspecified. In particular, we probably assumed that there were too few initially infected and exposed individuals, which would explain both the excessive rate of exposure early on, and the unexpected dearth of exposures after the outbreak

peak (the result of having to interrupt dynamics under which sustained transmission would otherwise be expected). Assigning a more diffuse prior the initial compartment volumes would be an appropriate way to remedy the problem. For example, we could allow for some over-dispersion in the initial compartment volumes by using the normal approximation to a dirichlet–multinomial distribution instead of the normal approximation to a multinomial (Section B.2).

In addition to misspecification of the initial distribution for Sierra Leone, the rate of infection from outside the country could also be misspecified. This rate is modeled differently in the joint model, where it covaries with the prevalence in neighboring countries, than in the country-specific model, where it is more nearly constant. It is possible that tying the rate of cross-border transmission to the prevalence in neighboring countries is having the effect of depressing the adjusted reproduction number in order to match the timing of the exponential growth of the outbreak. Transmission in Guinea and Liberia was ongoing at the beginning of the modeling period for Sierra Leone. Therefore, the rate of cross-border transmission might be too high and artificially depressing the adjusted reproduction number. One solution might be to reparameterize the model so that the rate of cross-border transmission into each country is tied to the adjusted population size in that country, as opposed to having it be a normalized rate of importation per number of infected individuals in neighboring countries.

Comparison with ODE Models

The estimated latent and infectious period durations in single country models were longer under the ODE, particularly in the case of Sierra Leone. This resulted in higher estimates of the adjusted reproduction number under the ODE, again in particular for Sierra Leone. Estimates obtained using the ODE and LNA are more similar for the joint models than for the country-specific models. Similar to findings in [123], posterior predictive intervals produced by ODE models are generally narrower than the corresponding predictive intervals produced by their stochastic counterparts (Figure 4.21). The ODE posterior predictive p-values for Guinea and Liberia are generally comparable, though the the data fall slightly

farther in the tails of the ODE PPIs for Liberia. The LNA PPIs for Sierra Leone under the joint model are narrower at the peak of the outbreak and wider during its exponential growth and decay. Note that we should not over-interpret the posterior predictive diagnostic and conclude based on the posterior predictive p-values that the sub-model for Sierra Leone is necessarily a better fit for the data. The incidence paths in the posterior predictive distributions of ODE models are still conditioned on the data, whereas the paths in the full posterior predictive distributions of LNA models are not. The partial posterior predictive distributions for the LNA models clearly seem to match the data better than the posterior predictive distributions under the ODE, even for Sierra Leone.

Full and partial posterior predictive distributions for single country LNA Ebola models

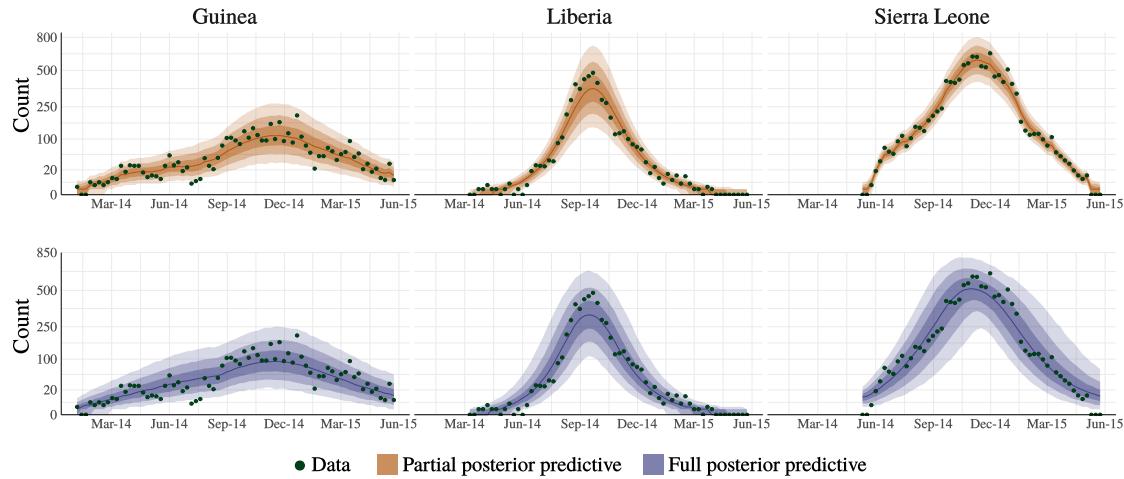


Figure 4.16: Partial and full posterior predictive distributions for country-specific SEIR LNA models for the West Africa Ebola outbreak. Green points correspond to the observed incidence. The shaded bands, in order of lightest to darkest, correspond to pointwise posterior 95%, 80%, and 50% credible intervals, with the pointwise posterior median given by the corresponding colored solid line.

Full and partial posterior predictive distributions for a joint West Africa LNA Ebola model

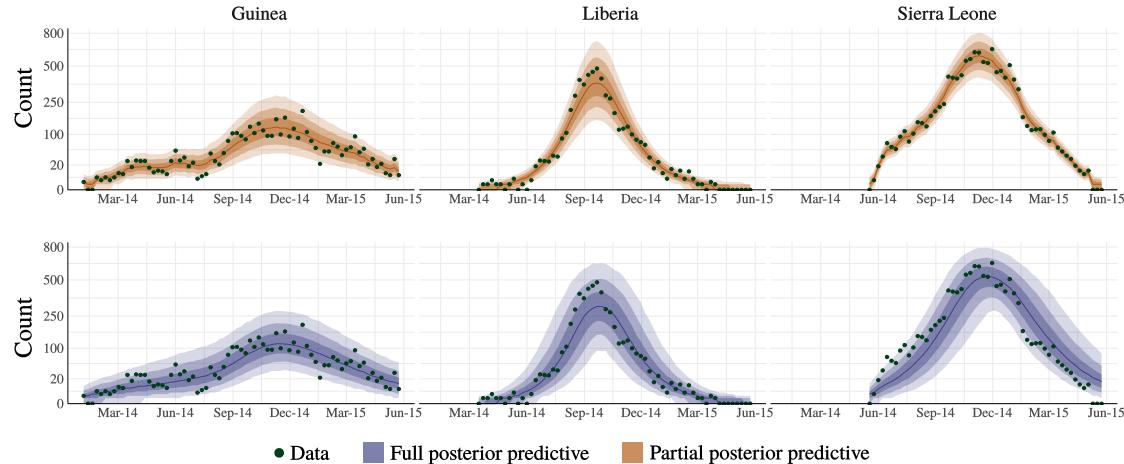


Figure 4.17: Partial and full posterior predictive distributions for a stratified SEIR LNA model for jointly modeling the Ebola outbreak in West Africa. Green points correspond to the observed incidence. The shaded bands, in order of lightest to darkest, correspond to pointwise posterior 95%, 80%, and 50% credible intervals, with the pointwise posterior median given by the corresponding colored solid line.

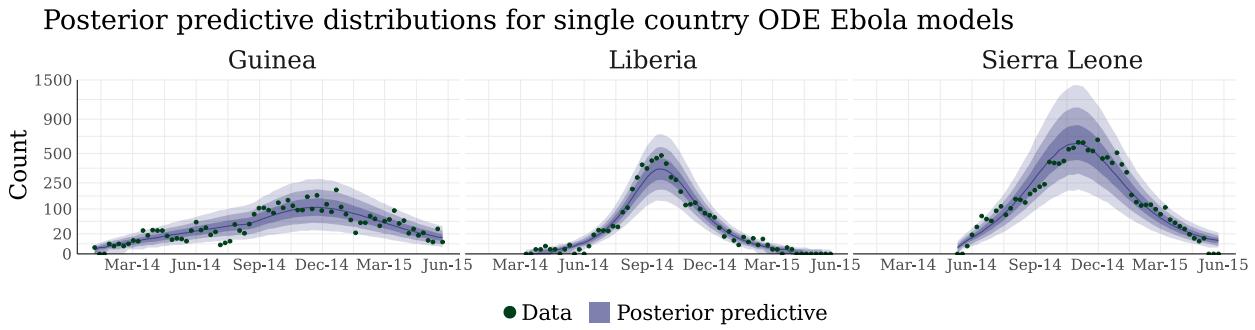


Figure 4.18: Partial and full posterior predictive distributions for country-specific SEIR ODE models for the West Africa Ebola outbreak. Green points correspond to the observed incidence. The shaded bands, in order of lightest to darkest, correspond to pointwise posterior 95%, 80%, and 50% credible intervals, with the pointwise posterior median given by the corresponding colored solid line.

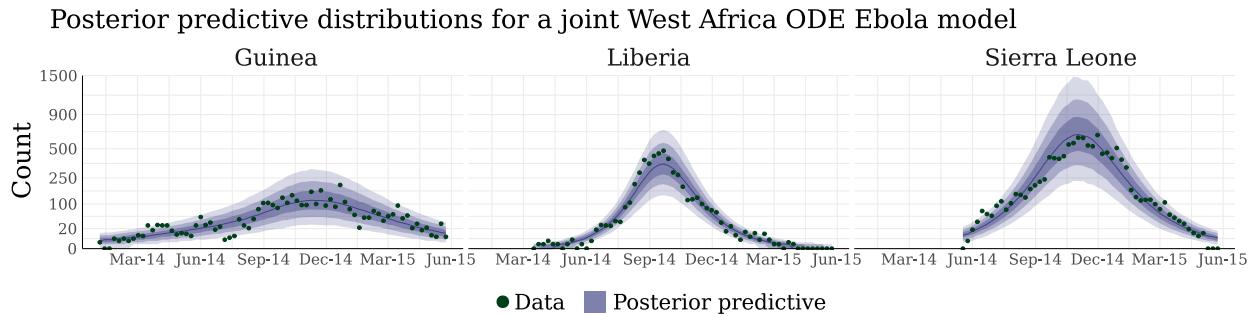


Figure 4.19: Partial and full posterior predictive distributions for a stratified SEIR ODE model for jointly modeling the Ebola outbreak in West Africa. Green points correspond to the observed incidence. The shaded bands, in order of lightest to darkest, correspond to pointwise posterior 95%, 80%, and 50% credible intervals, with the pointwise posterior median given by the corresponding colored solid line.

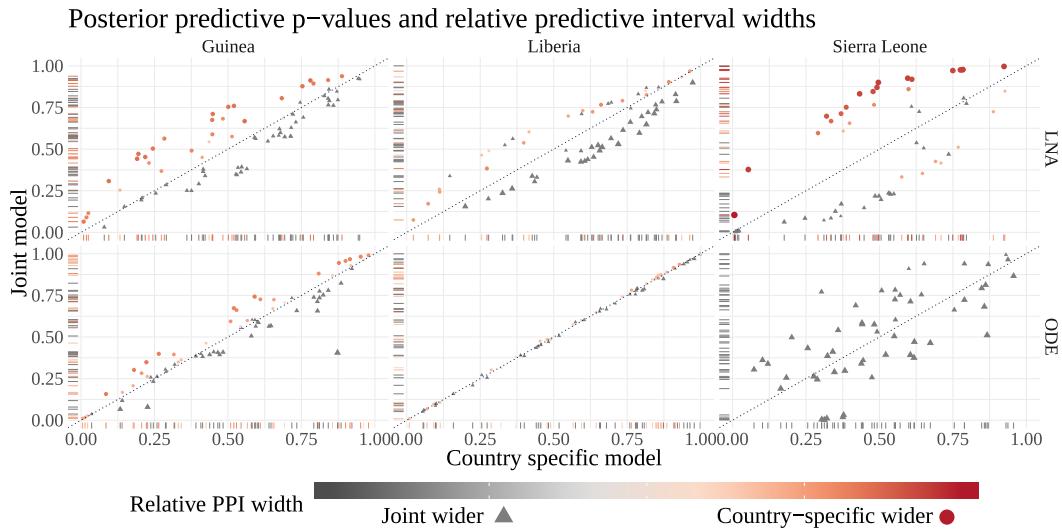


Figure 4.20: Comparison of country-specific and joint SEIR models for Ebola using posterior predictive p-values (PPPs) and relative posterior predictive interval (PPI) widths. The top row compares country-specific and joint models fit via the LNA, while the bottom row compares deterministic country-specific and joint models fit using ODE. Each point corresponds to the observed incidence in a given week. The X-Y coordinates give the PPPs of the data under the country-specific and joint models, respectively. The size and color of each point corresponds to the relative PPI width, computed as $(\hat{\sigma}_{post,\ell}^{country-specific} - \hat{\sigma}_{post,\ell}^{joint})/\hat{\sigma}_{post,\ell}^{joint}$, and the sign of the relative width is further emphasized by the shape of the point. Dots indicate that PPIs for country-specific models are wider, while triangles correspond to observations for which the PPI produced by the joint model was wider.

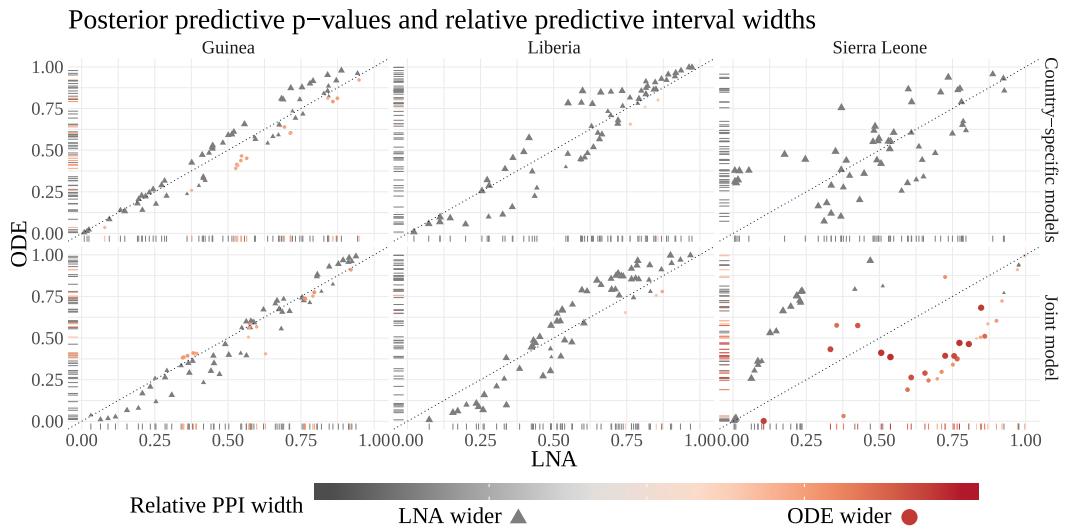


Figure 4.21: Comparison of SEIR models for Ebola fit using the LNA and ODE using posterior predictive p-values (PPPs) and relative posterior predictive interval (PPI) widths. The top row compares LNA and ODE country-specific models, while the bottom row compares LNA and ODE joint. Each point corresponds to the observed incidence in a given week. The X-Y coordinates give the PPPs of the data under LNA and ODE, respectively. The size and color of each point corresponds to the relative PPI width, computed as $(\hat{\sigma}_{post,\ell}^{ode} - \hat{\sigma}_{post,\ell}^{lna})/\hat{\sigma}_{post,\ell}^{lna}$, and the sign of the relative width is further emphasized by the shape of the point. Dots indicate that PPIs under the ODE are wider, while triangles correspond to observations for which the PPI produced by the LNA is wider.

4.5 Discussion

We have presented a framework for efficiently fitting SEMs to partially observed incidence data that is appropriate in large population settings where the time–evolution of the MJP can reasonably be approximated by LNA transition densities. The framework is neither specific to a particular set of model dynamics, nor is it dependent on the particular emission distributions that we used in the examples explored in this work. Therefore, it is possible to incorporate a SEM approximated via the LNA as part of a larger model in which multiple data streams, including incidence data, are synthesized to more precisely estimate the transmission dynamics of an outbreak. One advantage in our approach is that we can obtain approximate estimates of SEM parameters that appropriately account for the stochastic aspects of the MJP without resorting to computationally intensive simulation–based methods. Thus, we are able to fit relatively complex models that are fully stochastic in all aspects of their transmission dynamics, such as the one used to jointly model the 2013–2016 outbreak of Ebola in Guinea, Liberia, and Sierra Leone.

Our main contributions in this work were to demonstrate how a SEM can be reparameterized to admit latent epidemic paths that are compatible with emission distributions for incidence data, and how the LNA approximation can be reparameterized to take advantage of efficient MCMC machinery that can accommodate 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. Moreover, the computational advantage of the LNA allowed us to fit a relatively complex model, that we were unable to fit using MMTL within a pseudo–marginal framework, to data from the Ebola outbreak in West Africa. We also presented interpretable and easily implemented model diagnostics that we were able to lean on for diagnosing model misspecification in our Ebola models.

While this work has addressed various aspects of how to efficiently fit models that can be

used to describe the transmission dynamics of an outbreak, there are also several important issues regarding the application and scientific validity of these models that we have not addressed in depth. Critically, it is frequently, if not universally, unreasonable to assume that the transmission and surveillance dynamics are constant over time. We will address this issue in the next chapter, where we will fit models with time-varying dynamics. Other topics that we have not covered, in part because of the complexity and breadth of the work already presented here, are formal assessments of the predictive performance, issues of model selection, and the effects of model misspecification beyond the use of the LNA to approximate the MJP representation of a SEM.

Chapter 5

MODELS WITH TIME-VARYING DYNAMICS FOR PANDEMIC A(H1N1) INFLUENZA IN FINLAND

5.1 Overview

To this point, we have largely worked with stochastic epidemic models (SEMs) where the transmission dynamics of an outbreak are time-homogeneous. This may be reasonable for short outbreaks in closed, relatively “well-mixed” populations, and is often an attractive modeling choice as SEMs with static dynamics are easier to interpret and fit. Incidence data typically arise in settings where the outbreak milieu changes with environmental factors, spatio-temporal heterogeneity in the contact patterns of subpopulations as they are exposed, or behavioral responses as awareness of an outbreak increases and wanes. Furthermore, we are often interested in understanding the effects of time-varying interventions, such as vaccination campaigns, on the transmission dynamics.

In this chapter, we will demonstrate how the linear noise approximation (LNA) and ordinary differential equation (ODE) SEM representations can be used to fit models with time-varying dynamics. Section 5.2 presents a motivating example that illustrates the importance of allowing for the possibility of heterogeneity in the outbreak dynamics. We then develop in Section 5.3 an age-vaccination stratified SEM with time-varying dynamics for the spread of pandemic A(H1N1) influenza in Finland. We will analyze incidence data, described in Section 1.3.3, from two epidemic seasons. Our goals will be to quantify the transmission dynamics of the outbreak, to estimate the true incidence, and to understand what effect a national vaccination campaign had in mitigating the outbreak severity.

5.2 On the Importance of Allowing for Time–Varying Dynamics

A critical aspect of modeling outbreaks over multiple seasons is that we must account for changes in rates of infectious contacts both within, and between, seasons. The decline in transmission at the end of one season may be attributable, in some combination, to stochastic extinction, a decline in the rate of infectious contact, and a reduction in the effective number of susceptible individuals via immunity acquired from natural exposure or vaccination. It is difficult to explain the emergence of the second season without allowing for stochastic reemergence, changes in the force of infection (FOI), or waning immunity. Put another way, it is highly unlikely an outbreak that died off in a population protected by herd immunity would reemerge absent changes in FOI or repletion of susceptibles in the population.

Before delving into details of how we intend to accommodate time–inhomogeneity in the outbreak dynamics, we briefly highlight why we should bother. To make the point, we compare results for two SIRS models fit to incidence data simulated from an SIRS model where the rate of infectious contact varied sinusoidally over the course of two epidemic waves (depicted in Figure 5.2). While both models allowed for loss of immunity, the per-contact infectious rate was held constant in one model, and in the other was allowed to vary in time, with changes penalized via a first order Gaussian Markov random field (GMRF) shrinkage prior (details presented in Sections 5.3 and C.1). We used the LNA framework developed in Chapter 4 to fit both models.

Although both models are misspecified vis–a–vis the data generating model, the model with time–varying dynamics is clearly better able to describe the dynamics of the outbreak (Figures 5.1 and 5.2). The model with time–homogeneous dynamics fails to recover the static parameters governing the outbreak dynamics and sampling process, whereas the model with time varying dynamics is able to recover the true parameter values used to simulate the data. In particular, it is troubling that the time–homogeneous model is unable to recover the mean duration of immunity and the case detection rate. The model allowing for time–varying dynamics also does a much better job at estimating the basic and effective reproduction

numbers, and is also able to recover the true incidence. In contrast, the model with homogeneous dynamics fails to accurately estimate the basic and effective reproduction numbers between seasons and during the second wave. It also completely fails to capture the incidence during the second wave. Finally, the posterior predictive distributions for the model with time-varying dynamics are more accurate, and more precise, than posterior predictive distributions for the model with homogeneous dynamics (Figures 5.2 and 5.3).

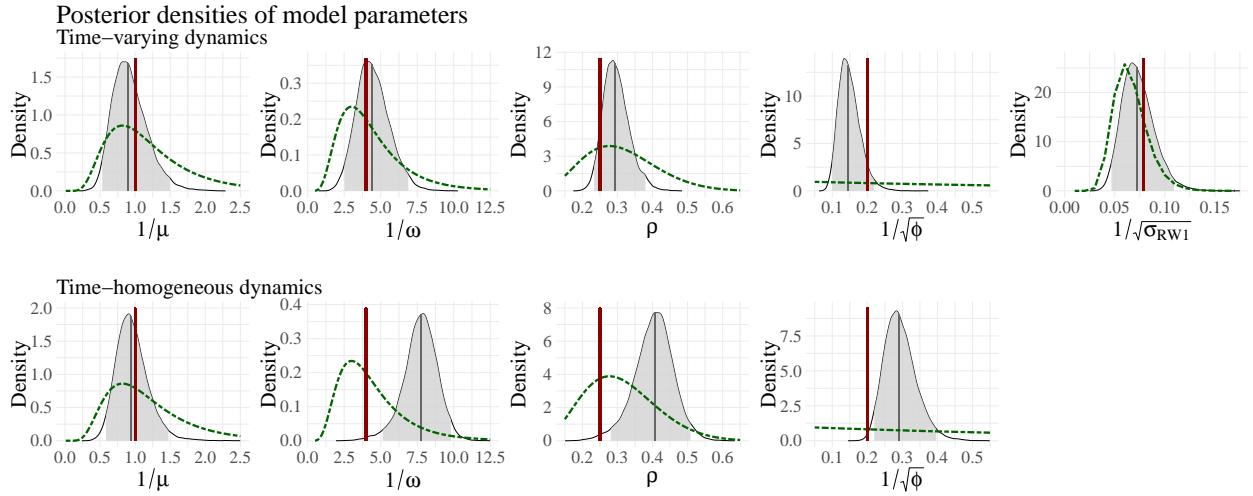


Figure 5.1: Posterior distributions of SIRS model parameters fit to data from an outbreak with time-varying dynamics. From left to right: $1/\mu$, the mean infectious period duration; $1/\omega$, the mean duration of immunity; ρ , the mean case detection rate; ϕ , negative binomial overdispersion parameter; σ_{GMRF} , standard deviation of log-differences of time-varying basic reproduction numbers. True values are given by solid red lines and priors by dashed green curves. Solid grey lines are posterior medians, and shaded regions correspond to 95% Bayesian credible intervals.

Basic and effective reproduction numbers, posterior incidence, partial, and full posterior predictive distributions

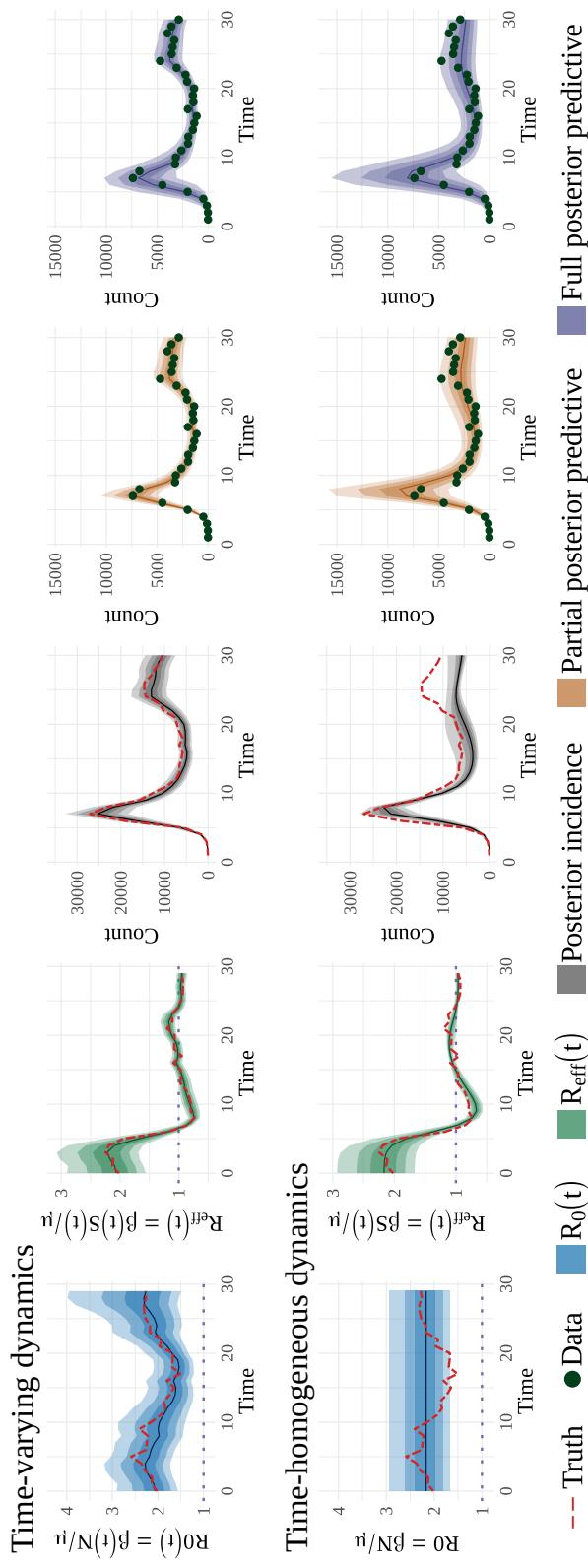


Figure 5.2: Posterior estimates of time-varying quantities. From left to right: basic reproduction numbers, effective reproduction numbers, latent incidence, partial and full posterior predictive distributions. The top row corresponds to estimates obtained using an SIRS model where the basic reproduction number, and hence the per-contact infectivity rate, was allowed to vary in time, with differences penalized according to a first order GMRF. The second row shows estimates obtained from an SIRS model where the per-contact infectivity rate was constant over time. Shaded bands correspond to pointwise 50%, 80%, and 95% Bayesian credible intervals and posterior predictive intervals, with the pointwise posterior/predictive median drawn as a solid line.

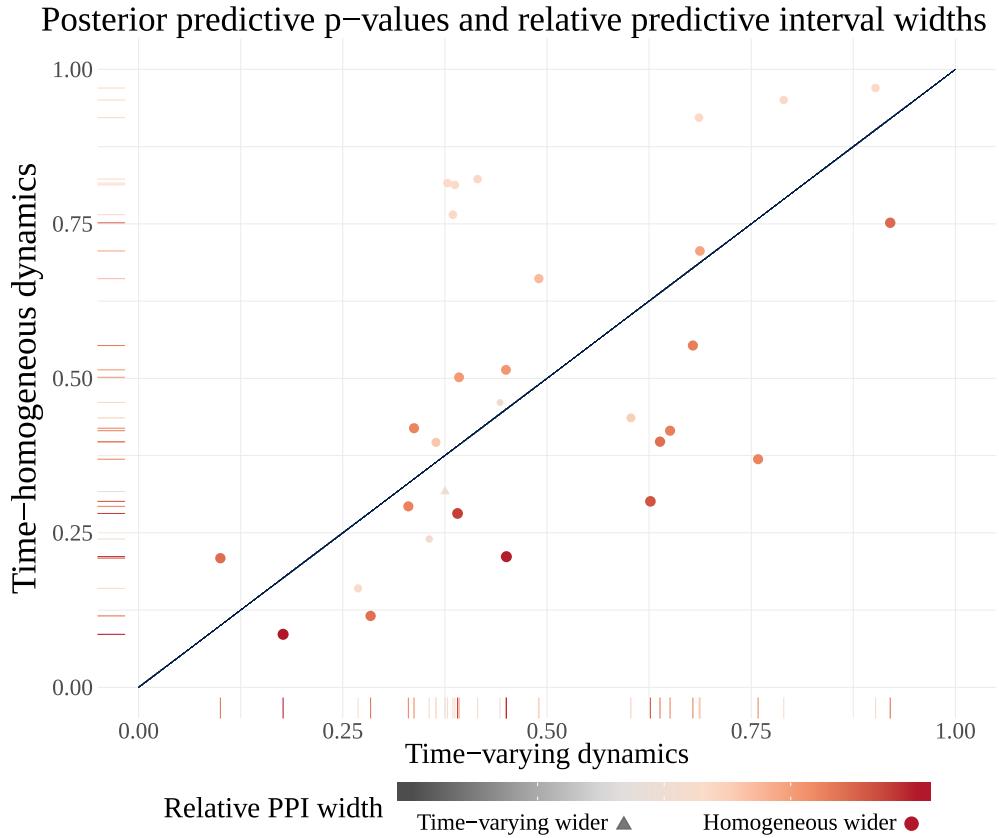


Figure 5.3: Comparison of models with time–varying and constant force of infection using posterior predictive p-values (PPPs) and relative posterior predictive interval (PPI) widths. Each point corresponds to the observed incidence in a given week. The X–Y coordinates give the PPPs under time–varying dynamics and time homogeneous dynamics, respectively. The size and color of each point corresponds to the relative PPI width, computed as $(\hat{\sigma}_{post,\ell}^{constant} - \hat{\sigma}_{post,\ell}^{GMRF})/\hat{\sigma}_{post,\ell}^{GMRF}$, and the sign of the relative width is further emphasized by the shape of the point. Dots indicate that PPIs with constant FOI are wider, the lone triangle corresponds to the one data point where the PPI for the model with time–varying FOI was wider.

5.3 Modeling the Spread of A(H1N1)pdm09 in Finland

5.3.1 Data and Vaccination

We will model the time series of weekly incidence among youths (Y), ages 0–19, and adults (A), ages 20+, over a one year period beginning in epiweek 15, 2009, one month prior to the first observed case in the first season, and a 42 week period beginning in epiweek 33, 2010, corresponding to the start of the 2010–2011 Finnish school year [1]. Data from the inter–season period, epiweeks 15–32, 2010, were aggregated over the inter–season period and accrued at epiweek 33, 2010. We denote the data as,

$$\mathbf{Y} = \left((Y_{Y,1}, Y_{A,1}), \dots, (Y_{Y,52}, Y_{A,52}), (Y'_{Y,71}, Y'_{A,71}), \dots, (Y_{Y,113}, Y_{A,113}) \right),$$

where the index corresponds to weeks elapsed from week zero, i.e., epiweek 15, 2009. The observed incidence in age–stratum j at week ℓ is modeled as a negative binomial sample of the latent incidence

$$Y_{j,\ell} \sim \text{Neg.Binom} \left(\mu_{j,\ell} = \rho_j \left(\Delta N_{SI,j}^{(u)}(t_\ell) + \Delta N_{SI,j}^{(v)}(t_\ell) \right), \sigma^2 = \mu_{j,\ell} + \mu_{j,\ell}^2/\phi_j \right), \quad (5.1)$$

where $N_{SI,j}^{(k)}(t_\ell)$ is the cumulative incidence in age stratum $j \in \{Y, A\}$ among individuals with vaccination status $k \in \{U, V\}$, ρ_j is the age–specific mean case detection rate, ϕ_j is an age–specific overdispersion parameter, and Δ is a difference operator for the cumulative incidence in an age–vaccination stratum, e.g., $\Delta N_{SI,Y}^{(u)}(t_\ell) = N_{SI,Y}^{(u)}(t_\ell) - N_{SI,Y}^{(u)}(t_{\ell-1})$ is the change in cumulative incidence among unvaccinated youths between times $t_{\ell-1}$ and t_ℓ . Three cases, two among adults and one among youths, were detected over the inter–season period and were treated as if they were accrued at epiweek 33, 2010.

5.3.2 Model Structure

We fit an age–vaccination stratified susceptible–infected–recovered–susceptible (SIRS) model in which individuals transitioned stochastically, and continuously in time, between disease states. Individuals were assumed to become infectious immediately upon becoming infected,

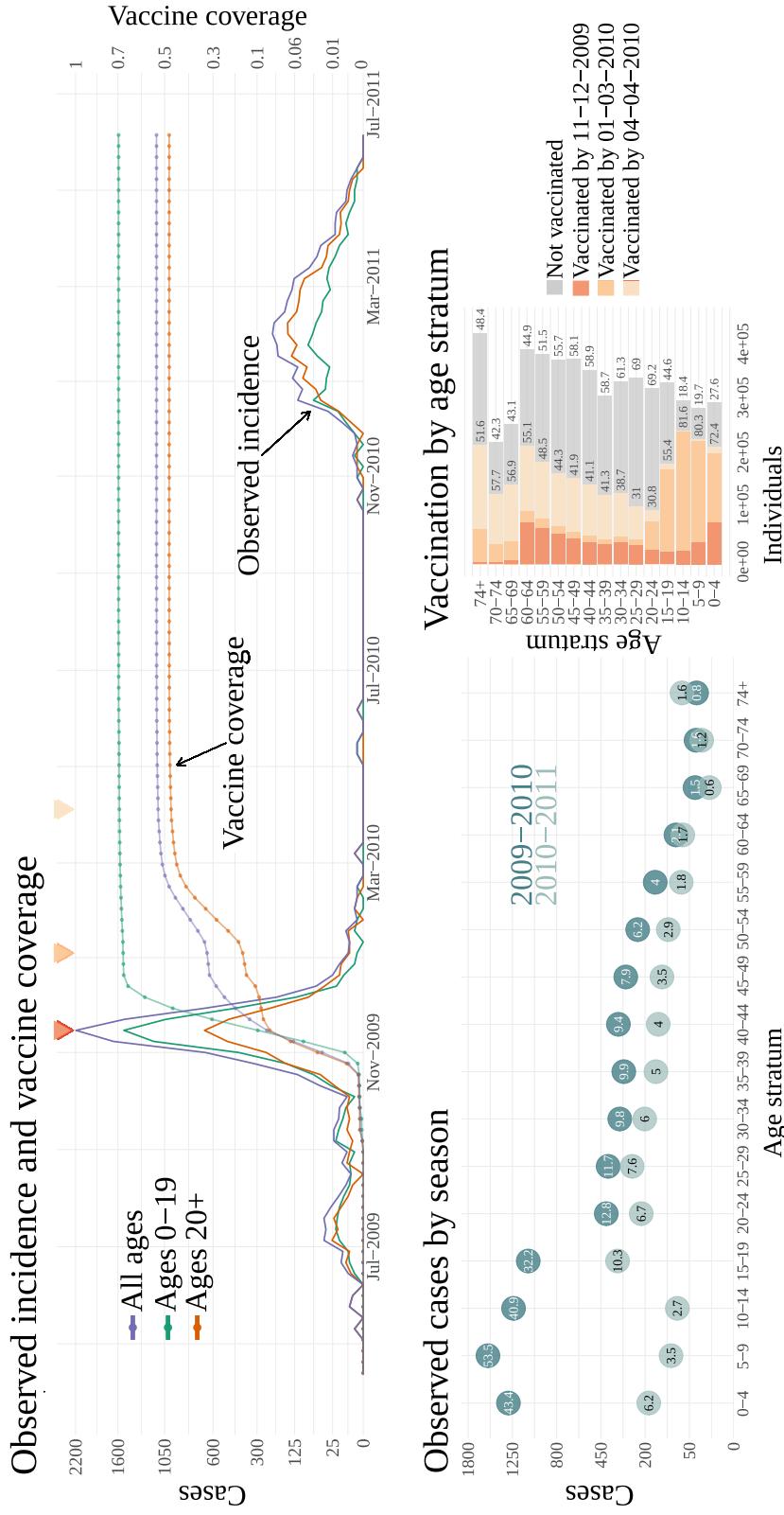


Figure 5.4: Influenza data summaries. *Top:* Observed incidence (solid lines) and vaccine coverage (lines with points). *Bottom left:* Observed cases by season and age stratum. The 2009–2010 season (dark green) corresponds to the period from April 15, 2009 through April 4, 2010. The 2010–2011 season (light green) corresponds to the period from September 12, 2010 through June 5, 2011. Numbers in points give the attack rate within each stratum for the corresponding season. *Bottom right:* Vaccination coverage by age stratum, colored by vaccine coverages at the times of peak incidence in the first season, tail of the major outbreak in the first season, and end of the vaccination campaign. The numbers inside and outside the histograms denote the percentage of individuals in each stratum that were vaccinated and unvaccinated, respectively, by the end of the vaccination campaign. The times at which vaccine coverages are summarized, denoted by colors of histogram bars, are also identified by corresponding triangles above the top figure.

and acquire temporary, though potentially long lasting, protection upon recovery. Individuals who lost immunity were assumed to become fully susceptible. We estimated the initial number of susceptibles as a parameter in the model and assumed that individuals who were not initially susceptible were detached (denoted, D) from the transmission processes but could still be vaccinated. The implications of estimating the initial numbers of susceptibles are discussed in Section C.5.4. Following [194], we assume a closed population and ignore demographic changes or mortality. The model is diagrammed in Figure 5.5, and Table 5.1 lists the model parameters and their interpretations.

We take sojourn time in each disease state to be exponentially distributed and leverage the exchangeability of individuals within the model to represent the time–evolution of the epidemic as a Markov jump process (MJP),

$$\mathbf{X}^c = \left\{ S_Y^{(u)}, I_Y^{(u)}, R_Y^{(u)}, D_Y^{(u)}, S_Y^{(v)}, I_Y^{(v)}, R_Y^{(v)}, D_Y^{(v)}, S_A^{(u)}, I_A^{(u)}, R_A^{(u)}, D_A^{(u)}, S_A^{(v)}, I_A^{(v)}, R_A^{(v)}, D_A^{(v)} \right\},$$

where $S_j^{(k)}$, $I_j^{(k)}$, $R_j^{(k)}$, $D_j^{(k)}$ are the numbers of susceptible, infected, recovered, and detached (susceptible, but not part of the transmission process) individuals in age stratum j with vaccination status k . The state space of \mathbf{X}^c , is the set of compartment counts, \mathcal{S}_X^c . \mathbf{X}^c is also coupled to the cumulative incidence process \mathbf{N}^c on the state space of cumulative incidence counts \mathcal{S}_N^c , as described in Section 4.2.4.

The contact rate between individuals in age stratum j and age stratum k , denoted C_{jk} , was based on estimated mean contact rates, appropriately standardized (see Section C.2), from the Finnish arm of the POLYMOD survey [156, 157] and accessed via the **socialmixr** R package [79]. Sixty percent of contacts in the 0–19 age group were from other individuals age 0–19, while eighty-five percent of the contacts in the 20+ age group were from other adults. The rates, $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_Y, \boldsymbol{\lambda}_A)$, at which individuals in age–stratum $j \in \{Y, A\}$ transition between disease states are

Table 5.1: Summary of notation for influenza models.

Parameter	Interpretation	Time-varying
$\alpha_j(t)$	Rate of exogenous infectious contact, age stratum j	Yes
$\beta_j(t)$	Per-contact rate of endogenous infection, age stratum j	Yes
ν	Rel. rate of infectious contact for vaccinated (1-VE for susceptibility)	No
$1/\mu_j$	Mean infectious period duration, age stratum j	No
$1/\omega$	Mean duration of immunity	No
ρ_j	Mean case detection rate, age stratum j	No
ϕ_j	Negative binomial overdispersion parameter, age stratum j	No
$\mathbf{X}(t)$	Compartment counts at time t	Yes
$\mathbf{N}(t)$	Cumulative incidence by time t	Yes
Variable	Interpretation	Time-varying
$Y_{j,\ell}$	Observed incidence, age stratum j , time t_ℓ	Yes
\mathcal{T}	Observation times, numbered by week: $\{t_\ell : \ell = 1, \dots, 52, 71, \dots, 113\}$	No
N	Population size	No
$N_j^{(k)}(t_\ell)$	Size of age stratum $j \in \{Y, A\}$ with vaccination status $k \in \{U, V\}$ in week ℓ	Yes
s_j	Initially susceptible fraction of age stratum j , $s_j = S_j^{(u)}(t_0)/N_j^{(u)}(t_0)$	No
$V_j(t_\ell)$	# vaccine doses to individuals in age stratum j in week ℓ	Yes
$P_j^v(t_\ell)$	Vaccination coverage in age stratum j by week ℓ	Yes
$\xi_{k,j}^{(uv)}(t_\ell)$	Vaccination forcing for compartment k in age stratum j in week ℓ	Yes
C_{jk}	Relative contact rate to age stratum j from stratum k	No
$R_0(t)$	Basic reproduction number at time t	Yes
$R_{eff}(t)$	Effective reproduction number at time t	Yes
$T_{0,j}(t)$	Basic type reproduction number at time t , stratum j	Yes
$T_{eff,j}(t)$	Effective type reproduction number at time t , stratum j	Yes
$\psi_j(t)$	Intrinsic reproduction number at time t , stratum j	Yes

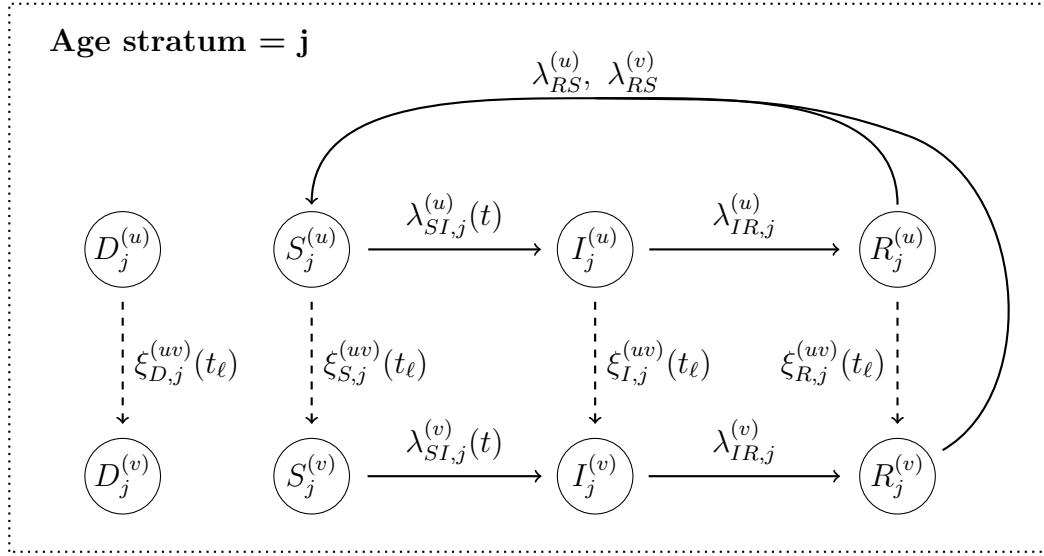


Figure 5.5: Diagram of state transitions for an age–vaccination stratified SIRS model for influenza in Finland. Both age strata, 0–19 and 20+, have the same compartmental structure. Nodes in circles denote model compartments, which are subscripted with age stratum and superscripted with vaccination status. Solid lines indicate stochastic transitions that occur continuously in time. Rates of disease state transitions, denoted by λ , are subscripted with the states from which, and to which, individuals flow and by the age stratum. Superscripts for transition rates indicate vaccination status. Dashed lines represent deterministic forcings from unvaccinated to vaccinated compartments that occur at discrete times. The mass of the forcing at time t , denoted $\xi(t)$, is subscripted by the compartment and age stratum, and superscripted by the direction of the forcing (unvaccinated to vaccinated). Individuals in the detached compartments of age stratum j , denoted $D_j^{(u)}$ and $D_j^{(v)}$, are detached from the transmission process but may still be vaccinated.

$$\boldsymbol{\lambda}_j = \begin{cases} \lambda_{SI,j}^{(u)}(t) = \left[\alpha_j(t) + \beta_j(t) \left(C_{jk} \left(I_j^{(u)} + I_j^{(v)} \right) + (1 - C_{jk}) \left(I_k^{(u)} + I_k^{(v)} \right) \right) \right] S_j^u \\ \lambda_{SI,j}^{(v)}(t) = \nu \left[\alpha_j(t) + \beta_j(t) \left(C_{jk} \left(I_j^{(u)} + I_j^{(v)} \right) + (1 - C_{jk}) \left(I_k^{(u)} + I_k^{(v)} \right) \right) \right] S_j^v \\ \lambda_{IR,j}^{(u)} = \mu_j I_j^{(u)} \\ \lambda_{IR,j}^{(v)} = \mu_j I_j^{(v)} \\ \lambda_{RS,j}^{(u)} = \omega R_j^{(u)} \\ \lambda_{RS,j}^{(v)} = \omega R_j^{(v)} \end{cases} \quad (5.2)$$

At weekly intervals, we deterministically force individuals from unvaccinated model compartments to the corresponding vaccinated compartments. Following [194], we assume that vaccine doses were distributed proportionally to the number of individuals in each unvaccinated compartment of the corresponding age stratum. In [194], it was assumed vaccination resulted in either full protection with probability 0.8, or no change in susceptibility with probability 0.2, two weeks after vaccine administration. Here, we will assume that the vaccine is partially protective in that it reduces the rate at which vaccinated individuals become infected. We do not assume that VE for susceptibility varies by age, and also do not model vaccine efficacy for infectiousness or recovery due to the limited extent of the data. To account for the time required to elicit an immune response, we apply the deterministic forcing at the beginning of the week after the time at which each dose count was indexed.

Algorithmically, the forcing of individuals from unvaccinated compartments to the corresponding vaccinated compartments is applied after sampling the latent path at the end of each time interval. So, suppose that at the end of week ℓ there are

$$\mathbf{X}_j^{(u)}(t_\ell) = \left(S_j^{(u)}(t_\ell), I_j^{(u)}(t_\ell), R_j^{(u)}(t_\ell), D_j^{(u)}(t_\ell) \right)$$

unvaccinated susceptible, infected, recovered, and detached individuals in age stratum j and

$$\mathbf{X}_j^{(v)}(t_\ell) = \left(S_j^{(v)}(t_\ell), I_j^{(v)}(t_\ell), R_j^{(v)}(t_\ell), D_j^{(v)}(t_\ell) \right)$$

vaccinated individuals, and that $V_j(t_\ell)$ vaccine doses were recorded for stratum j in that

week. The vaccination count will then modify the initial state for the following week,

$$\begin{aligned}\mathbf{X}_j^{(u)}(t_\ell^+) &= \mathbf{X}_j^{(u)}(t_\ell) - V_j(t_\ell) \frac{\mathbf{X}_j^{(u)}(t_\ell)}{N_j^{(u)}(t_\ell)}, \\ \mathbf{X}_j^{(v)}(t_\ell^+) &= \mathbf{X}_j^{(v)}(t_\ell) + V_j(t_\ell) \frac{\mathbf{X}_j^{(u)}(t_\ell)}{N_j^{(u)}(t_\ell)}.\end{aligned}\tag{5.3}$$

The procedure for mapping LNA draws onto an LNA path using the non-centered parameterization with vaccination forcings is given in Algorithm 11 (Section C.3). ODE paths are obtained using a similar procedure by omitting the stochastic terms in the LNA.

Reproduction numbers

The basic reproduction number, R_0 , and its variants describe the propensity of an outbreak to spread through a population [102, 209]. R_0 is loosely interpreted as the average number of secondary infections arising from a single infected individual in a completely susceptible population. An outbreak will fail to sustain itself (almost surely when it evolves deterministically, but in reality with high probability due to the inherently stochastic nature of epidemics) when $R_0 < 1$. For this reason, epidemiologists often quantify the effectiveness of a control measure by estimating the extent to which the intervention reduces R_0 . In settings, such as this, where the force of infection also includes pressure from outside the density dependent transmission process, e.g., due to importation of infection from outside the population, we can interpret the basic reproduction number as an intrinsic reproduction number [24]. From a modeling perspective, it is important to understand not only how R_0 affects the transmission dynamics, but also how various aspects of the model interact to affect R_0 .

The basic reproduction number of a structured SEM can be obtained by computing the spectral radius (absolute value of the dominant eigenvalue) of the next generation matrix (NGM) for the linearized system of ODEs specifying the flow in and out of states at infection [102, 208, 209]. Element (i, j) of the NGM gives the expected number of secondary infections in age stratum i given an index infection in stratum j . Noting that vaccinated and unvacci-

nated individuals in this model are equally transmissive and have the same mean infectious period durations, we obtain the age-structured next generation matrix

$$\mathbf{K} = \begin{pmatrix} K_{YY} & K_{YA} \\ K_{AY} & K_{AA} \end{pmatrix} = \begin{pmatrix} \frac{C_{YY}\beta_Y(N_Y^{(u)} + \nu_Y N_Y^{(v)})}{\mu_Y} & \frac{C_{YA}\beta_Y(N_Y^{(u)} + \nu_Y N_Y^{(v)})}{\mu_A} \\ \frac{C_{AY}\beta_A(N_A^{(u)} + \nu_A N_A^{(v)})}{\mu_Y} & \frac{C_{AA}\beta_A(N_A^{(u)} + \nu_A N_A^{(v)})}{\mu_A} \end{pmatrix}. \quad (5.4)$$

The basic reproduction number is

$$R_0 = \frac{1}{2}(K_{YY} + K_{AA}) + \frac{1}{2}\sqrt{(K_{YY} - K_{AA})^2 + 4K_{YA}K_{AY}}. \quad (5.5)$$

Note that it is possible for the outbreak to persist, i.e., $R_0 > 1$, when the intrinsic reproduction number in one of the strata is below 1, e.g., $K_{YY} < 1$, even though most transmission arises from within-group contacts. However, $R_0 <$ will be less than 1 if both $K_{YY} < 1$ and $K_{AA} < 1$.

When assigning priors for the intrinsic reproduction numbers, it is important to consider that K_{YY} K_{AA} interact in a somewhat complicated way to induce a prior over the basic reproduction number. R_0 is greater than the average of intrinsic reproduction numbers, K_{YY} and K_{AA} , since the second term on the right hand side of (5.5) will typically be positive. Furthermore, R_0 does not exactly behave like the average of intrinsic reproduction numbers since the cross-stratum contact rates were negligible ($C_{AA} = 0.6$, $C_{YY} = 0.85$). Rather, R_0 is proportional to the average of K_{YY} and K_{AA} , plus a non-linear function of K_{YY} and K_{AA} .

5.3.3 Flexible Models for the Force of Infection with Gaussian Markov Random Fields

We will model the time-varying force of infection using first order GMRFs for the log intrinsic reproduction numbers and the effective rate of exogenous infectious contacts in each age stratum. The GMRFs are separately specified over three epochs corresponding to the period one month before the first detected case until the start of the Finnish school year in 2009, the beginning of the Finnish school year in 2009 until its start in 2010 (including the inter-season period), and the second epidemic season beginning at the start of the Finnish school year, 2010. These epochs correspond to epiweeks 15, 2009 through epiweek 34, 2009, epiweek

35, 2009 through epiweek 32, 2010, and epiweek 33, 2010 through epiweek 22, 2011. We will assume that the intrinsic reproduction numbers are homogeneous over the inter-season period, from epiweek 16, 2010 – epiweek 32, 2010.

Let $\psi_j = \{\psi_{j,\ell} : \ell \in 0, \dots, 52, 71, \dots, 112\}$ be the vector of intrinsic reproduction numbers for stratum $j \in \{Y, A\}$, ignoring the relative contact rates, C_{jk} , and vaccination. The reproduction number in time interval ℓ , indexed in weeks from epiweek 15, 2009, is

$$\psi_{j,\ell} = \frac{\beta_j(t_\ell)N_j s_j}{\mu_j}, \quad (5.6)$$

where $\beta_j(t_\ell)$ is the per-contact infection rate, μ_j is the recovery rate, N_j is the size of the age stratum, and s_j is the fraction of individuals in stratum j that are susceptible.

When transmission is sustained at near-endemic levels, we should expect that the effective reproduction numbers are near one, or perhaps slightly above. Let ν be the relative rate of infectious contact for vaccinated vis-a-vis unvaccinated individuals, i.e., one minus the vaccine efficacy for susceptibility. The effective reproduction number, which accounts for the effects of vaccination and depletion of susceptibles, is

$$\psi_{j,\ell}^{eff} = \frac{\beta_j(t_\ell)_j \left(S_j^{(u)}(t_\ell) + S_j^{(v)}(t_\ell)\nu \right)}{\mu_j}. \quad (5.7)$$

Assuming that vaccination does not increase the rate of infectious contact (and we will assume it does not), then $\psi_{j,\ell}^{eff} < \psi_{j,\ell}$ since $\nu \in [0, 1]$ and $S_j^{(u)}(t_\ell) + S_j^{(v)}(t_\ell) < N_j s_j$. We must account for this when specifying priors for the basic intrinsic reproduction numbers at the start of each epoch. It would not be appropriate to assign a prior for the intrinsic basic reproduction number at the beginning of the third epoch that nominally concentrates mass near one since this would imply that the effective reproduction numbers are far below one due to vaccination and depletion of susceptibles in the first wave of the outbreak. Our approach will be to assign priors to the vaccination adjusted intrinsic reproduction numbers at the beginning of each epoch, $\psi'_{j,\ell} = \psi_{j,\ell} (1 - P_j^v(t_\ell) + \nu P_j^v(t_\ell))$, where $P_j^v(t_\ell)$ is the vaccination coverage in stratum j at t_ℓ , $\ell = 0, 19, 71$.

We denote by Δ the first-order forward difference operator, i.e., $\Delta X_j = X_{j+1} - X_j$. We will penalize the magnitude of week-to-week changes in the force of infection by assigning

a Gaussian prior with mean zero and variance $\sigma_{j,\ell}^2$ to the first-order differences in the log intrinsic reproduction numbers for each age stratum in each epoch. The sub-model for intrinsic reproduction numbers in stratum j is

$$\begin{aligned}
\log(\psi_{j,\ell}) &\sim \mathcal{N}(\mu_{j,\ell}, \tau_{j,\ell}^2), & \ell = 0, 19, 71, \\
\log(\sigma_{j,\ell}) &\sim \mathcal{N}(m_{j,\ell}, s_{j,\ell}^2), & \ell = 0, 19, 71, \\
\Delta \log(\psi_{j,\ell}) &\sim \mathcal{N}(0, \sigma_{j,0}^2), & \ell = 0, \dots, 17, \\
\Delta \log(\psi_{j,\ell}) &\sim \mathcal{N}(0, \sigma_{j,19}^2), & \ell = 19, \dots, 51, \\
\Delta \log(\psi_{j,\ell}) &\sim \mathcal{N}(0, \sigma_{j,71}^2), & \ell = 71, \dots, 111, \\
\log(\psi_{j,\ell}) &= \log(\psi_{j,0}) + \sum_{k=0}^{\ell-1} \Delta \log(\psi_{j,k}), & \ell = 1, \dots, 18, \\
\log(\psi_{j,\ell}) &= \log(\psi_{j,19}) + \sum_{k=0}^{\ell-1} \Delta \log(\psi_{j,19+k}), & \ell = 20, \dots, 52, \\
\log(\psi_{j,\ell}) &= \log(\psi_{j,71}) + \sum_{k=0}^{\ell-1} \Delta \log(\psi_{j,71+k}), & \ell = 72, \dots, 112. \\
\beta_j(t_\ell) &= \psi_{j,\ell} N_j / \mu_j, & \ell = 0, \dots, 52, 71, \dots, 112.
\end{aligned}$$

The induced GMRF prior for the vaccination adjusted basic reproduction numbers is shown in Figure C.6. In practice, we will use the following partially non-centered parameterization:

$$\begin{aligned}
\log(\psi_{j,\ell}) &\sim \mathcal{N}(\mu_{j,\ell}, \tau_{j,\ell}^2), & \ell &= 0, 19, 71, \\
Z_{\log(\sigma_{j,\ell})} &\sim \mathcal{N}(0, 1), & \ell &= 0, 19, 71, \\
Z_{\Delta \log(\psi_{j,\ell})} &\sim \mathcal{N}(0, 1), & \ell &= 0, \dots, 17, 19, \dots, 51, 71, \dots, 111, \\
\log(\sigma_{j,\ell}) &= m_{j,\ell} + s_{j,\ell} Z_{\log(\sigma_{j,\ell})}, & \ell &= 0, 19, 71, \\
\Delta \log(\psi_{j,\ell}) &= \sigma_{j,0} Z_{\Delta \log(\psi_{j,\ell})}, & \ell &= 0, \dots, 17, \\
\Delta \log(\psi_{j,\ell}) &= \sigma_{j,19} Z_{\Delta \log(\psi_{j,\ell})}, & \ell &= 19, \dots, 51, \\
\Delta \log(\psi_{j,\ell}) &= \sigma_{j,71} Z_{\Delta \log(\psi_{j,\ell})}, & \ell &= 71, \dots, 111, \\
\log(\psi_{j,\ell}) &= \log(\psi_{j,0}) + \sum_{k=0}^{\ell-1} \Delta \log(\psi_{j,k}), & \ell &= 1, \dots, 18, \\
\log(\psi_{j,\ell}) &= \log(\psi_{j,19}) + \sum_{k=0}^{\ell-1} \Delta \log(\psi_{j,19+k}), & \ell &= 20, \dots, 52, \\
\log(\psi_{j,\ell}) &= \log(\psi_{j,71}) + \sum_{k=0}^{\ell-1} \Delta \log(\psi_{j,71+k}), & \ell &= 72, \dots, 112. \\
\beta_j(t_\ell) &= \psi_{j,\ell} N_j / \mu_j, & \ell &= 0, \dots, 52, 71, \dots, 112.
\end{aligned}$$

We similarly assign a first order GMRF prior to the effective number of exogenous infectious contacts for each age stratum (Figure C.7). Note that this term is somewhat of a catch-all for infectious contacts that arise outside of the density dependent mixing of susceptible and infected individuals, and is loosely interpreted as either the baseline rate of infectious contact or as the rate of infection from outside the population. Let $\alpha'_{j,\ell} = \alpha_{j,\ell} N_j s_j$ denote the rate of exogenous infection for age stratum j in week ℓ , scaled by the effective stratum size. The GMRF prior for $\boldsymbol{\alpha}'_j = \{\alpha'_{j,\ell} : \ell = 0, \dots, 52, 71, \dots, 112\}$ is

$$\begin{aligned}
\log(\alpha'_{j,\ell}) &\sim \mathcal{N}(\mu_{j,\ell}, \tau_{j,\ell}^2), & \ell = 0, 19, 71, \\
\log(\sigma_{j,\ell}) &\sim \mathcal{N}(m_{j,\ell}, s_{j,\ell}^2), & \ell = 0, 19, 71, \\
\Delta \log(\alpha'_{j,\ell}) &\sim \mathcal{N}(0, \sigma_{j,0}^2), & \ell = 0, \dots, 17, \\
\Delta \log(\alpha'_{j,\ell}) &\sim \mathcal{N}(0, \sigma_{j,19}^2), & \ell = 19, \dots, 51, \\
\Delta \log(\alpha'_{j,\ell}) &\sim \mathcal{N}(0, \sigma_{j,71}^2), & \ell = 71, \dots, 111, \\
\log(\alpha'_{j,\ell}) &= \log(\alpha'_{j,0}) + \sum_{k=0}^{\ell-1} \Delta \log(\alpha'_{j,k}), & \ell = 1, \dots, 17, \\
\log(\alpha'_{j,\ell}) &= \log(\alpha'_{j,19}) + \sum_{k=0}^{\ell-1} \Delta \log(\alpha'_{j,19+k}), & \ell = 19, \dots, 52, \\
\log(\alpha'_{j,\ell}) &= \log(\alpha'_{j,71}) + \sum_{k=0}^{\ell-1} \Delta \log(\alpha'_{j,71+k}), & \ell = 72, \dots, 112. \\
\alpha_{j,\ell} &= \alpha'_{j,\ell}/(N_j s_j), & \ell = 0, \dots, 52, 71, \dots, 112.
\end{aligned}$$

Again, we will use a partially non-centered parameterization in practice.

$$\begin{aligned}
\log(\alpha'_{j,\ell}) &\sim \mathcal{N}(\mu_{j,\ell}, \tau_{j,\ell}^2), & \ell = 0, 19, 71, \\
Z_{\log(\sigma_{j,\ell})} &\sim \mathcal{N}(0, 1), & \ell = 0, 19, 71, \\
Z_{\Delta \log(\alpha'_{j,\ell})} &\sim \mathcal{N}(0, 1), & \ell = 0, \dots, 17, 19, \dots, 51, 71, \dots, 111, \\
\log(\sigma_{j,\ell}) &= m_{j,\ell} + s_{j,\ell} Z_{\log(\sigma_{j,\ell})}, & \ell = 0, 19, 71, \\
\Delta \log(\alpha'_{j,\ell}) &= \sigma_{j,0} Z_{\Delta \log(\alpha'_{j,\ell})}, & \ell = 0, \dots, 17, \\
\Delta \log(\alpha'_{j,\ell}) &= \sigma_{j,19} Z_{\Delta \log(\alpha'_{j,\ell})}, & \ell = 19, \dots, 51, \\
\Delta \log(\alpha'_{j,\ell}) &= \sigma_{j,71} Z_{\Delta \log(\alpha'_{j,\ell})}, & \ell = 71, \dots, 111, \\
\log(\alpha'_{j,\ell}) &= \log(\alpha'_{j,0}) + \sum_{k=0}^{\ell-1} \Delta \log(\alpha'_{j,k}), & \ell = 1, \dots, 17, \\
\log(\alpha'_{j,\ell}) &= \log(\alpha'_{j,19}) + \sum_{k=0}^{\ell-1} \Delta \log(\alpha'_{j,19+k}), & \ell = 19, \dots, 52, \\
\log(\alpha'_{j,\ell}) &= \log(\alpha'_{j,71}) + \sum_{k=0}^{\ell-1} \Delta \log(\alpha'_{j,71+k}), & \ell = 72, \dots, 112. \\
\alpha_{j,\ell} &= \alpha'_{j,\ell}/(N_j s_j), & \ell = 0, \dots, 52, 71, \dots, 112
\end{aligned}$$

The initial reproduction numbers, for which we retain the centered parameterization, are blocked with other model parameters and updated using a multivariate normal slice sampler

(MVNSS). The non-centered GMRF differences and their standard deviations are updated using elliptical slice sampling (Algorithm 6). Readers familiar with GMRFs will note the, somewhat unconventional, Gaussian prior for the standard deviation of the GMRF increments. This choice was made for computational reasons in order to facilitate joint updates of the GMRF and its hyper-parameters. Alternating between field and hyper-parameter updates is known to result in poorly mixing MCMC chains as large updates to the hyper-parameters quickly result in fields that are not concordant with the data [126, 158].

5.3.4 Computation and Inference

Due to the complexity of the model, we will fit the model using the deterministic ODE framework that was explored as a comparator for the LNA in Chapter 4. Recall that is the infinite population limit of the MJP representation of the SEM. This choice was made to allow us to more quickly explore different parameterizations of MCMC estimation scales and priors.

Let $\mathbf{Z} = (\mathbf{Z}^F, \mathbf{Z}^{\boldsymbol{\theta}_F})$ denote the vector of i.i.d. standard normal draws for the GMRF increments and GMRF hyperparameters, $\text{doGMRF}(\mathbf{Z}; \boldsymbol{\theta})$ be the operation for computing the GMRF from its draws and hyperparameters, and $\pi(\boldsymbol{\theta})$ be the prior distributions for the other model parameters. Let $\mathcal{T} = \{t_\ell : \ell = 1, \dots, 52, 71, \dots, 113\}$ denote the set of observation times. Our MCMC will target the posterior

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{Z} | \mathbf{Y}) &\propto \pi(\mathbf{Z}) \pi(\boldsymbol{\theta}) \pi(\mathbf{Y} | \boldsymbol{\theta}, \mathbf{Z}) \\ &= \pi(\mathbf{Z}) \pi(\boldsymbol{\theta}) \prod_{t_\ell \in \mathcal{T}} \prod_{j \in \{Y, A\}} \Pr(Y_j(t_\ell) | \text{doGMRF}(\mathbf{Z}), \boldsymbol{\theta}, \mathcal{I}) \end{aligned} \quad (5.8)$$

MCMC proceeds by alternately updating $\mathbf{Z} | \mathbf{Y}, \boldsymbol{\theta}$ using ElliptSS (Algorithm 12) and MVNSS updates for $\boldsymbol{\theta} | \mathbf{Y}, \mathbf{Z}$ (Algorithm 5). Priors for model parameters were informative, as detailed in Section C.5.2, and we made an effort to choose priors that were based on published estimates from other studies. Additional MCMC details are provided in Section C.5.1.

5.4 Results

5.4.1 Incidence

Estimates of cumulative incidence and attack rates by season and age group are reported in Table 5.2. We estimate that there were approximately 532,000 (95% BCI: 393,000, 703,000) infections in the first epidemic season, and an additional 240,000 (95% BCI: 172,000, 333,000) cases during the second season. These estimates would correspond to estimated attack rates of 7.4%–13.1% during the first season, and 3.2%–6.2% in the second season if each individual was infected only once. The estimated attack rates were substantially higher among youths than adults in the first epidemic season, and slightly higher during the second season. The estimated incidence over the inter-season period was low; 110 cases (95% BCI: 50, 250) among youths, and 330 cases (95% BCI: 150, 710) among adults.

5.4.2 Transmission Dynamics and Vaccination

The period just prior to the first detected cases until the start of the Finnish school year in 2009 was characterized by near-endemic effective reproduction numbers and modest rates of exogenous infectious contact (second and fourth rows of Figure 5.6). We estimate that at the start of the Finnish school year in 2009, there were 390 (95% BCI: 230, 670) infected individuals in the population, roughly half of whom were youths. The median estimated effective reproduction numbers in the first season exceeded 1.1 starting at the beginning of October, 2009, and peaked around 1.25 (95% BCI: 1.2, 1.3) at the beginning of November, which corresponded to the start of exponential growth in the number of detected cases. The effective reproduction number first dropped below one in the third week of November, corresponding to the time at which case counts began to decline. The rate of exogenous infectious contact was higher during the epidemic wave compared to the preceding months, though the contribution of exogenous contact to the force of infection was modest.

The second epidemic wave was longer in duration, but less severe, than the first wave. We estimate that there were only a handful of infected individuals at the start of the third

Table 5.2: Estimated infections (thousands) and attack rates by season and age stratum. Attack rates are calculated as the number of infections divided by the size of each stratum, assuming that cases are unique.

Time varying dynamics						
	Season 1		Season 2		Both Seasons	
	Cases ($\times 10^3$)	Attack rate (%)	Cases ($\times 10^3$)	Attack rate (%)	Cases ($\times 10^3$)	Attack rate (%)
Ages 0-19	174 (127, 231)	14.2 (10.4, 18.9)	68.6 (46.6, 100)	5.6 (3.8, 8.2)	244 (181, 321)	19.9 (14.8, 26.2)
Ages 20+	356 (243, 501)	8.6 (5.9, 12.1)	171 (117, 245)	4.1 (2.8, 5.9)	530 (378, 720)	12.8 (9.2, 17.4)
All ages	532 (393, 703)	9.9 (7.4, 13.1)	240 (172, 333)	4.5 (3.2, 6.2)	774 (586, 1,010)	14.5 (11, 18.8)

Piecewise homogeneous dynamics						
	Season 1		Season 2		Both Seasons	
	Cases ($\times 10^3$)	Attack rate (%)	Cases ($\times 10^3$)	Attack rate (%)	Cases ($\times 10^3$)	Attack rate (%)
Ages 0-19	150 (107, 203)	12.3 (8.7, 16.6)	55.1 (34.4, 82)	4.5 (2.8, 6.7)	206 (154, 267)	16.8 (12.6, 21.9)
Ages 20+	263 (182, 375)	6.4 (4.4, 9.1)	181 (128, 242)	4.4 (3.1, 5.9)	445 (324, 598)	10.8 (7.9, 14.5)
All ages	414 (305, 556)	7.7 (5.7, 10.4)	236 (178, 309)	4.4 (3.3, 5.8)	653 (499, 840)	12.2 (9.3, 15.7)

epoch (posterior median, 12; 95% BCI: 5, 140), about half of whom were youths. The median effective reproduction number during the second wave exceeded one at the end of September, 2010, rose above 1.1 in early November, and peaked at 1.16 (95% BCI: 1.1, 1.2) in early December. The effective reproduction number only fell below one in early February, 2011. The rate of exogenous infectious contact was lower during the second epidemic wave than during the first wave.

Figure 5.6 (second and third rows) presents the effective and basic type reproduction numbers for the 0–19 and 20+ age strata, computed as the column sums of the basic and effective NGMs and which are interpreted as a measure of the expected number of secondary cases of any age caused by a single index case in a given age stratum [89]. Estimates of effective type reproduction numbers suggest that transmission was largely sustained by adults transmitting infections to others, rather than by infections that originated among youths. This is a reasonable finding given that the adults made up 77% of the population and that the majority of all contacts were with adults (85% of contacts in the adult population were with other adults, while 40% of youth contacts were with adults). Note that this definition differs from the definition of type reproduction number in [101], which is the expected number of secondary cases of a particular type given an index case of that type.

The mean infectious period of youths was slightly longer than that of adults, 2.5 days (95% BCI: 1.9 days, 3.5 days) compared with 2.1 days (1.8 days, 5.8 days), respectively. Published point estimates of the serial interval for A(H1N1)pdm09, which is the average time between an index case and a secondary case, and which we might expect to be a bit shorter than the mean infectious period, have mostly been in the 2–3.5 day range [214]. Immunity following infection was relatively long-lasting, although a sizable percentage of individuals who were infected lost immunity; 19% (95% BCI: 12%, 28%) of individuals who had become infected before the start of the second season had lost protection by the beginning of the second season, while the percentage of infected individuals who became infected over both seasons and lost immunity by the end of season two was 27% (95% BCI: 18%, 38%). The vaccine was estimated to reduce the rate of infectious contact by 81% (95% BCI: 46%, 96%).

Estimates of time-varying quantities under time-varying dynamics

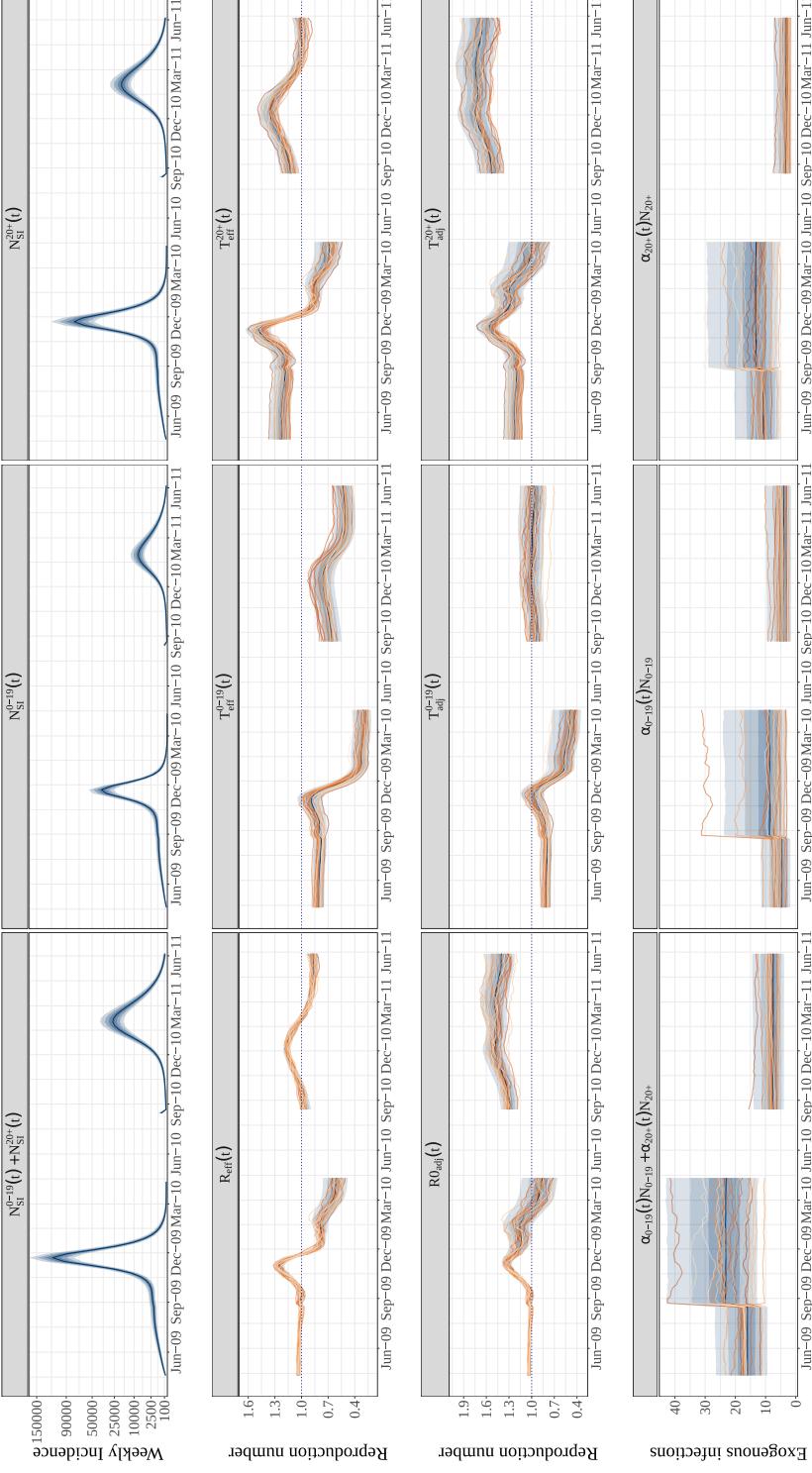


Figure 5.6: Posterior estimates of time-varying quantities under time-varying dynamics. Estimated incidence (top row), effective reproduction numbers (second row), vaccination-adjusted basic reproduction numbers (third row), and exogenous infections (bottom row). $N_S^j(t)$ is the weekly incidence in age stratum j , $R_{eff}(t)$ and $R_{adj}(t)$ are the effective and vaccination-adjusted basic reproduction numbers, $T_{eff}^j(t)$ is the type reproduction number for stratum j (expected # of secondary infections in all strata for a single infected in stratum j), and $\alpha_j(t)$ is the rate of exogenous infectious contacts for stratum j .

Table 5.3: Posterior medians (95% Bayesian credible intervals) of SIRS model parameters for pandemic A(H1N1) influenza in Finland. The corresponding priors are given in Tables C.3 and C.6.

Parameter	Interpretation	Dynamics	
		Time varying	Piecewise homogeneous
$\psi_{Y,0}$	Intrinsic R_0 for youths at epiweek 15, 2009	1.0 (0.99, 1.1)	1.0 (0.95, 1.2)
$\psi_{Y,19}$	Intrinsic R_0 for youths at epiweek 35, 2009	1.1 (1.0, 1.1)	1.5 (1.1, 1.8)
$\psi_{Y,71}^{adj}$	Vaccination adjusted intrinsic R_0 for youths at epiweek 33, 2010	1.2 (1.0, 1.3)	1.2 (0.95, 1.5)
$\psi_{A,0}$	Intrinsic R_0 for adults at epiweek 15, 2009	1.0 (1.0, 1.1)	1.0 (0.94, 1.1)
$\psi_{A,19}$	Intrinsic R_0 for adults at epiweek 35, 2009	1.0 (0.98, 1.1)	1.1 (1.0, 1.2)
$\psi_{A,71}^{adj}$	Vaccination adjusted intrinsic R_0 for adults at epiweek 33, 2010	1.3 (1.2, 1.5)	1.4 (1.3, 1.6)
$1/\mu_Y$	Mean infectious period for youths (days)	2.5 (1.9, 3.5)	5.5 (3.5, 7.5)
$1/\mu_A$	Mean infectious period for adults (days)	2.1 (1.8, 2.4)	2.6 (2.2, 3.1)
$1/\omega$	Mean duration of immunity (years)	3.6 (2.4, 5.8)	4.5 (2.8, 7.3)
ν	1 - VE for susceptibility	0.19 (0.04, 0.54)	0.05 (0.01, 0.19)
s_Y	% susceptible youths, $S_Y^{(u)}(t_0)/N_Y$	0.28 (0.21, 0.36)	0.34 (0.26, 0.42)
s_A	% susceptible adults, $S_A^{(v)}(t_0)/N_A$	0.28 (0.21, 0.35)	0.32 (0.24, 0.40)
ρ_Y	Mean case detection rate for youths	0.013 (0.01, 0.02)	0.015 (0.01, 0.02)
ρ_A	Mean case detection rate for adults	0.008 (0.006, 0.01)	0.0008 (0.006, 0.01)
$1/\sqrt{\phi_Y}$	Negative binomial overdispersion for youths	0.71 (0.55, 0.92)	1.1 (0.86, 1.3)
$1/\sqrt{\phi_A}$	Negative binomial overdispersion for adults	0.56 (0.37, 0.76)	1 (0.82, 1.20)

The effective reproduction numbers in the first season fell below one several months before the vaccination adjusted basic reproduction numbers. This suggests that the end of the first wave outbreak largely came about due to the depletion of susceptibles, rather than due to the aggressive vaccination campaign. Nevertheless, there is evidence to suggest that the vaccination campaign mitigated the severity of the epidemic. Table 5.4 presents naive estimates of the increases in incidence and attack rates by age stratum and season, which are obtained by setting vaccine coverage to zero and simulating outbreaks from the posterior. Without vaccination, we estimate that an additional 8.1% (95% PPI: 4.0%, 12.5%) of the population would have become infected over the course of both seasons. The vaccination campaign likely mitigated the severity of the second wave, in particular. We estimate that the attack rate would have been 1.3-fold (95% PPI: 1.1, 1.5) higher in the first season, and 2.1-fold (95% PPI: 1.4, 3.3) higher in the second season absent the effects of vaccination.

In addition to mitigating the outbreak severity, we also find that the vaccination campaign may have delayed the second wave of the outbreak by roughly four months. Incidence peaked during the second season in late January to early February (posterior median: epiweek 6, 2011; 95% BCI: epiweek 4, epiweek 7). Our model predicts that, absent vaccination, incidence would have peaked between the September and mid–November (posterior median: epiweek 41, 2010; 95% PPI: epiweek 38, epiweek 47). A similar finding was reported in [194]. We caution that this result should not be over–interpreted as it is most likely an artifact of the procedure used to predict the counterfactual outcomes and the lack of data on the vaccination status of cases that could be used to construct a true counterfactual prediction. The per–contact infection rates at the start of each epoch are calibrated so that the effective reproduction numbers (5.7), which account for vaccination, are roughly one (see Figure 5.6, second row). Setting vaccination to zero has the effect of artificially inflating the effective reproduction number, and hence the per–contact infection rate, by the effective fraction of the susceptible population that is removed due to vaccination. This is why the model predicts that the epidemic takes off immediately in the third epoch.

We can make a crude adjustment to account for the artificial inflation of the FOI by

rescaling the intrinsic reproduction numbers by the reduction in R_0 that would have been achieved with vaccination in the absence of any infections or recoveries. Let $\psi_{j,\ell}^{adj}$ denote the vaccination adjusted reproduction number in week ℓ , which scaled as

$$\psi_{j,\ell}^{adj} = \psi_{j,\ell} (1 - P_j^v(t_\ell) + \nu P_j^v(t_\ell)) . \quad (5.9)$$

We set vaccination counts to zero as before and simulate outbreaks using the posterior distribution of the adjusted reproduction numbers. After adjustment, we predict that the overall attack rate would have been roughly 1.2-fold higher (95% PPI: 1.1, 1.5) in the absence of vaccination, which is lower than the unadjusted estimate. We predict that the attack rates would have been roughly 1.2-fold higher (95% PPI: 1.1, 1.3) in the first season, and 1.3-fold higher (95% PPI: 1.0, 1.8) in the second season were it not for the vaccination campaign. We also predict that the second season would have peaked between late December and late February (posterior median: epiweek 2, 2011; 95% PPI: epiweek 51, 2010, epiweek 9, 2011). Thus, after adjusting for the artificial inflation of contact rates, we do not find evidence to suggest that the second epidemic wave would have taken off in the late summer or early fall.

The posterior predictive distributions in Figure 5.7 suggest that the model does a better job reconstructing the second epidemic season than the first, particularly among individuals ages 0–19. The model appears to underestimate the peak of the outbreak among youths in the first season. One possibility is that heterogeneity in detection rates is confounding our estimates of the dynamics. In their analysis of this data, [194] estimated that detection was positively associated with incidence, and it is known that heterogeneity in detection rates can confound estimates of outbreak dynamics [40, 89, 221]. Another possibility is that the assumption of constant within-epoch variance of log differences of time-varying reproduction numbers is leading to underestimation of the peak reproduction numbers. It is challenging to tune the GMRFs so that they capture changes in the dynamics over time but are not so flexible that we lose identifiability. It is possible that introducing additional changepoints in the time-varying dynamics, or using a model that allows for non-constant variance in the GMRF increments might alleviate this problem. Finally, it is possible that dichotomizing

Table 5.4: Naive predictions of incidence and attack rates in the absence of vaccination. Increases in cumulative incidence (thousands), absolute percent increase, and relative fold-increase in attack rates without vaccination by season and age stratum.

<i>Season 1</i>			
	Incidence ($\times 10^3$)	Attack rate (Abs. % diff.)	Attack rate (Fold-change)
Ages 0-19	45.4 (21.4, 71.8)	3.7 (1.7, 5.9)	1.3 (1.1, 1.4)
Ages 20+	107 (51.9, 162)	2.6 (1.3, 3.9)	1.3 (1.1, 1.5)
All ages	154 (74.9, 226)	2.9 (1.4, 4.2)	1.3 (1.1, 1.5)

<i>Season 2</i>			
	Incidence ($\times 10^3$)	Attack rate (Abs. % diff.)	Attack rate (Fold-change)
Ages 0-19	66.1 (25.3, 125)	5.4 (2.1, 10.2)	2.0 (1.3, 3.1)
Ages 20+	202 (85.9, 359)	4.9 (2.1, 8.7)	2.2 (1.4, 3.5)
All ages	272 (121, 456)	5.1 (2.3, 8.5)	2.1 (1.4, 3.3)

<i>Both seasons</i>			
	Incidence ($\times 10^3$)	Attack rate (Abs. % diff.)	Attack rate (Fold-change)
Ages 0-19	113 (50.7, 192)	9.3 (4.1, 15.7)	1.5 (1.2, 1.9)
Ages 20+	314 (156, 509)	7.6 (3.8, 12.3)	1.6 (1.3, 2.1)
All ages	432 (216, 668)	8.1 (4, 12.5)	1.6 (1.3, 2.0)

Table 5.5: Effects of vaccination adjusted for artificial inflation of reproduction numbers in the absence of vaccination. Increases in cumulative incidence (thousands), absolute percent increase, and relative fold-increase in attack rates without vaccination by season and age stratum.

<i>Season 1</i>			
	Incidence ($\times 10^3$)	Attack rate (Abs. % diff.)	Attack rate (Fold-change)
Ages 0-19	34 (16, 52)	2.8 (1.3, 4.3)	1.2 (1.1, 1.3)
Ages 20+	75 (36, 110)	1.8 (0.9, 2.7)	1.2 (1.1, 1.4)
All ages	110 (54, 160)	2 (1.0, 3.0)	1.2 (1.1, 1.3)

<i>Season 2</i>			
	Incidence ($\times 10^3$)	Attack rate (Abs. % diff.)	Attack rate (Fold-change)
Ages 0-19	12 (-2.5, 37)	1.0 (-0.2, 3.1)	1.2 (0.97, 1.6)
Ages 20+	52 (2.7, 140)	1.3 (0.1, 3.4)	1.3 (1, 2)
All ages	65 (3.3, 170)	1.2 (0.1, 3.3)	1.3 (1, 1.8)

<i>Both seasons</i>			
	Incidence ($\times 10^3$)	Attack rate (Abs. % diff.)	Attack rate (Fold-change)
Ages 0-19	47 (18, 84)	3.8 (1.5, 6.9)	1.2 (1.1, 1.4)
Ages 20+	130 (48, 240)	3.1 (1.2, 5.7)	1.2 (1.1, 1.5)
All ages	180 (70, 310)	3.3 (1.3, 5.9)	1.2 (1.1, 1.5)

the population into two age strata does not adequately capture the age structure in contact patterns, incidence, and vaccination. A more granular model with additional age strata might be better able to capture the dynamics of the outbreak.

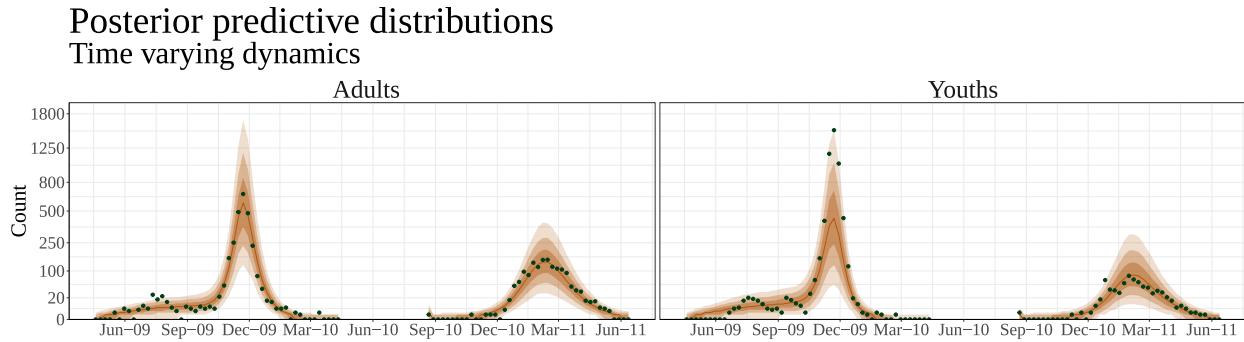


Figure 5.7: Posterior predictive distributions of a stratified SIRS ODE model with time-varying dynamics for A(H1N1)pdm09 in Finland.

5.4.3 Comparison with Piecewise Homogeneous Dynamics

For comparison, we fit a model where the outbreak dynamics were piecewise homogeneous in each epoch. The basic reproduction numbers and rates of exogenous infection were estimated for each of the three epochs, but unlike the previous model were held constant within each epoch. The remaining aspects of the model structure and emission distributions were similar to the time-varying model. Estimates of incidence and attack rates are presented in Table 5.2, posterior distributions of time-varying quantities are presented in Figure 5.8, and posterior estimates of model parameters are summarized in Table 5.3.

Attack rates under piecewise homogeneous dynamics were slightly lower than under time-varying dynamics (Table 5.2). Average dwell times in the infected and recovered states were longer, and the reduction in the rate of infectious contact for vaccinated individuals was substantially greater under piecewise homogeneous dynamics. Two striking differences in the time-varying aspects of the piecewise homogeneous model are in the rates of exogenous infection in each epoch, and the contributions of youths and adults to the force of infection.

Estimates of time varying quantities under piecewise homogeneous dynamics

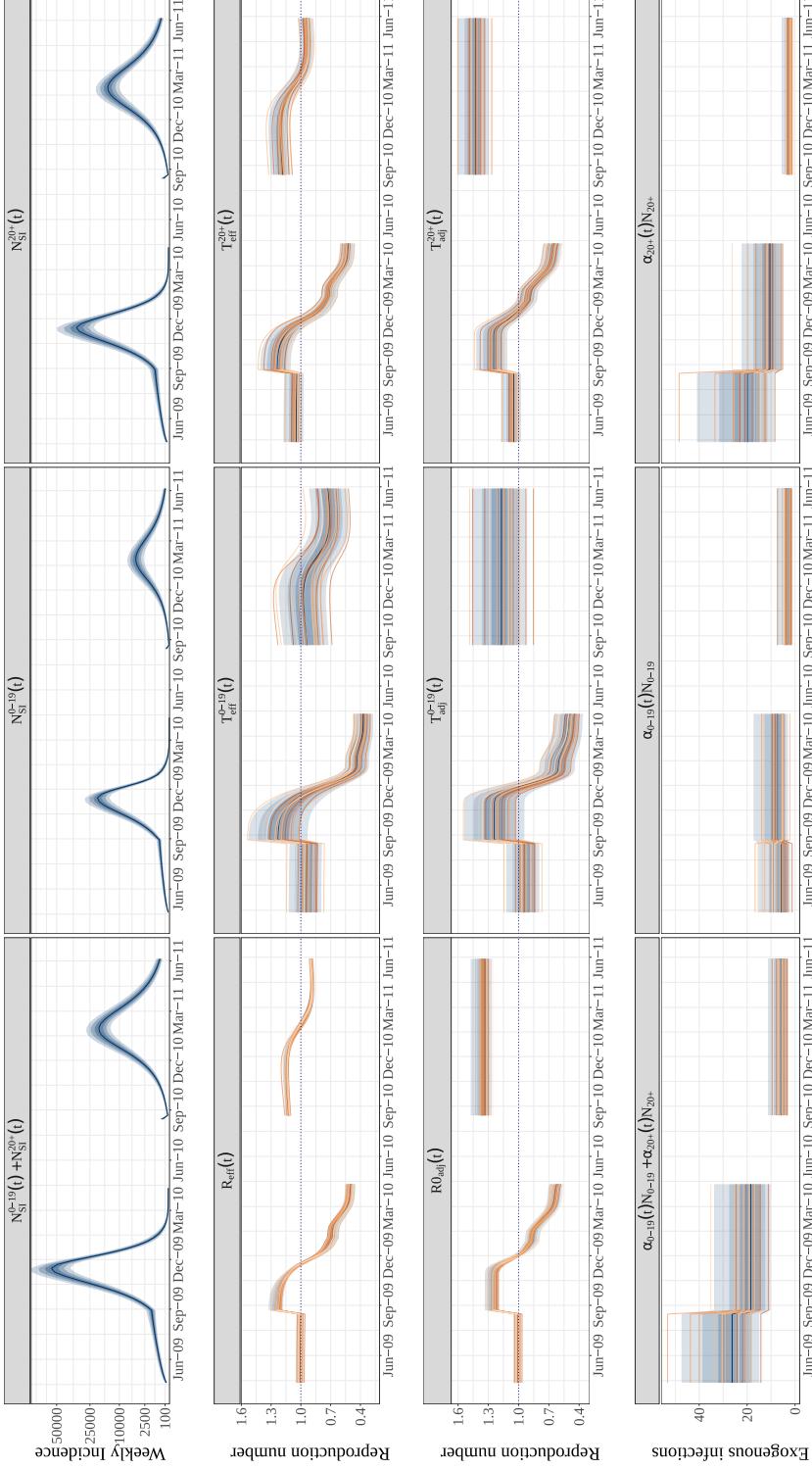


Figure 5.8: Posterior estimates of time-varying quantities under piecewise homogeneous dynamics. Estimated incidence (top row), effective reproduction numbers (second row), vaccination-adjusted basic reproduction numbers (third row), and exogenous infections (bottom row). $N_{SI}^j(t)$ is the weekly incidence in age stratum j , $R_{eff}(t)$ and $R_{adj}(t)$ are the effective and vaccination-adjusted basic reproduction numbers, $T_{eff}^j(t)$ is the type reproduction number for stratum j (expected # of secondary infections in all strata for a single infected in stratum j), and $\alpha_j(t)$ is the rate of exogenous infectious contacts for stratum j .

The highest rates of exogenous infectious contact under piecewise homogeneous dynamics are in the first epoch in the adult sub-population. This might seem paradoxical, but in fact it is an artifact of the inability of the piecewise homogeneous model to accommodate the sub-exponential dynamics that were observed in the first five months of the modeling period. It is somewhat more surprising that the piecewise homogeneous model suggests that transmission from infectious contact with youths alone could possibly have sustained the outbreak in the first epidemic season, and also possibly in the second season. Youths comprised roughly 20% of the population and only 15% of the contacts made by adults. Hence, it seems unlikely that the effective type reproduction numbers for youths in the first season, and possibly in the second epidemic season, would have been above one.

Unsurprisingly, the GMRF model is much better able to capture the shape to the outbreak than is the model with piecewise homogeneous dynamics. The posterior predictive distribution (Figure 5.9) indicates that the model with piecewise homogeneous dynamics has particular difficulty modeling the first wave of the outbreak and predicted that incidence begin to increase exponentially much sooner than was observed. It is possible that delaying the second modeling epoch several weeks to, say, late September could mitigate this behavior. Directly comparing the posterior predictive p-values and posterior predictive interval widths with those of the GMRF model shows that the piecewise homogeneous produces substantially wider predictive intervals with no obvious improvement in accuracy (Figure 5.10).

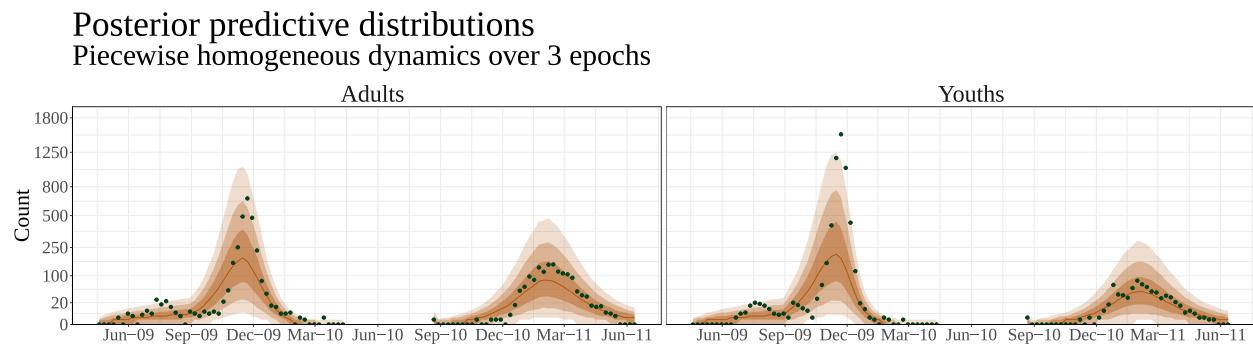


Figure 5.9: Posterior predictive distributions of a stratified SIRS ODE model with piecewise homogeneous dynamics over three epochs for A(H1N1)pdm09 in Finland.

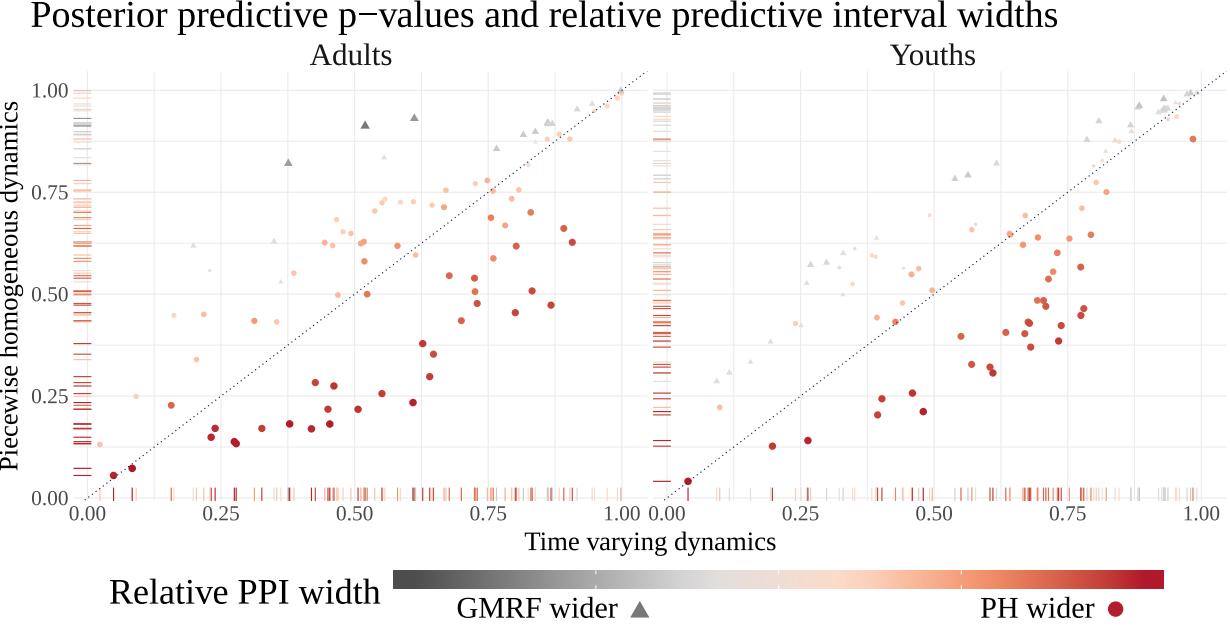


Figure 5.10: Comparison of models with time–varying (GMRF) and piecewise homogeneous (PH) dynamics using posterior predictive p–values (PPPs) and relative posterior predictive interval (PPI) widths. Each point corresponds to the observed incidence in a given week. The X–Y coordinates give the PPPs under time–varying and piecewise homogeneous dynamics, respectively. The size and color of each point corresponds to the relative PPI width, computed as $(\hat{\sigma}_{post,\ell}^{GMRF} - \hat{\sigma}_{post,\ell}^{PH})/\hat{\sigma}_{post,\ell}^{GMRF}$, and the sign of the relative width is further emphasized by the shape of the point. Red dots indicate that PPIs under piecewise homogeneous dynamics are wider and grey triangles indicate that PPIs under time–varying dynamics are wider.

5.5 Discussion

We began this chapter by highlighting the importance of allowing for the possibility that transmission dynamics might change over time. We used the LNA framework developed in Chapter 4 to fit an SIRS model with time–varying dynamics, where time–varying reproduction numbers were flexibly modeled using a Gaussian Markov random field that penalized the magnitude of week–to–week changes in basic reproduction numbers. We compared the results to those obtained using an SIRS model with time homogeneous dynamics and found that the model with time homogeneous dynamics, which unlike the flexible model was un-

able to accurately reconstruct the unobserved incidence, the time-varying aspects of the transmission dynamics, or even static model parameters.

We then developed an age-vaccination stratified SIRS model for the spread of pandemic A(H1N1) influenza in Finland over two epidemic waves. We allowed for the possibility that various static and temporal aspects of transmission dynamics were different in youths and adults, and incorporated information about the contact patterns between individuals of different ages. The model also incorporated uncertainty about vaccine efficacy and loss of immunity, rather than fix these quantities as was done in previous analyses [194]. Critically, the model allowed for time-varying dynamics within each of three epochs by assigning GMRF shrinkage priors to the intrinsic reproduction numbers and rates of exogenous infection in each age stratum. This enabled the model to greatly outperform a piecewise homogeneous model that, in some aspects, yielded somewhat unreasonable estimates.

Due to the complexity of the model, we opted to use the ODE representation in fitting it. The lower computational burden involved in fitting the ODE model allowed us to more quickly iterate through various parameterizations and to explore the effects of different priors, and in particular of alternative GMRF formulations. Run times for the model were on the order of 1–2 hours. We are confident that the model could also be fit using the LNA with some additional effort. One future line of inquiry is in comparing the statistical performance of models with varying dynamics fit via the LNA and ODE. It would be useful to know the extent to which randomness in the latent epidemic process and the time-varying aspects of its dynamics are separately identifiable from incidence data, especially if there are situations where the latent process and its dynamics are not separately identifiable.

A strength of the analysis is that the two epidemic seasons were jointly modeled along with the pre-epidemic and inter-season periods. This provides us with a principled structure for incorporating uncertainty about the state of the population at the start of each epidemic wave. This is critical to making accurate inferences about the true incidence and model dynamics, especially during the second epidemic season where the dynamics inherently depend on the attack rate during the first season and the fraction of individuals who lost immunity

during the inter–season period. Moreover, the assumption that vaccination was independent of the infection process implies that the distribution of vaccine doses at the beginning of the second season is also tied up with the attack rate in the first season (e.g., a low attack rate in the first season implies that there are many vaccinated susceptibles to start the second season). For these reasons, it is important to jointly model both epidemic seasons.

There are a few aspects of the analysis that we point to as weaknesses and that should be improved in future work. First, it is important to recognize that our estimates of the cumulative incidence are predicated on the correctness of our assumptions about the effective population size. In a supplementary analysis, we found that centering the prior for the susceptible fraction of youths at 50% instead of 25%, while holding the prior for the susceptible fraction of adults constant, yielded estimates of cumulative incidence among youths that were roughly double the estimates reported in this chapter (Section C.5.4). Frankly, it is doubtful that the reporting rates and true incidence can be unambiguously estimated from partially observed incidence data without strong assumptions when detection rates are low and in the absence of prevalence data to anchor estimates of the total outbreak size. We attempted to set the priors for the effective population size in a principled way based on known final size relations for SIR models combined with, what we felt, were reasonable assumptions about detection rates (Section C.5.2). Our estimates are also in line with estimated attack rates for A(H1N1)pdm09 in Finland [57, 195, 195] and in other settings [59, 169, 197]. In this chapter, we did not incorporate data on severe cases and hospitalizations, which would help to calibrate inference of the total outbreak size, assuming that the case–hospitalization ratio is stable. This data is available and was previously analyzed in [194].

Another aspect of the model that could be improved is the coarseness of the age structure. We dichotomized the population for simplicity and computational expediency. However, further stratifying the population would allow us to more faithfully represent age structure in contact rates, vaccination, and transmission dynamics. We suggest that, at a minimum, individuals age 0–4 could be split from individuals age 5–19, and that individuals of ages 65+ be split from other adults based on differences in the timing and coverage of vaccinations,

observed patterns in attack rates, and contact patterns.

Chapter 6

DISCUSSION AND FUTURE WORK

This dissertation has contributed computational methods for fitting stochastic epidemic models to partially observed incidence and prevalence data. Despite the ubiquity and importance of this data setting, and the historical contributions of stochastic epidemic models to the study of disease transmission, there has remained a need for computational tools that are simple, broadly applicable, and robust.

In Chapter 3, we developed an agent-based Bayesian data augmentation algorithm for fitting stochastic epidemic models to prevalence data in small to moderate size populations. This work was previously published in [74]. Historically, agent-based data augmentation algorithms for fitting stochastic epidemic models have relied on reversible-jump Markov chain Monte Carlo schemes to sample subject-level disease histories. The data agnostic proposals used in these methods are inefficient and perform poorly in the absence of subject-level data, which has all but precluded their use in the analysis of epidemic count data. To our knowledge, our algorithm is the first agent-based data augmentation algorithm for tractably fitting stochastic epidemic models in the absence of subject level information. Future lines of inquiry based on the methods developed in Chapter 3 could pursue extensions of the algorithm to other data settings, in particular to incidence data and to datasets that include both aggregate counts and incomplete subject-level data, and improvements of the computational efficiency of the algorithm.

Chapter 4 developed a computationally efficient framework based on the linear noise approximation for approximate Bayesian inference of stochastic epidemic models fit to partially observed incidence. Though the linear noise approximation has previously been used in outbreak modeling, its application has been restricted to the analysis prevalence data (or

cumulative incidence data, wrongly treated as prevalence data) where the data are normally distributed. We demonstrated how, through a series of reparameterizations, the linear noise approximation could be used in combination with state of the art Markov chain Monte Carlo samplers to efficiently fit stochastic epidemic models. We demonstrated through simulations that the resulting estimates were approximately equivalent to those obtained by more faithful approximations to the Markov jump process, and were superior to estimates obtained using deterministic methods. We also presented a series of easily implemented model diagnostics that could be used to assess in-sample model fit and compare the posterior predictive distributions of different models. We used our methods to fit several models to data from the 2014–2015 outbreak of Ebola in West Africa.

Chapter 5 explored how the linear noise approximation and ordinary differential equation frameworks could be used to fit stochastic epidemic models with time-varying dynamics. We allowed for the possibility of time-heterogeneity in the dynamics by modeling the basic reproduction number of the outbreak using Gaussian Markov random fields. This was an attractive choice for computational and statistical reasons, as the sparsity of the sub-models for the time-varying dynamics and their interpretations discretized versions of continuous-time processes made them particularly compatible with the other aspects of the modeling framework. We demonstrated in a simulation that the model with time-varying dynamics was able to recover the time-varying aspects of the outbreak along with time-homogeneous parameters of the model, whereas a time-homogeneous model yielded misleading estimates. Finally, we fit a complex age-vaccination stratified model to two seasons of data from the 2009–2011 A(H1N1) influenza pandemic in Finland. Due to the complexity of the model, we fit the model using the ordinary differential equation representation of the latent epidemic process. We used our model to obtain estimates of the transmission dynamics, outbreak size, and effects of a national vaccination campaign in mitigating the severity of the pandemic.

On a personal level, one of the great lessons I learned, or at least have tried to learn, in writing this dissertation is to be forgiving of the shortcomings in my work. I am deeply grateful to my advisors, who repeatedly reminded me to stay resilient and who gave me the

confidence to persist in our work. I wish to conclude this dissertation by pointing to some applied and methodological areas of general interest that I wish I had addressed. Were I to start this dissertation over (and I cannot emphasize strongly enough that I do not wish to do so), these are the problems I would work on.

Assimilation of data from multiple sources. The limited extent of partially observed epidemic count data, and particularly of incidence data, severely limits the strength of the conclusions we would like to draw from the data. In chapters 4 and 5, estimates of the outbreak size and detection processes were only weakly identifiable from incidence counts. The models presented in those chapters would be greatly improved by the assimilation of additional data; at a minimum, mortality data in the case of Ebola, and data on severe and hospitalized cases for the analysis of pandemic influenza. Questions of how to jointly model and weight different sources of data are of great practical importance [60].

Assessing model fit, predictive performance, and model comparison. The model diagnostics presented in this dissertation were useful for assessing the in-sample fit, but did not address the generalizability of the models and their adequacy for out-of-sample prediction. The computational cost to fitting our models is the main challenge involved in assessing their out-of-sample predictive ability since iteratively holding out data and refitting the model is problematic for anything but simple models fit to short time series. One potentially useful line of work is the development of methods for minimizing the number of refits, e.g., [39].

Combining mechanistic models. One challenge in working with mechanistic compartmental models is that we are required to make choices in specifying multifaceted models, every one of which is highly consequential in determining the model's validity. Even in the case of a model with SIR dynamics, we must make decisions about how to model the hazards, the contact structure of the population, the emission distribution, and the initial state of the population, on top of which we must decide if and how to model stochasticity in the latent epidemic process. Then, if we are Bayesian, we must assign priors to all of the parameters. In light of the obvious limitations of any particular model, an area of ongoing research is in combining models to draw more robust conclusions and improve predictions [177, 179, 227].

BIBLIOGRAPHY

- [1] Finland school calendar, 2009–2010 and 2010–2011. <https://www.feiertagskalender.ch/ferien.php?geo=3290&jahr=2010&hl=en>. Accessed: 2018-11-03.
- [2] H. Abbey. An examination of the Reed-Frost theory of epidemics. *Human Biology*, 24:201, 1952.
- [3] 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.
- [4] K. Ahnert and M. Mulansky. Odeint—solving ordinary differential equations in C++. In *AIP Conference Proceedings*, volume 1389, pages 1586–1589. AIP, 2011.
- [5] L.J.S. Allen. An introduction to stochastic epidemic models. In *Mathematical Epidemiology*, pages 81–130. Springer, New York, 2008.
- [6] L.J.S. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2:128–142, 2017.
- [7] C.S. Ambrose, T. Yi, R.E. Walker, and E.M. Connor. Duration of protection provided by live attenuated influenza vaccine in children. *The Pediatric Infectious Disease Journal*, 27:744–748, 2008.
- [8] R.M. Anderson and R.M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, 1992.
- [9] H. Andersson and T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis*. Lecture Notes in Statistics. Springer, New York, 2000.

- [10] 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.
- [11] C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Statistics and Computing*, 18:343–373, 2008.
- [12] Anon. Influenza in a boarding school. *The British Medical Journal*, 1:587, 1978.
- [13] K.E. Atkins, N.S. Wenzel, M. Ndeffo-Mbah, F.L. Altice, J.P. Townsend, and A.P. Galvani. Under-reporting and case fatality estimates for emerging epidemics. *BMJ*, 350:h1115, 2015.
- [14] K. Auranen, E. Arjas, T. Leino, and A.K. Takala. Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data. *Journal of the American Statistical Association*, 95:1044–1053, 2000.
- [15] J.A. Backer and J. Wallinga. Spatiotemporal analysis of the 2014 Ebola epidemic in West Africa. *PLoS Computational Biology*, 12:e1005210, 2016.
- [16] C. Bauer, J. Wakefield, H. Rue, S. Self, Z. Feng, and Y. Wang. Bayesian penalized spline models for the analysis of spatio-temporal count data. *Statistics in Medicine*, 35:1848–1865, 2016.
- [17] N.G. Becker. On a general stochastic epidemic model. *Theoretical Population Biology*, 11:23–36, 1977.
- [18] J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and D. West. Non-centered parameterisations for hierarchical models and data augmentation. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, volume 307. Oxford University Press, USA, 2003.

- [19] M. Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. [arXiv preprint arXiv:1701.02434](#), 2017.
- [20] M. Betancourt. Calibrating model-based inferences and decisions. [arXiv preprint arXiv:1803.08393](#), 2018.
- [21] M. Biggerstaff, S. Cauchemez, C. Reed, M. Gambhir, and L. Finelli. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. [BMC Infectious Diseases](#), 14:480, 2014.
- [22] 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.
- [23] A.J. Black and A.J. McKane. Stochastic amplification in an epidemic model with seasonal forcing. [Journal of Theoretical Biology](#), 267:85–94, 2010.
- [24] J.C. Blackwood and L.M. Childs. An introduction to compartmental modeling for the budding infectious disease modeler. [Letters in Biomathematics](#), 5:195–221, 2018.
- [25] B. Bloem-Reddy and J. Cunningham. Slice sampling on Hamiltonian trajectories. In [International Conference on Machine Learning](#), pages 3050–3058, 2016.
- [26] R.J. Boys and P.R. Giles. Bayesian inference for stochastic epidemic models with time-inhomogeneous removal rates. [Journal of Mathematical Biology](#), 55:223–247, 2007.
- [27] F. Brauer. Compartmental Models in Epidemiology. In [Mathematical Epidemiology](#), pages 18–80. Springer, New York, 2008.
- [28] F. Brauer. Mathematical epidemiology: Past, present, and future. [Infectious Disease Modelling](#), 2:113–127, 2017.
- [29] P. Brémaud. [Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues](#), volume 31. Springer Science & Business Media, New York, 1999.

- [30] 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.
- [31] 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.
- [32] T. Britton. Stochastic epidemic models: a survey. *Mathematical Biosciences*, 225:24–35, 2010.
- [33] T. Britton. Basic stochastic transmission models and their inference. *ArXiv e-prints*, January 2018.
- [34] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2013.
- [35] S.P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1998.
- [36] S.P. Brooks, A. Gelman, G. Jones, and X. Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, 2011.
- [37] P. Buchholz. Exact and ordinary lumpability in finite Markov chains. *Journal of Applied Probability*, 31:59–75, 1994.
- [38] E. Buckingham-Jeffery, V. Isham, and T. House. Gaussian process approximations for fast inference from infectious disease data. *Mathematical biosciences*, 2018.
- [39] P. Buerkner, A. Vehtari, and J. Gabry. Psis assisted m-step-ahead predictions for time-series models. Technical report, Technical report. URL <http://mc-stan.org/loo/articles/m-step-ahead-predictions.html> 999, 1000, 2018. Accessed: 2018-11-21.

- [40] P. Caley, D.J. Philp, and K. McCracken. Quantifying social distancing arising from pandemic influenza. *Journal of The Royal Society Interface*, 5:631–639, 2008.
- [41] O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer, New York, 2006.
- [42] F. Carrat, E. Vergu, N.M. Ferguson, M. Lemaitre, S. Cauchemez, S. Leach, and A.J. Valleron. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American Journal of Epidemiology*, 167:775–785, 2008.
- [43] S. Cauchemez, F. Carrat, C. Viboud, and A.J. Valleron. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in Medicine*, 23:3469–3487, 2004.
- [44] S. Cauchemez and N.M. Ferguson. Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London. *Journal of the Royal Society Interface*, 5:885–897, 2008.
- [45] Centers for Disease Control and Prevention. 2014–2016 Ebola outbreak in West Africa: Case counts. <https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html>. Last accessed: July 28, 2018.
- [46] Centers for Disease Control and Prevention. remembering the 1918 influenza pandemic. <https://www.cdc.gov/features/1918-flu-pandemic/index.html>. Last accessed: September 26, 2018.
- [47] Centers for Disease Control and Prevention. How flu spreads, 2014. Accessed on January 3, 2016.
- [48] G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2:379–398, 2017.

- [49] G. Chowell and H. Nishiura. Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC medicine*, 12:196, 2014.
- [50] G. Chowell, L. Sattenspiel, S. Bansal, and C. Viboud. Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, 18:66–97, 2016.
- [51] J.P. Chretien, S. Riley, and D.B. George. Mathematical modeling of the West Africa Ebola epidemic. *eLife*, 4:e09186, 2015.
- [52] 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.
- [53] A.J.K. Conlan, P. Rohani, A.L. Lloyd, M. Keeling, and B.T. Grenfell. Resolving the impact of waiting time distributions on the persistence of measles. *Journal of the Royal Society Interface*, page rsif20090284, 2009.
- [54] A. Cori, A.J. Valleron, F. Carrat, G.S. Tomba, G. Thomas, and P.Y. Boëlle. Estimating influenza latency and infectious period durations using viral excretion data. *Epidemics*, 4:132–138, 2012.
- [55] S.L. Cotter, G.O. Roberts, A.M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, pages 424–446, 2013.
- [56] D.R. Cox and E.J. Snell. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 248–275, 1968.
- [57] J.G. Cuesta, P. Aavitsland, H. Englund, O. Gudlaugsson, S.H. Hauge, O. Lytykäinen, G. Sigmundsdóttir, A. Tegnell, M. Virtanen, and T.G. Krause. Pandemic vaccination strategies and influenza severe outcomes during the influenza A(H1N1)pdm09 pandemic and the post-pandemic influenza season: the Nordic experience. *Eurosurveillance*, 21, 2016.

- [58] J.R. Davies, A.J. Smith, E.A. Grilli, and T.W. Hoskins. Christ's Hospital 1978–79: An account of two outbreaks of influenza A H1N1. *Journal of Infection*, 5:151–156, 1982.
- [59] F.S. Dawood, A.D. Iuliano, C. Reed, M.I. Meltzer, D.K. Shay, P. Cheng, D. Ban-daranayake, R.F. Breiman, W.A. Brooks, P. Buchy, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A(H1N1) virus circulation: a modelling study. *The Lancet infectious diseases*, 12:687–695, 2012.
- [60] D. De Angelis, A.M. Presanis, P.J. Birrell, G.S. Tomba, and T. House. Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics*, 10:83–87, 2015.
- [61] 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.
- [62] O. Diekmann, J.A.P. Heesterbeek, and M.G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, page rsif20090386, 2009.
- [63] 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.
- [64] 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.
- [65] D. Eddelbuettel and R. François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40:1–18, 2011.

- [66] D. Eddelbuettel and C. Sanderson. RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063, 2014.
- [67] J. Elf and M. Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome research*, 13:2475–2484, 2003.
- [68] P. Fearnhead, V. Giagos, and C. Sherlock. Inference for reaction networks using the linear noise approximation. *Biometrics*, 70:457–466, 2014.
- [69] 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.
- [70] B.F. Finkenstädt, O.N. Bjørnstad, and B.T. Grenfell. A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks. *Biostatistics*, 3:493–510, 2002.
- [71] B.F. Finkenstädt and B.T. Grenfell. Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49:187–205, 2000.
- [72] J. Fintzi. *ECctmc: Simulation from endpoint-conditioned continuous time Markov chains*, 2017. R package, version 0.2.3.
- [73] J. Fintzi. *stemr: Fit stochastic epidemic models via Bayesian data augmentation*, 2018. R package, version 0.2.1.
- [74] 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.

- [75] L. Fisher, J. Wakefield, C. Bauer, and S. Self. Time series modeling of pathogen-specific disease probabilities with subsampled data. *Biometrics*, 73:283, 2017.
- [76] J.M. Flegal, J. Hughes, D.Vats, and N. Dai. *mcmcse: Monte Carlo standard errors for MCMC*, 2017. R package version 1.3-2.
- [77] M.M. Folia and M. Rattray. Trajectory inference and parameter estimation in stochastic models with temporally aggregated data. *Statistics and Computing*, pages 1–20, 2017.
- [78] C. Fuchs. *Inference for Diffusion Processes: With Applications in Life Sciences*. Springer Science & Business Media, New York, 2013.
- [79] S. Funk. socialmixr: Social mixing matrices for infectious disease modelling. 2018.
- [80] T. Garske, A. Cori, A. Ariyarajah, I.M. Blake, I. Dorigatti, T. Eckmanns, C. Fraser, W. Hinsley, T. Jombart, H.L. Mills, et al. Heterogeneities in the case fatality ratio in the West African Ebola outbreak 2013–2016. *Philosophical Transactions of the Royal Society Series B*, 372:20160308, 2017.
- [81] A. Gelman, W.R. Gilks, and G.O. Roberts. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7:110–120, 1997.
- [82] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472, 1992.
- [83] A. Gelman, H.S. Stern, J.B. Carlin, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*. CRC Press, Boca Raton, third edition edition, 2013.
- [84] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984.

- [85] V. Giagos. Inference for auto-regulatory genetic networks using diffusion process approximations. PhD dissertation, Lancaster University, 2010.
- [86] G.J. Gibson and E. Renshaw. Estimating parameters in stochastic compartmental models using Markov chain methods. Mathematical Medicine and Biology, 15:19–40, 1998.
- [87] 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.
- [88] D.T. Gillespie. The chemical Langevin equation. The Journal of Chemical Physics, 113:297–306, 2000.
- [89] K. Glass, G.N. Mercer, H. Nishiura, E.S. McBryde, and N.G. Becker. Estimating reproduction numbers for adults and children from case data. Journal of The Royal Society Interface, 8:1248–1259, 2011.
- [90] 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.
- [91] J.R. Glynn, H. Bower, S. Johnson, C. Turay, D. Sesay, S.H. Mansaray, O. Kamara, A.J. Kamara, M.S. Bangura, and F. Checchi. Variability in intrahousehold transmission of Ebola virus, and estimation of the household secondary attack rate. The Journal of Infectious Diseases, 217:232–237, 2017.
- [92] A. Golightly and C.S. Gillespie. Simulation of stochastic kinetic models. In In Silico Systems Biology, pages 169–187. Springer, 2013.
- [93] 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.

- [94] A. Golightly and T. Kypraios. Efficient SMC2 schemes for stochastic kinetic models. *Statistics and Computing*, 28:1215–1230, 2018.
- [95] A. Golightly and C. Sherlock. Efficient sampling of conditioned markov jump processes. *arXiv preprint arXiv:1809.07139*, 2018.
- [96] M. Greenwood. On the statistical measure of infectiousness. *Epidemiology & Infection*, 31:336–351, 1931.
- [97] P.E. Greenwood and L.F. Gordillo. Stochastic epidemic modeling. In Mathematical and Statistical Estimation Approaches in Epidemiology, pages 31–52. Springer, New York, 2009.
- [98] R. Grima. A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *The Journal of Chemical Physics*, 136:04B616, 2012.
- [99] P. Guttorp. Stochastic modeling of scientific data. CRC Press, Boca Raton, 1995.
- [100] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. 1970.
- [101] J.A.P. Heesterbeek and M.G. Roberts. The type-reproduction number T in models for infectious disease control. *Mathematical biosciences*, 206:3–10, 2007.
- [102] J.M. Heffernan, R.J. Smith, and L.M. Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2:281–293, 2005.
- [103] 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.
- [104] L. Held and S. Meyer. Forecasting based on surveillance data. *arXiv preprint arXiv:1809.03735*, 2018.

- [105] L. Held and M. Paul. Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal*, 54:824–843, 2012.
- [106] K.L. Hey, H. Momiji, K. Featherstone, J.R.E. Davis, M.R.H. White, D.A. Rand, and B. Finkenstädt. A stochastic transcriptional switch model for single cell imaging data. *Biostatistics*, 16:655–669, 2015.
- [107] M.W. Hirsch, S. Smale, and R.L. Devaney. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press. Academic Press, Waltham, 2013.
- [108] L.S.T. Ho, F.W. Crawford, and M.A. Suchard. Direct likelihood-based inference for discretely observed stochastic compartmental models of infectious disease. *arXiv preprint arXiv:1608.06769*, 2016.
- [109] 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.
- [110] M. Höhle and E. Jørgensen. Estimating parameters for stochastic epidemics. Technical Report 102, The Royal Veterinary and Agricultural University, November 2002.
- [111] H. Hu, K. Nigmatulina, and P. Eckhoff. The scaling of contact rates with population density for the infectious disease models. *Mathematical Biosciences*, 244:125–134, 2013.
- [112] E.L. Ionides, A. Bhadra, Y. Atchadé, A.A. King, et al. Iterated filtering. *The Annals of Statistics*, 39:1776–1802, 2011.
- [113] E.L. Ionides, C. Bretó, and A.A. King. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103:18438–18443, 2006.
- [114] A.D. Iuliano, K.M. Roguski, H.M. Chang, D.J. Muscatello, R. Palekar, S. Tempia, C. Cohen, J.M. Gran, D. Schanzer, B.J. Cowling, et al. Estimates of global seasonal

influenza-associated respiratory mortality: a modelling study. *The Lancet*, 391:1285–1300, 2018.

- [115] A. Jacks, J. Ollgren, T. Ziegler, and O. Lyytikäinen. Influenza-associated hospitalisations in Finland from 1996 to 2010: unexpected age-specific burden during the influenza A(H1N1)pdm09 pandemic from 2009 to 2010. *Eurosurveillance*, 17:20276, 2012.
- [116] R. Jandarov, M. Haran, O. Bjørnstad, and B. Grenfell. Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63:423–444, 2014.
- [117] C.P. Jewell, T. Kypraios, P. Neal, and G.O. Roberts. Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4:465–496, 2009.
- [118] M.D. Karcher, J.A. Placios, T. Bedford, M.A. Suchard, and V.N. Minin. Quantifying and mitigating the effect of preferential sampling on phylodynamic inference. *PLOS Computational Biology*, 12:e1004789, 2015.
- [119] M.J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, 2008.
- [120] T.H. Keitt. *odeintr: C++ ODE Solvers Compiled on-Demand*, 2017. R package version 1.7.1.
- [121] H. Kelly, H.A. Peck, K.L. Laurie, P. Wu, H. Nishiura, and B.J. Cowling. The age-specific cumulative incidence of infection with pandemic influenza H1N1 2009 was similar in various countries prior to vaccination. *PLoS One*, 6:e21828, 2011.
- [122] 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.

- [123] 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.
- [124] A.A. King, D. Nguyen, and E.L. Ionides. Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69:1–43, 2016.
- [125] A.A. King, D. Nguyen, and E.L. Ionides. Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69:1–43, 2016.
- [126] L. Knorr-Held and H. Rue. On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, 29:597–614, 2002.
- [127] A.A. Koepke, I.M. Longini Jr., M.E. Halloran, J. Wakefield, and V.N. Minin. Predictive modeling of Cholera outbreaks in Bangladesh. *The Annals of Applied Statistics*, 10:575–595, 2016.
- [128] 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.
- [129] T.G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7:49–58, 1970.
- [130] T.G. Kurtz. Limit theorems for sequences of jump Markov processes. *Journal of Applied Probability*, 8:344–356, 1971.
- [131] T.G. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57:2976–2978, 1972.

- [132] T.G. Kurtz. Approximation of population processes, volume 36. Society for Industrial and Applied Mathematics, Philadelphia, 1981.
- [133] T. Kypraios and P.D. O'Neill. Bayesian nonparametrics for stochastic epidemic models. Statistical Science, 33:44–56, 2018.
- [134] J.M. Lange and V.N. Minin. Fitting and interpreting continuous-time latent Markov models for panel data. Statistics in Medicine, 32:4581–4595, 2013.
- [135] L.E. Lansbury, S. Smith, W. Beyer, E. Karamehic, E. Pasic-Juhas, H. Sikira, A. Mateus, H. Oshitani, H. Zhao, C.R. Beck, et al. Effectiveness of 2009 pandemic influenza A(H1N1) vaccines: A systematic review and meta-analysis. Vaccine, 35:1996–2006, 2017.
- [136] E.H.Y. Lau and P.S.F. Yip. Estimating the basic reproductive number in the general epidemic model with an unknown initial number of susceptible individuals. Scandinavian Journal of Statistics, 35:650–663, 2008.
- [137] M.S.Y. Lau, G. Marion, G. Streftaris, and G.J. Gibson. New model diagnostics for spatio-temporal systems in epidemiology and ecology. Journal of The Royal Society Interface, 11:20131093, 2014.
- [138] 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.
- [139] J. Lessler and D.A.T. Cummings. Mechanistic models of infectious disease and their impact on public health. American Journal of Epidemiology, 183:415–422, 2016.
- [140] F. Liang, C. Liu, and R. Carroll. Advanced Markov chain Monte Carlo methods: learning from past samples, volume 714. John Wiley & Sons, Hoboken, 2011.

- [141] D. Lindenstrand and Å. Svensson. Estimation of the Malthusian parameter in an stochastic epidemic model using martingale methods. *Mathematical Biosciences*, 246:272–279, 2013.
- [142] E.T. Lofgren, M.E. Halloran, C.M. Rivers, J.M. Drake, T.C. Porco, B. Lewis, W. Yang, A. Vespignani, J. Shaman, J.N.S. Eisenberg, M.C. Eisenberg, M. Marathe, S.V. Scarpino, K.A. Alexander, R. Meza, M.J. Ferrari, J.M. Hyman, L.A. Meyers, and S. Eubank. Opinion: Mathematical models: A key tool for outbreak response. *Proceedings of the National Academy of Sciences*, 111:18095–18096, 2014.
- [143] I.M. Longini Jr. and J.S. Koopman. Household and community transmission parameters from final distributions of infections in households. *Biometrics*, 38:115–126, 1982.
- [144] O. Lyytikäinen, M. Kuusi, M. Snellman, M.J. Virtanen, J. Eskola, E. Rönkkö, N. Ikonen, I. Julkunen, T. Ziegler, and P. Ruutu. Surveillance of influenza in finland during the 2009 pandemic, 10 may 2009 to 8 march 2010. *Eurosurveillance*, 16:19908, 2011.
- [145] M. Martinez-Bakker, A. A. King, and P. Rohani. Unraveling the transmission ecology of polio. *PLoS Biology*, 13, 2015.
- [146] H. McCallum, N. Barlow, and J. Hone. How should pathogen transmission be modelled? *Trends in ecology & evolution*, 16(6):295–300, 2001.
- [147] T. McKinley, A.R. Cook, and R. Deardon. Inference in epidemic models without likelihoods. *The International Journal of Biostatistics*, 5:1–40, 2009.
- [148] T.J. McKinley, J.V. Ross, R. Deardon, and A.R. Cook. Simulation-based Bayesian inference for epidemic models. *Computational Statistics & Data Analysis*, 71:434–447, 2014.
- [149] T.J. McKinley, I. Vernon, I. Andrianakis, N. McCreesh, J.E. Oakley, R.N. Nsubuga, M. Goldstein, and R.G. White. Approximate Bayesian computation and simulation-

based inference for complex stochastic epidemic models. *Statistical Science*, 33:4–18, 2018.

- [150] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092, 1953.
- [151] S. Meyer and L. Held. Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, 18:338–351, 2017.
- [152] J.C. Miller. A note on the derivation of epidemic final sizes. *Bulletin of Mathematical Biology*, 74:2125–2141, 2012.
- [153] G. Minas and D.A. Rand. Long-time analytic approximation of large stochastic oscillators: Simulation, analysis and inference. *PLoS Computational Biology*, 13:e1005676, 2017.
- [154] A. Mira and L. Tierney. Efficiency and convergence properties of slice samplers. *Scandinavian Journal of Statistics*, 29:1–12, 2002.
- [155] 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.
- [156] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G.S. Tomba, J. Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, 5:e74, 2008.
- [157] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G.S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W.J. Edmunds. Polymod social contact data, 2017. Version 1.1.

- [158] I. Murray and R.P. Adams. Slice sampling covariance hyperparameters of latent Gaussian models. In *Advances in Neural Information Processing Systems*, pages 1732–1740, 2010.
- [159] I. Murray, R.P. Adams, and D.J.C. MacKay. Elliptical slice sampling. *JMLR: W&CP*, 9:541–548, 2010.
- [160] P. Neal and G.O. Roberts. A case study in non-centering for data augmentation: stochastic epidemics. *Statistics and Computing*, 15:315–327, 2005.
- [161] P.J. Neal and G.O. Roberts. Statistical inference and model selection for the 1861 Hageloch measles epidemic. *Biostatistics*, 5:249–261, 2004.
- [162] R.M. Neal. Regression and classification using gaussian process priors. *Bayesian Statistics*, 6:475, 1998.
- [163] R.M. Neal. Slice sampling. *Annals of Statistics*, pages 705–741, 2003.
- [164] B. Øksendal. *Stochastic Differential Equations*. Springer, New York, 2003.
- [165] P.D. O'Neill. A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Mathematical Biosciences*, 180:103–114, 2002.
- [166] P.D. O'Neill. Bayesian inference for stochastic multitype epidemics in structured populations using sample data. *Biostatistics*, 10:779–791, 2009.
- [167] P.D. O'Neill. Introduction and snapshot review: relating infectious disease transmission models to data. *Statistics in Medicine*, 29:2069–2077, 2010.
- [168] 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.

- [169] L. Opatowski, C. Fraser, J. Griffin, E. de Silva, M.D. Van Kerkhove, E.J. Lyons, S. Cauchemez, and N.M. Ferguson. Transmission characteristics of the 2009 H1N1 influenza pandemic: comparison of 8 southern hemisphere countries. *PLoS Pathogens*, 7:e1002225, 2011.
- [170] O. Papaspiliopoulos, G.O. Roberts, and M. Sköld. Non-centered parameterisations for hierarchical models and data augmentation. *Bayesian Statistics*, 7:307–326, 2003.
- [171] O. Papaspiliopoulos, G.O. Roberts, and M. Sköld. A general framework for the parametrization of hierarchical models. *Statistical Science*, pages 59–73, 2007.
- [172] M. Paul and L. Held. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Statistics in Medicine*, 30:1118–1136, 2011.
- [173] M. Plummer, N. Best, K. Cowles, and K. Vines. Coda: Convergence diagnosis and output analysis for MCMC. *R News*, 6:7–11, 2006.
- [174] 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.
- [175] Z. Qin and C.R. Shelton. Auxiliary Gibbs sampling for inference in piecewise-constant conditional intensity models. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, 2015.
- [176] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.
- [177] E.L. Ray and N.G. Reich. Prediction of infectious disease epidemics via weighted density ensembles. *PLoS Computational Biology*, 14:e1005910, 2018.

- [178] N.P. Rebuli, N.G. Bean, and J.V. Ross. Hybrid Markov chain models of S–I–R disease dynamics. *Journal of Mathematical Biology*, 75:521–541, 2017.
- [179] N.G. Reich, L. Brooks, S. Fox, S. Kandula, C. McGowan, E. Moore, D. Osthuis, E.L. Ray, A. Tushar, T. Yamana, M. Biggerstaff, M.A. Johansson, R. Rosenfeld, and J. Shamman. Forecasting seasonal influenza in the US: A collaborative multi-year, multi-model assessment of forecast performance. *bioRxiv*, page 397190, 2018.
- [180] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Science & Business Media, New York, 2013.
- [181] G.O. Roberts and J.S. Rosenthal. Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61:643–660, 1999.
- [182] G.O. Roberts and J.S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16:351–367, 2001.
- [183] G.O. Roberts and J.S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18:349–367, 2009.
- [184] G.O. Roberts and S.K. Sahu. Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59:291–317, 1997.
- [185] G.O. Roberts and O. Stramer. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, 88:603–621, 2001.
- [186] J.V. Ross. On parameter estimation in population models III: Time-inhomogeneous processes and observation error. *Theoretical Population Biology*, 82:1–17, 2012.
- [187] 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.

- [188] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*. CRC Press, Boca Raton, 2005.
- [189] 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.
- [190] 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.
- [191] 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.
- [192] C.R. Shelton and G. Ciardo. Tutorial on structured continuous-time Markov processes. *Journal of Artificial Intelligence Research*, 51:725–778, 2014.
- [193] A.Y. Shestopaloff and R.M. Neal. Sampling latent states for high-dimensional non-linear state space models with the embedded HMM method. *Bayesian Analysis*, 2018.
- [194] M. Shubin, A. Lebedev, O. Lyytikäinen, and K. Auranen. Revealing the true incidence of pandemic A(H1N1) pdm09 influenza in Finland during the first two seasons — an analysis based on a dynamic transmission model. *PLoS computational biology*, 12:e1004803, 2016.
- [195] M. Shubin, M. Virtanen, S. Toikkanen, O. Lyytikäinen, and K. Auranen. Estimating the burden of A(H1N1) pdm09 influenza in Finland during two seasons. *Epidemiology & Infection*, 142:964–974, 2014.
- [196] V. Stathopoulos and M.A. Girolami. Markov chain Monte Carlo inference for Markov jump processes via the linear noise approximation. *Philosophical Transactions of the Royal Society Series A*, 371:20110541, 2013.

- [197] A. Steens, S. Waaijenborg, P.F.M. Teunis, J.H.J. Reimerink, A. Meijer, M. van der Lubben, M. Koopmans, M. van der Sande, J. Wallinga, and M. van Boven. Age-dependent patterns of infection and severity explaining the low impact of 2009 influenza A(H1N1): evidence from serial serologic surveys in the Netherlands. *American journal of epidemiology*, 174:1307–1315, 2011.
- [198] A. Sudbury. The proportion of the population never hearing a rumour. *Journal of Applied Probability*, 22:443–446, 1985.
- [199] R.K. Syrjänen, J. Jokinen, T. Ziegler, J. Sundman, M. Lahdenkari, T. Julkunen, and T.M. Kilpi. Effectiveness of pandemic and seasonal influenza vaccines in preventing laboratory-confirmed influenza in adults: a clinical cohort study during epidemic seasons 2009–2010 and 2010–2011 in Finland. *PloS One*, 9:e108538, 2014.
- [200] WHO Ebola Response Team. Ebola virus disease in West Africa — the first 9 months of the epidemic and forward projections. *New England Journal of Medicine*, 371:1481–1495, 2014.
- [201] P. Thomas, A.V. Straube, and R. Grima. The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. *BMC Systems Biology*, 6:39, 2012.
- [202] M.B. Thompson. *Slice Sampling with Multivariate Steps*. PhD dissertation, 2011.
- [203] J.P. Tian and D. Kannan. Lumpability and commutativity of Markov processes. *Stochastic Analysis and Applications*, 24:685–702, 2006.
- [204] 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.
- [205] A.C. Titman and L.D. Sharples. Semi-Markov models with phase-type sojourn distributions. *Biometrics*, 66:742–752, 2010.

- [206] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M.P.H. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6:187–202, 2009.
- [207] 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.
- [208] P. van den Driessche. Reproduction numbers of infectious disease models. *Infectious Disease Modelling*, 2:288–303, 2017.
- [209] P. van den Driessche and J. Watmough. Further notes on the basic reproduction number. In *Mathematical epidemiology*, pages 159–178. Springer, 2008.
- [210] N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, third edition edition, 2007.
- [211] M.D. Van Kerkhove, S. Hirve, A. Koukounari, A.W. Mounts, and H1N1pdm Serology Working Group. Estimating age-specific cumulative incidence for the 2009 influenza pandemic: a meta-analysis of A(H1N1) pdm09 serological studies from 19 countries. *Influenza and other respiratory viruses*, 7:872–886, 2013.
- [212] 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.
- [213] C. Viboud, K. Sun, R. Gaffey, M. Ajelli, L. Fumanelli, S. Merler, Q. Zhang, G. Chowell, L. Simonsen, A. Vespignani, et al. The RAPIDD Ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, 22:13–21, 2018.
- [214] M.A. Vink, M.C.J. Bootsma, and J. Wallinga. Serial intervals of respiratory infectious diseases: a systematic review and analysis. *American Journal of Epidemiology*, 180:865–875, 2014.

- [215] J. Wakefield, T.Q. Dong, and V.N. Minin. Spatio-temporal analysis of surveillance data. *arXiv preprint arXiv:1711.00555*, 2017.
- [216] 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.
- [217] R. Watson. An application of a martingale central limit theorem to the standard epidemic model. *Stochastic Processes and Their Applications*, 11:79–89, 1981.
- [218] H.J. Wearing, P. Rohani, and M.J. Keeling. Appropriate models for the management of infectious diseases. *PLOS Medicine*, 2:e174, 2005.
- [219] G.A. Whitaker. *Bayesian inference for stochastic differential mixed-effects models*. PhD dissertation, 2016.
- [220] G.A. Whitaker, A. Golightly, R.J. Boys, and C. Sherlock. Bayesian inference for diffusion-driven mixed-effects models. *Bayesian Analysis*, 12:435–463, 2017.
- [221] L.F. White and M. Pagano. Reporting errors in infectious disease outbreaks, with an application to Pandemic Influenza A/H1N1. *Epidemiologic Perspectives & Innovations*, 7:12, 2010.
- [222] D.J. Wilkinson. *Stochastic Modelling for Systems Biology*. CRC Press, Boca Raton, 2011.
- [223] World Health Organization. Situation report: Ebola virus disease – 30 March 2016. <http://apps.who.int/ebola/current-situation/ebola-situation-report-30-march-2016>. Last accessed: July 28, 2018.
- [224] 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.

- [225] X. Xu, T. Kypraios, and P.D. O'Neill. Bayesian non-parametric inference for stochastic epidemic models using Gaussian processes. *Biostatistics*, 17:619–633, 2016.
- [226] W. Yang, M. Lipsitch, and J. Shaman. Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences*, 112:2723–2728, 2015.
- [227] Y. Yao, A. Vehtari, D.P. Simpson, and A. Gelman. Using stacking to average Bayesian predictive distributions. *Bayesian Analysis*, 13:917–1007, 2018.
- [228] Y. Yu and X. Meng. To center or not to center: That is not the question — An Ancillarity–Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, 20:531–570, 2011.
- [229] C. Zimmer and S. Sahle. Deterministic inference for stochastic systems using multiple shooting and a linear noise approximation for the transition probabilities. *IET systems biology*, 9:181–192, 2015.
- [230] C. Zimmer, R. Yaesoubi, and R. Cohen. A likelihood approach for real-time calibration of stochastic compartmental epidemic models. *PLoS Computational Biology*, 13:e1005257, 2017.

Appendix A

APPENDIX TO CHAPTER 3

A.1 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

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

where $\boldsymbol{\Lambda}$ is the transition rate matrix for the CTMC and \mathbf{I} is an identity matrix of the same size as $\boldsymbol{\Lambda}$ [222]. Therefore, \mathbf{P} is computed using the matrix exponential solution of the above differential equation, $\mathbf{P} = \exp(t\boldsymbol{\Lambda})$. 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 $\boldsymbol{\Lambda}$ and caching the resulting eigenvalues and eigenvectors will be relatively efficient [155]. We outline this computation in the following two cases: when the eigenvalues of $\boldsymbol{\Lambda}$ are all real (e.g. as with the SIR and SEIR models), and when $\boldsymbol{\Lambda}$ has complex eigenvalues (e.g. as is possibly the case with the SIRS model).

A.1.1 Case 1: $\boldsymbol{\Lambda}$ has real eigenvalues

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

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

That \mathbf{U} is nonsingular yields

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

A.1.2 Case 2: Λ has complex eigenvalues

In the event that Λ has complex eigenvalues, we may obtain a real-valued TPM by transforming Λ into its real canonical form [107]. Suppose that Λ has r real eigenvalues, v_1, \dots, v_r , with corresponding real eigenvectors, $\mathbf{u}_1, \dots, \mathbf{u}_r$, 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 $j = r + 1, \dots, n$, and define the matrix $\mathbf{T} = (\mathbf{u}_1 | \dots | \mathbf{u}_r | \mathbf{u}_{r+1} | \mathbf{w}_{r+1} | \dots | \mathbf{u}_n | \mathbf{w}_n)$. The real canonical form for a rate matrix with complex eigenvalues can now be written as $\mathbf{V} = \mathbf{T}^{-1}\Lambda\mathbf{T}$, where $\mathbf{V} = \text{diag}(v_1, \dots, v_r, \mathbf{B}_{r+1}, \dots, \mathbf{B}_n)$, and each \mathbf{B}_j , $j = r + 1, \dots, n$ is given by

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

which implies that

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

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

A.2 Conjugate priors for SIR, SEIR, and SIRS model parameters

Parameter	Conjugate Prior Dist.	Prior Hyperparameters	Full Conditional Hyperparameters
R_0	Beta'	$a_\beta, a_\mu, 1, \frac{b_\mu N}{b_\beta}$	—
β	Gamma	a_β, b_β	$a_\beta + \sum_{j=1}^M \mathbb{1}_{\{\tau_j \leq I\}}, b_\beta + \sum_{j=1}^M S_{\tau_{j-1}} I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$
μ	Gamma	a_μ, b_μ	$a_\mu + \sum_{j=1}^M \mathbb{1}_{\{\tau_j \leq R\}}, b_\mu + \sum_{j=1}^M I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$
ρ	Beta	a_ρ, b_ρ	$a_\rho + \sum_{j=1}^L Y_{t_j}, b_\rho + \sum_{j=1}^L (I_{t_j} - Y_{t_j})$
\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 A.1: Prior and full conditional distributions for SIR model parameters. β is the per-contact infectivity rate, μ is the recovery rate, ρ 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 $\text{Gamma}(a, b)$ 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 β and μ . The indicators $\mathbb{1}_{\{\tau_j \leq I\}}$ and $\mathbb{1}_{\{\tau_j \leq R\}}$ equal 1 if τ_j corresponds to a time when an individual becomes infected or recovers.

Parameter	Conjugate Prior Dist.	Prior Hyperparameters	Full Conditional Hyperparameters
R_0	Beta'	$a_\beta, a_\mu, 1, \frac{b_\mu N}{b_\beta}$	—
β	Gamma	a_β, b_β	$a_\beta + \sum_{j=1}^M \mathbb{1}_{\{\tau_j \in E\}}, b_\beta + \sum_{j=1}^M S_{\tau_{j-1}} I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$
γ	Gamma	a_γ, b_γ	$a_\gamma + \sum_{j=1}^M \mathbb{1}_{\{\tau_j \in I\}}, b_\gamma + \sum_{j=1}^M E_{\tau_{j-1}} (\tau_j - \tau_{j-1})$
μ	Gamma	a_μ, b_μ	$a_\mu + \sum_{j=1}^M \mathbb{1}_{\{\tau_j \in R\}}, b_\mu + \sum_{j=1}^M I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$
ρ	Beta	a_ρ, b_ρ	$a_\rho + \sum_{j=1}^L Y_{t_j}, b_\rho + \sum_{j=1}^L (I_{t_j} - Y_{t_j})$
\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 A.2: Prior and full conditional distributions for SEIR model parameters. β is the per-contact infectivity rate, γ is the rate at which an exposed individual becomes infectious, μ is the recovery rate, ρ 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 $\text{Gamma}(a, b)$ 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 β and μ . The indicators $\mathbb{1}_{\{\tau_j \in E\}}$, $\mathbb{1}_{\{\tau_j \in I\}}$ and $\mathbb{1}_{\{\tau_j \in R\}}$ equal 1 if τ_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'	$a_\beta, a_\mu, 1, \frac{b_\mu N}{b_\beta}$	—
β	Gamma	a_β, b_β	$a_\beta + \sum_{j=1}^M \mathbb{1}_{\{\tau_j \leq E\}}, b_\beta + \sum_{j=1}^M S_{\tau_{j-1}} I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$
μ	Gamma	a_μ, b_μ	$a_\mu + \sum_{j=1}^M \mathbb{1}_{\{\tau_j \leq R\}}, b_\mu + \sum_{j=1}^M I_{\tau_{j-1}} (\tau_j - \tau_{j-1})$
γ	Gamma	a_γ, b_γ	$a_\gamma + \sum_{j=1}^M \mathbb{1}_{\{\tau_j \leq L\}}, b_\gamma + \sum_{j=1}^M R_{\tau_{j-1}} (\tau_j - \tau_{j-1})$
ρ	Beta	a_ρ, b_ρ	$a_\rho + \sum_{j=1}^L Y_{t_j}, b_\rho + \sum_{j=1}^L (I_{t_j} - Y_{t_j})$
\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 A.3: Prior and full conditional distributions for SIRS model parameters. β is the per-contact infectivity rate, μ is the recovery rate, γ is the rate at which a recovered individual loses immunity, ρ 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 $\text{Gamma}(a, b)$ 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 β and μ . The indicators $\mathbb{1}_{\{\tau_j \leq I\}}$, $\mathbb{1}_{\{\tau_j \leq R\}}$, and $\mathbb{1}_{\{\tau_j \leq L\}}$ equal 1 if τ_j corresponds to a time when an individual becomes infected, recovers, or loses immunity.

A.3 Simulation 1 — Inference Under Various Epidemic Dynamics — Setup and Additional Results

A.3.1 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 A.4) 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.

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

Table A.4: Prior distributions for SIR model and measurement process parameters. The prior for R_0 is the induced prior implied by β and μ . The per-contact infectivity rate is β , the recovery rate is μ , the binomial sampling probability is ρ , and the initial state probabilities are \mathbf{p}_{t_1} .

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 τ -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 β and μ , a logit transformation to ρ , and a generalized logit transformation to \mathbf{p}_{t_1} .

A.3.2 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 A.5: Run times, log-posterior 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 τ -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.

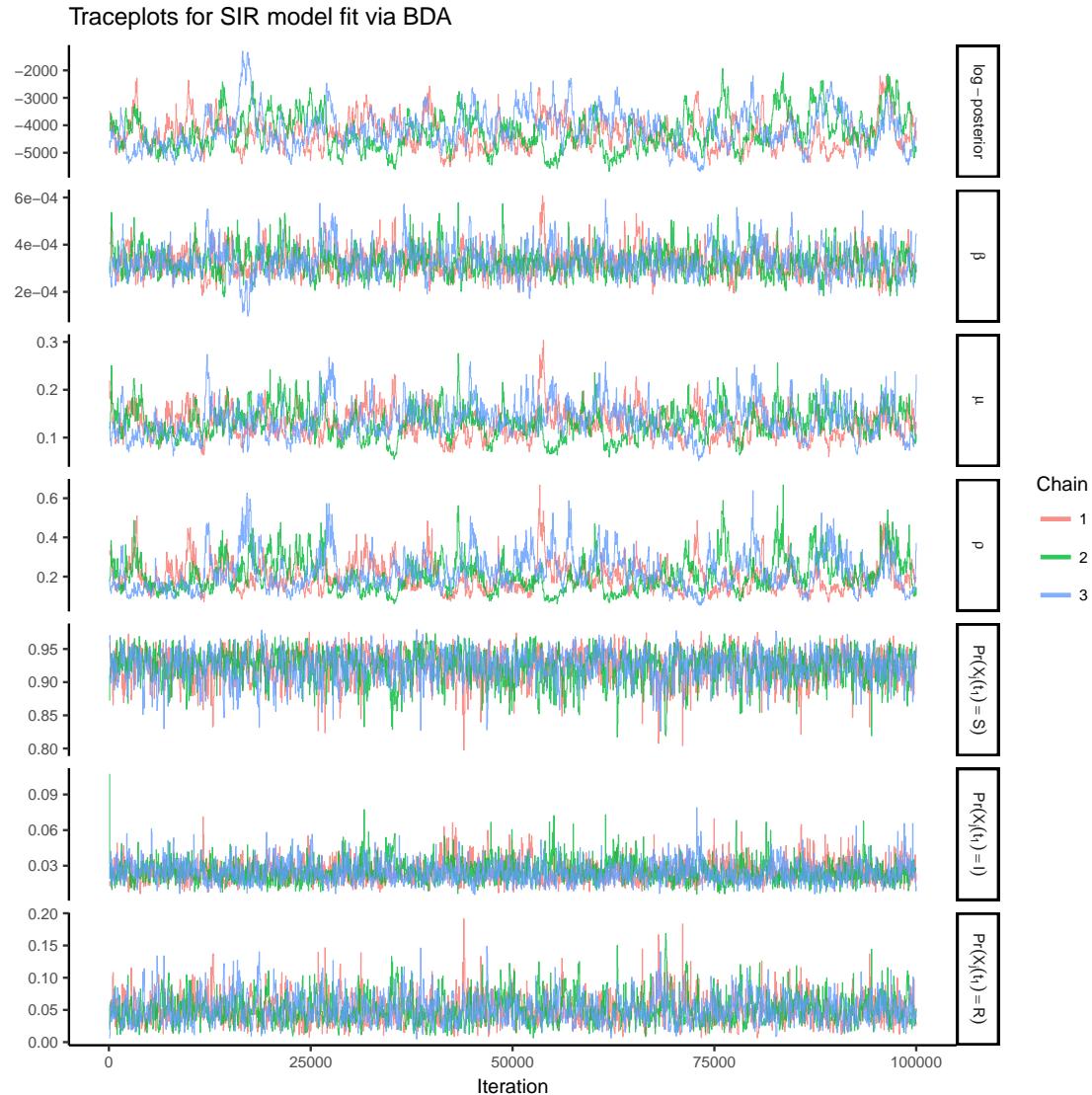


Figure A.1: Traceplots of the log-posterior and model parameters for the SIR model fit using Bayesian data augmentation following an initial burn-in of 10 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

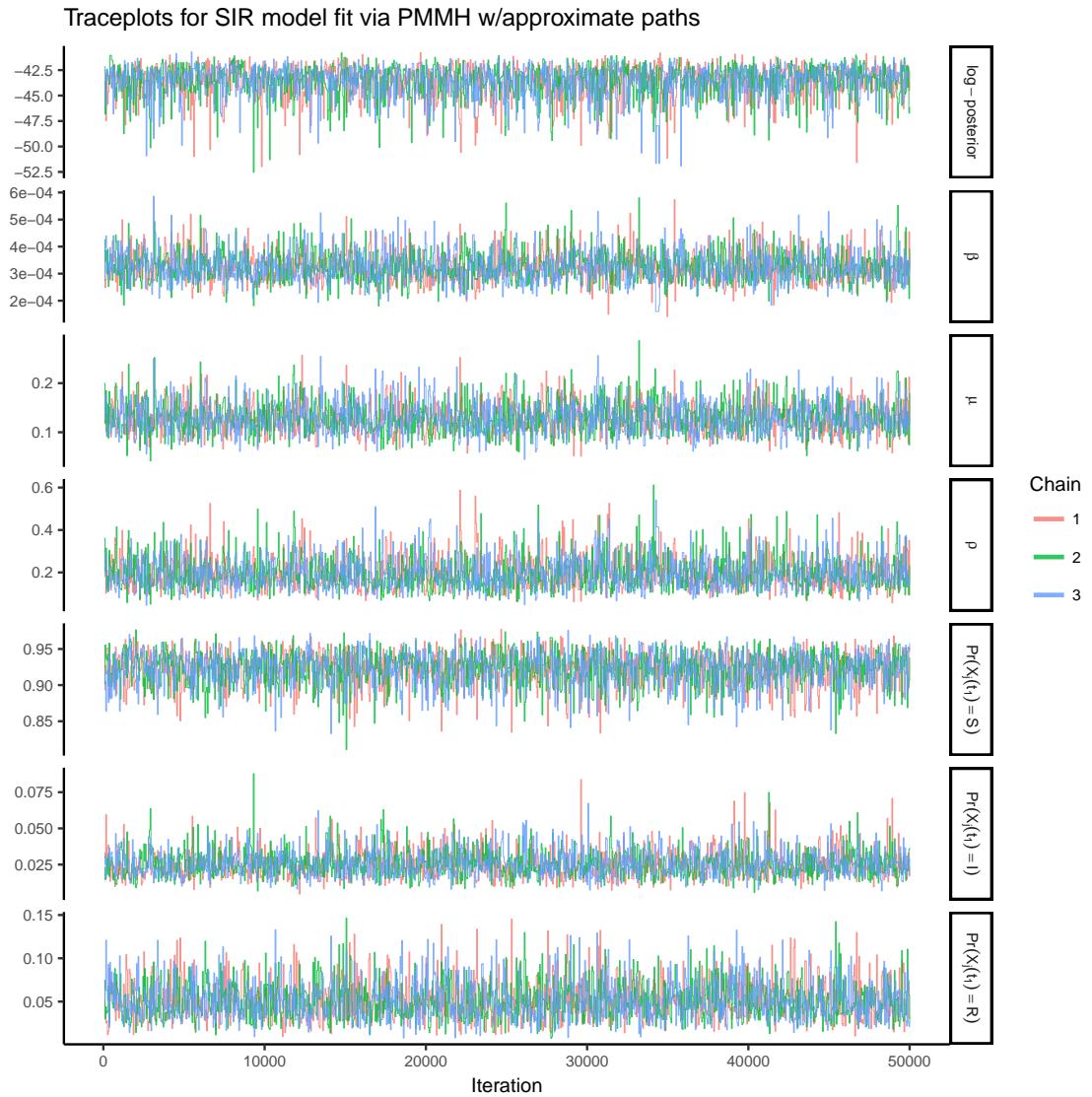


Figure A.2: Traceplots of the log-posterior and model parameters for the SIR model fit using PMMH with 100 particles and a time step of 8 hours, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH and an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

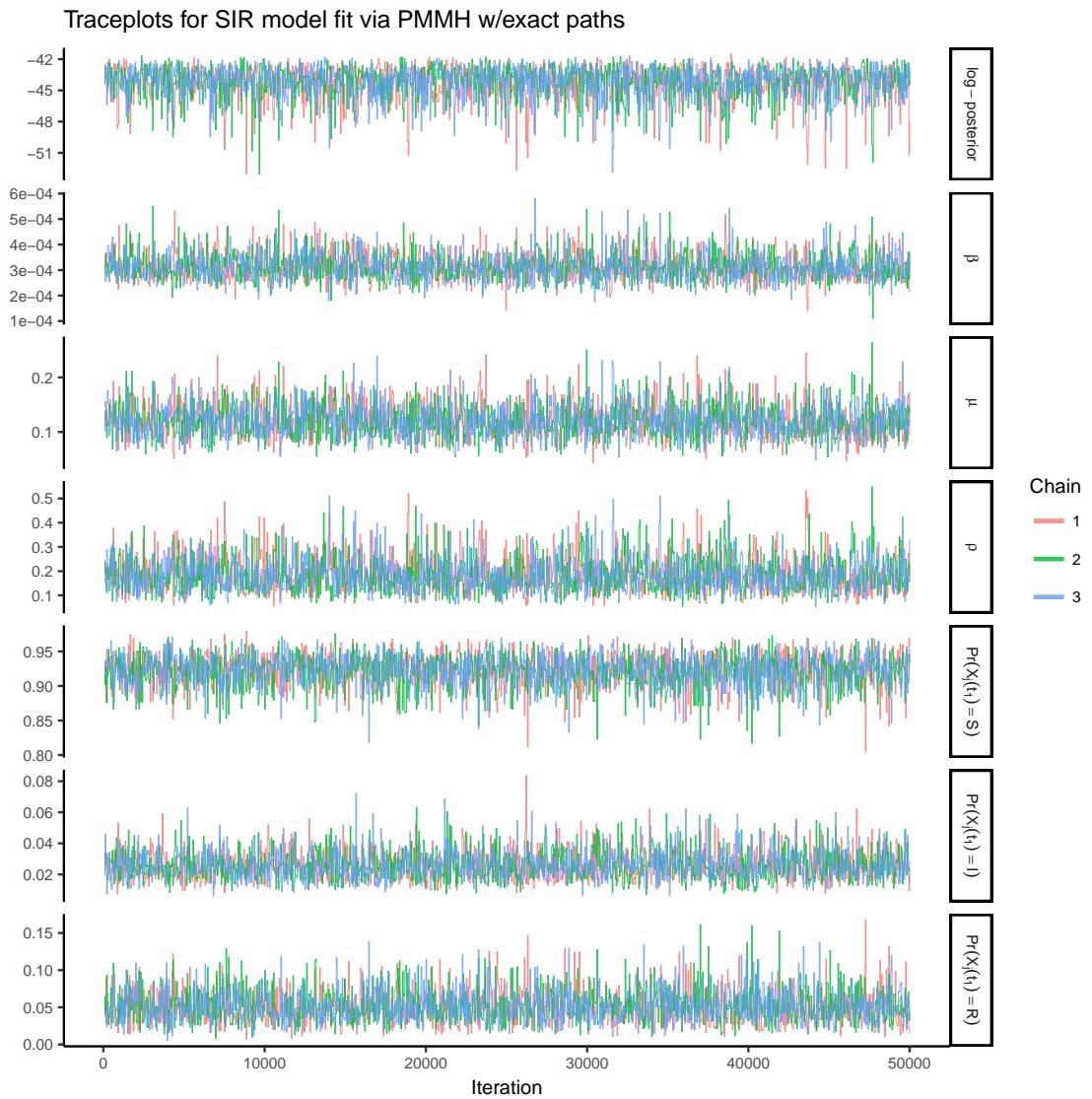


Figure A.3: Traceplots of the log–posterior and model parameters for the SIR model fit using PMMH with 100 particles, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH and an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

A.3.3 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 A.6) 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)
β	0.000075	Gamma(1, 10000)
γ	0.071	Gamma(1, 11)
μ	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)
ρ	0.3	Beta(3.5, 6.5)

Table A.6: Prior distributions for SEIR model and measurement process parameters. The prior for R_0 is the induced prior implied by β and μ . The per-contact infectivity rate is β , the rate at which an exposed individual becomes infectious is γ , the recovery rate is μ , the binomial sampling probability is ρ , 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 τ -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 β , γ , and μ , a logit transformation to ρ , and a generalized logit transformation to \mathbf{p}_{t_1} .

A.3.4 Additional results and MCMC diagnostics for the SEIR model

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 A.7: Run times, log-posterior 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 τ -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.

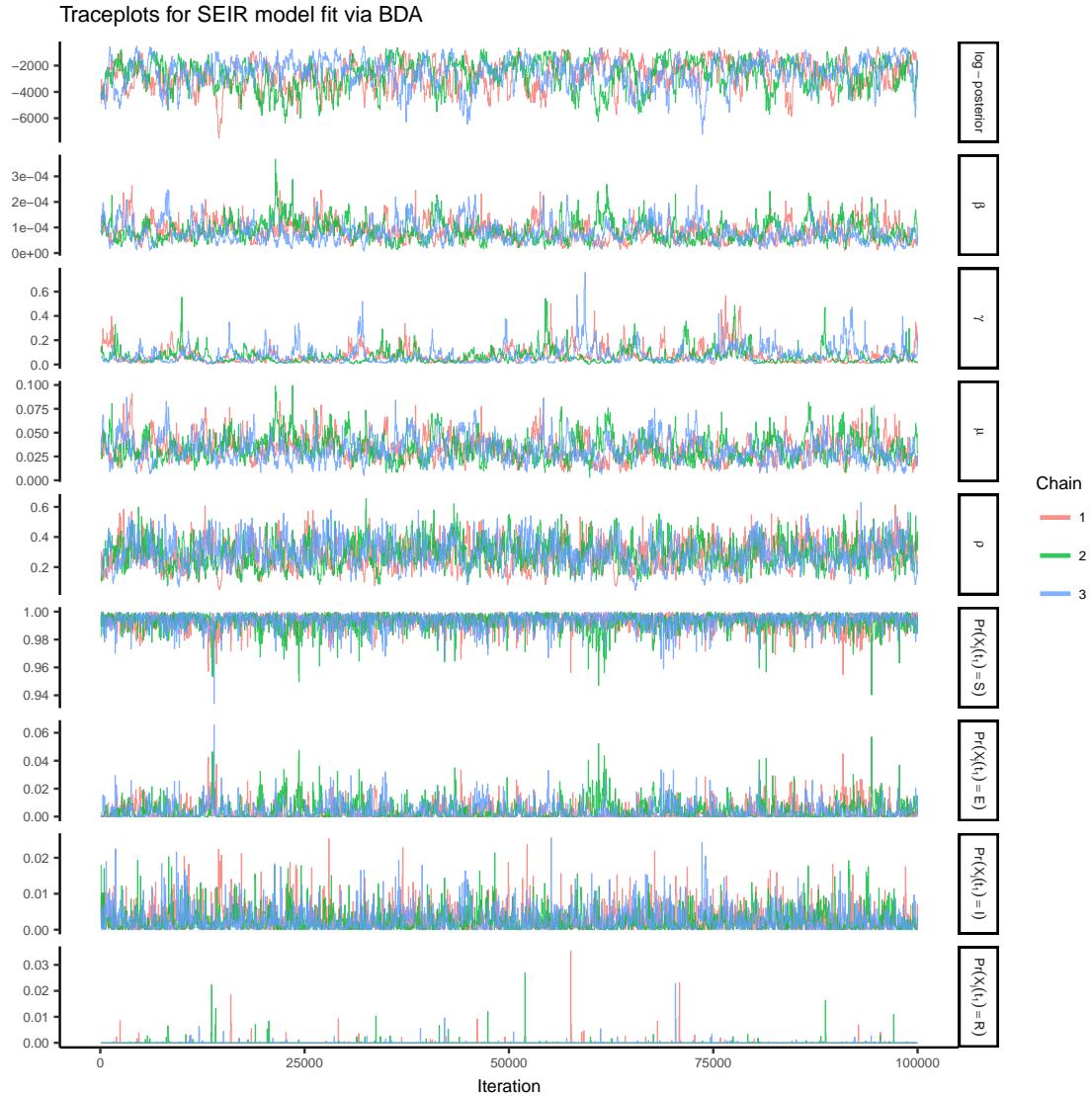


Figure A.4: Traceplots of the log–posterior and model parameters for the SEIR model fit using Bayesian data augmentation following an initial burn–in of 10 iterations. β denotes the per–contact infectivity rate, γ is the rate at which exposed individuals become infectious, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

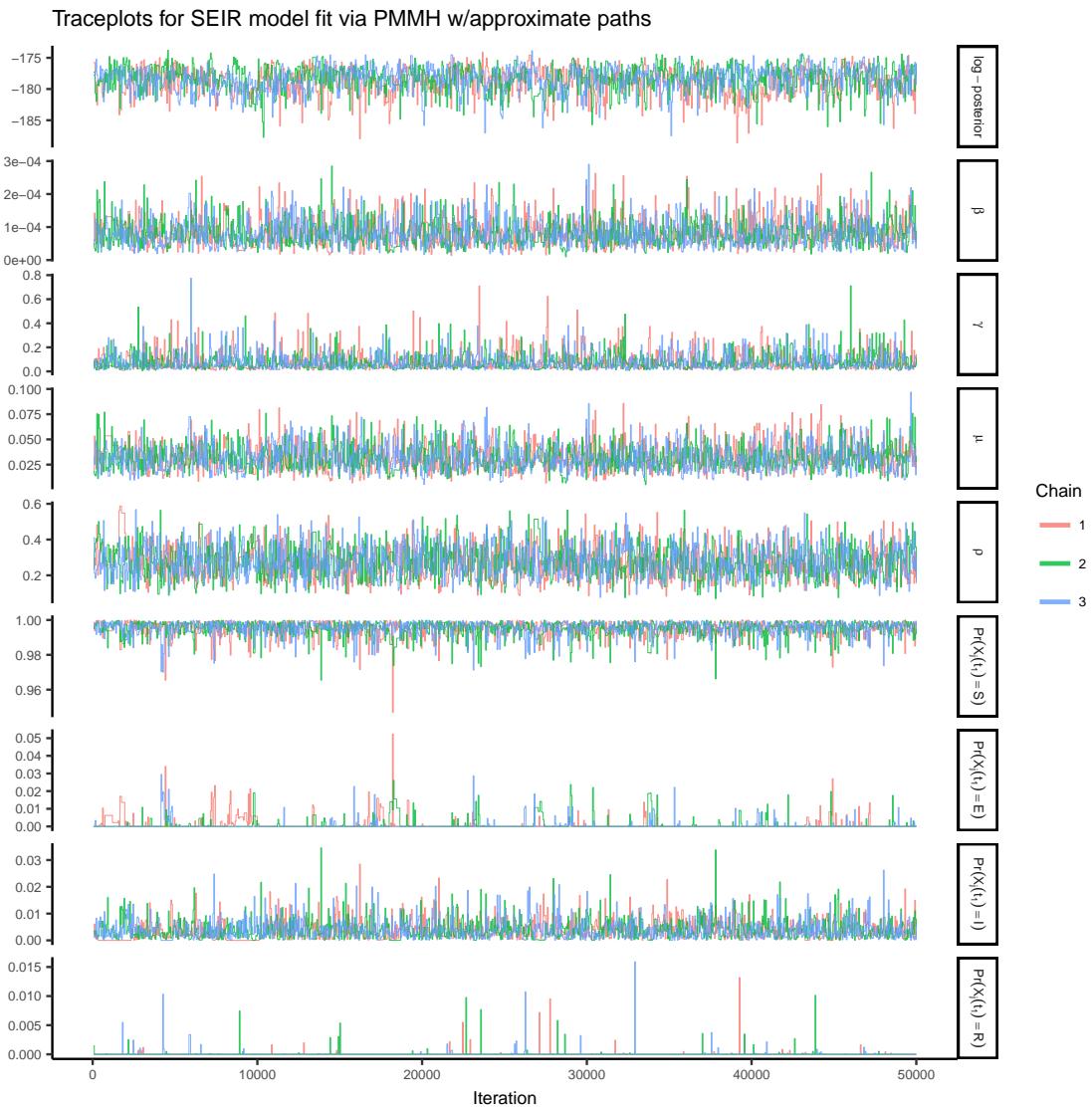


Figure A.5: Traceplots of the log-posterior and model parameters for the SEIR model fit using PMMH with 200 particles and a time step of 8 hours, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH and an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, γ is the rate at which exposed individuals become infectious, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

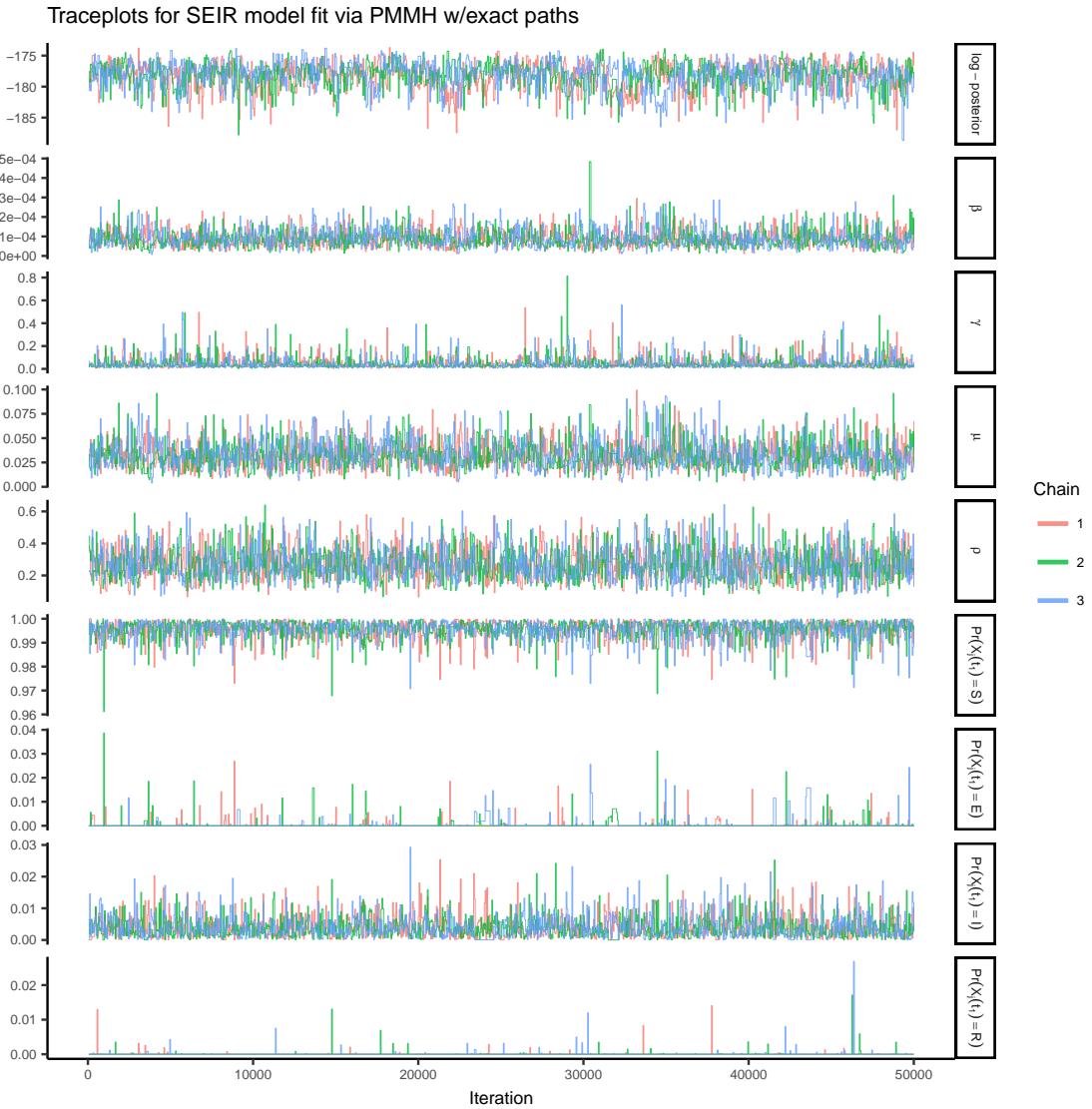


Figure A.6: Traceplots of the log-posterior and model parameters for the SIR model fit using PMMH with 200 particles, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH and an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, γ is the rate at which exposed individuals become infectious, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

A.3.5 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 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 A.8) 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.

Param.	True Value	Prior distribution
R_0	2.52	Beta'(0.1, 1.5, 1, 28)
β	0.1	Gamma(0.1, 100)
μ	0.036	Gamma(1.8, 14)
γ	0.071	Gamma(0.0625, 10)
\mathbf{p}_{t_1}	(0.99, 0.01, 0)	Dirichlet(90, 1.5, 0.01)
ρ	0.95	Beta(5, 1)

Table A.8: Prior distributions for SIRS model and measurement process parameters. The prior for R_0 is the induced prior implied by β and μ . The per-contact infectivity rate is β , the recovery rate is μ , the rate at which immunity is lost is γ , the binomial sampling probability is ρ , and the initial state probabilities are \mathbf{p}_{t_1} .

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 τ -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 β , μ , and γ , a logit transformation to ρ , and a generalized logit transformation to \mathbf{p}_{t_1} .

A.3.6 Additional results and MCMC diagnostics for the SIRS model

A.3.7 Estimated latent posterior distributions for all models

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 A.9: Run times, log-posterior 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 τ –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.

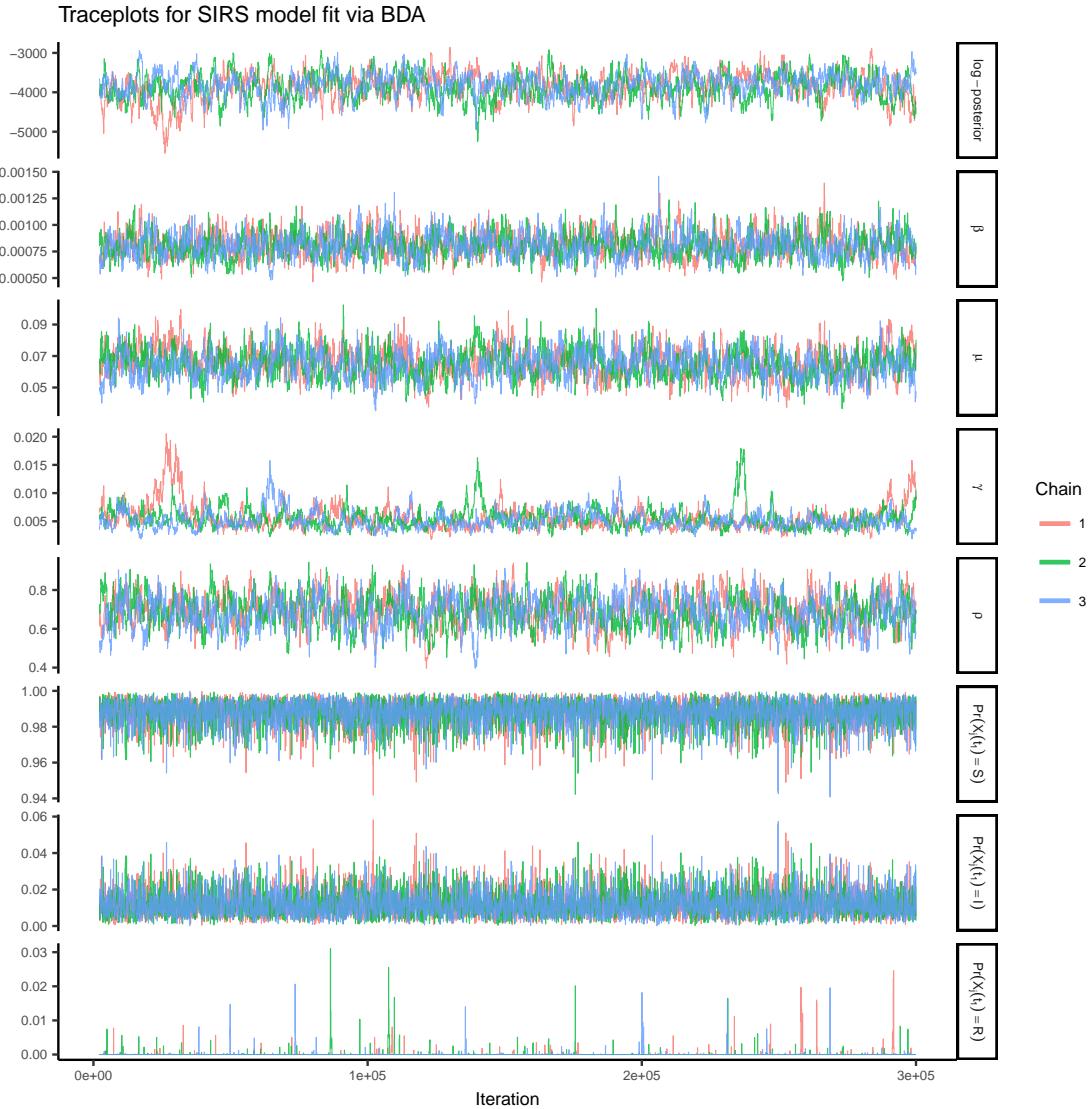


Figure A.7: Traceplots of the log-posterior and model parameters for the SIR model fit using Bayesian data augmentation following an initial burn-in of 2,000 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, γ is the rate at which immunity is lost, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

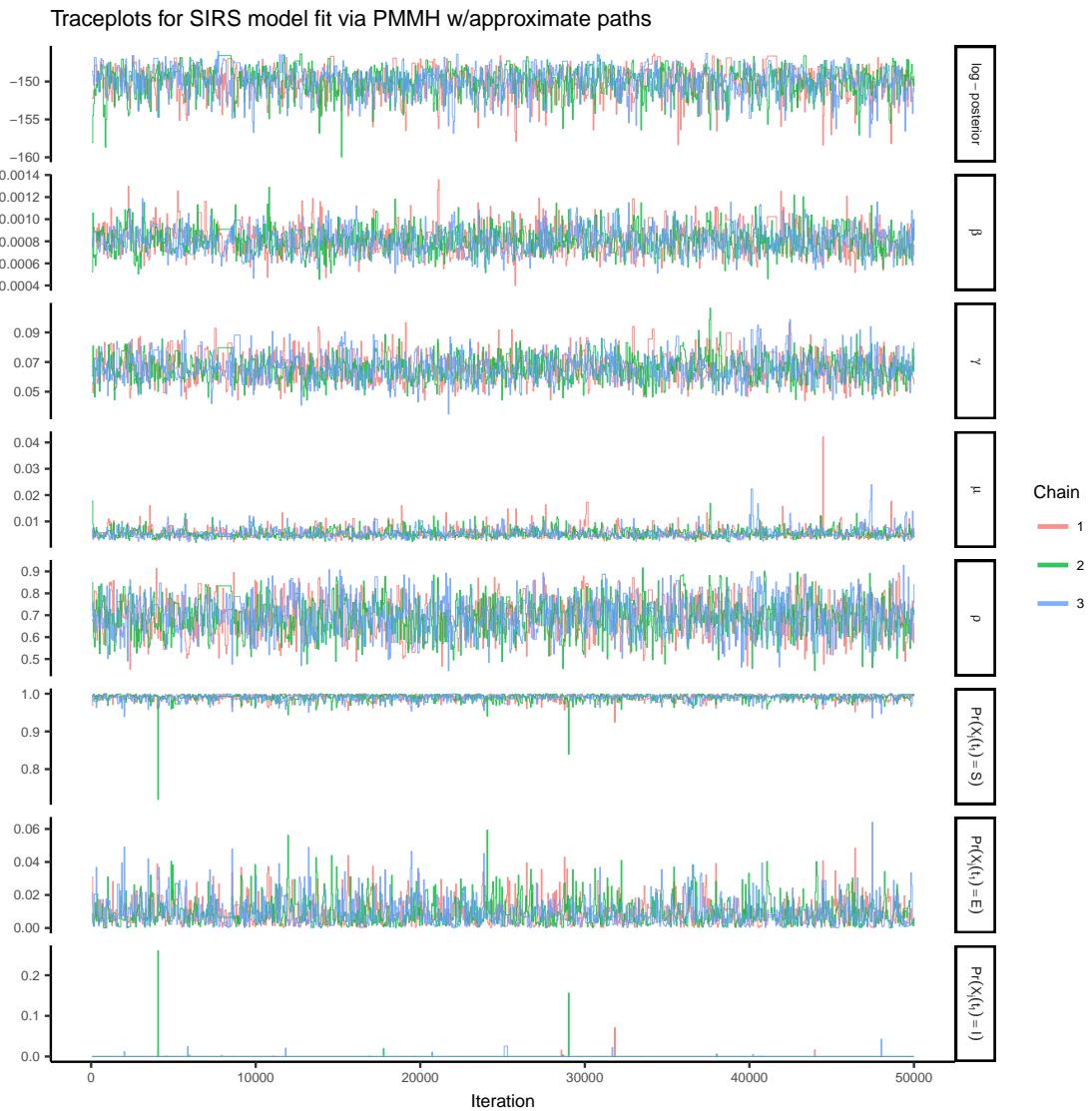


Figure A.8: Traceplots of the log-posterior and model parameters for the SIR model fit using PMMH with 500 particles per chain and a time-step of 8 hours in the approximate τ -leaping algorithm, following a tuning run of 5,000 iterations to estimate the RWMH covariance matrix and in initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, γ is the rate at which immunity is lost, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

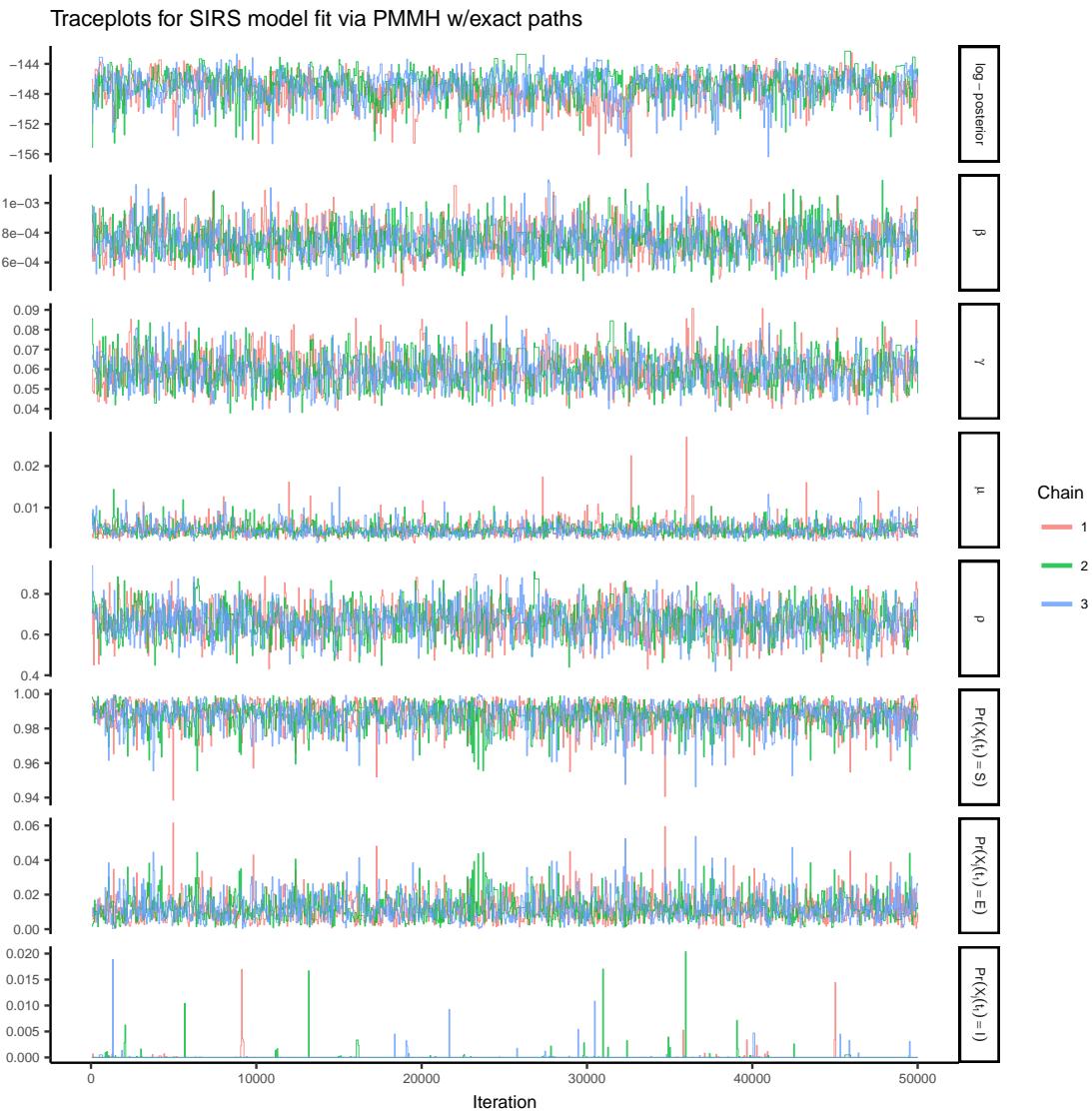


Figure A.9: Traceplots of the log-posterior and model parameters for the SIR model fit using PMMH with 500 particles per chain and particle paths simulated exactly via Gillespie's direct algorithm, following a tuning run of 5,000 iterations to estimate the RWMH covariance matrix and an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, γ is the rate at which immunity is lost, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

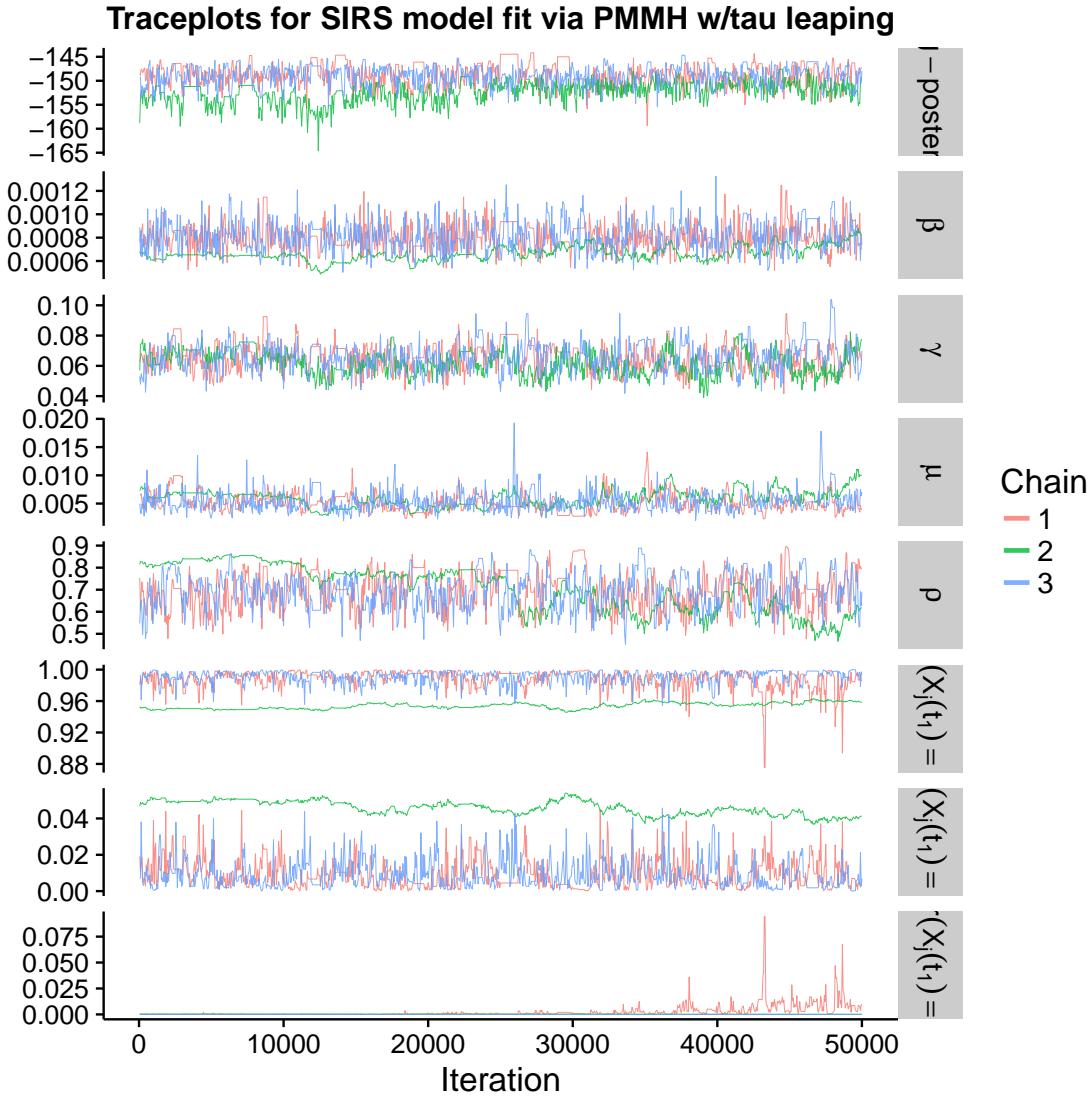


Figure A.10: Traceplots of the log-posterior and model parameters for the SIR model fit using PMMH with 200 particles per chain and a time-step of 8 hours in the approximate τ -leaping algorithm, following a tuning run of 5,000 iterations to estimate the RWMH covariance matrix and an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, γ is the rate at which immunity is lost, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

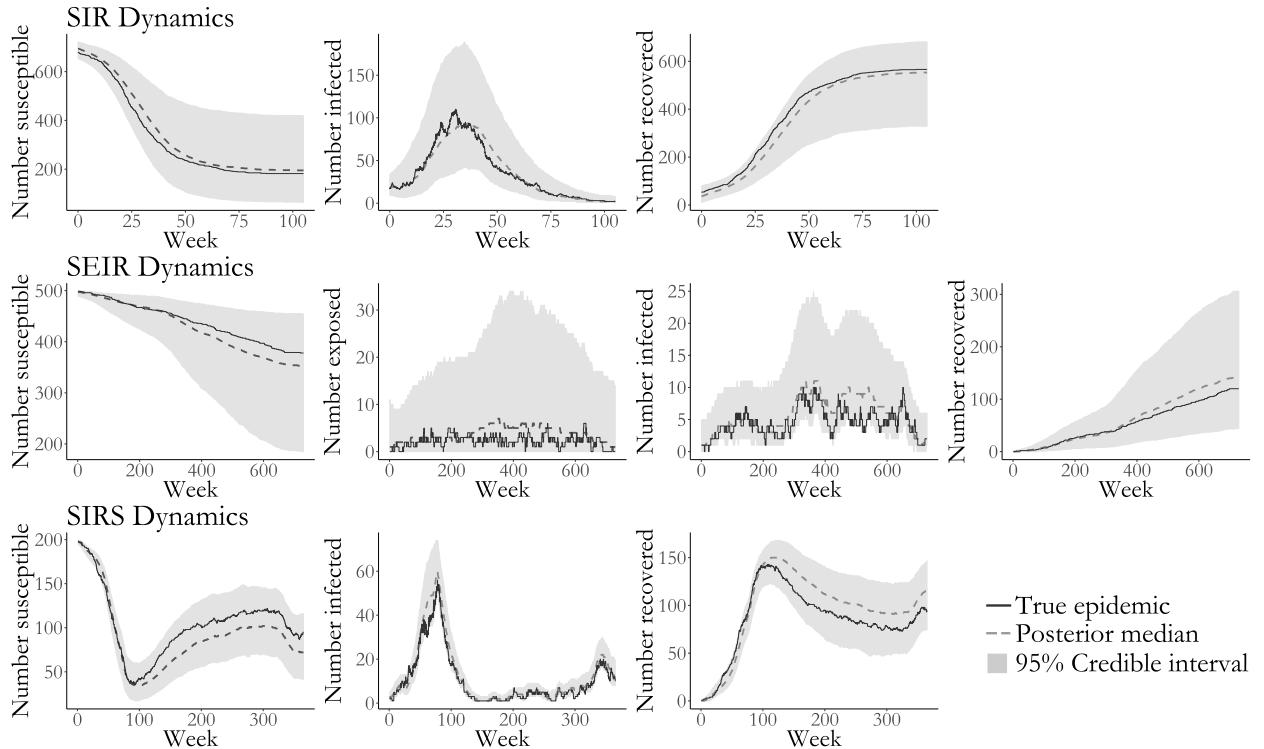


Figure A.11: Pointwise posterior medians (dashed lines) and pointwise 95% credible intervals for the numbers of individuals in each disease state for the SIR, SEIR, and SIRS models. True compartment counts are shown as solid lines. Estimates are based on a thinned sample, retaining the collection of disease histories at the end of every 250th MCMC iteration.

A.4 Simulation 2 — Inference under Model Misspecification — Setup and Additional Results

A.4.1 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 A.10. 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 τ -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
β	0.00025
γ	1/210
μ	1/150
ρ	0.95
$\mathbf{X}(t_0)$	$S_0 = 397, E_0 = 2, I_0 = 1, R_0 = 0$

Epoch 2: Weeks 26–105

β	0.0001
γ	1/210
μ	1/330
ρ	0.95
$\mathbf{X}(t_{26})$	$S_0 = 279, E_0 = 98, I_0 = 20, R_0 = 3$

Epoch 3: Weeks 105–167

β	0.00035
γ	1/90
μ	1/300
ρ	0.95
$\mathbf{X}(t_{105})$	$S_0 = 1, E_0 = 43, I_0 = 145, R_0 = 211$

Epoch 4: Weeks 167 – 209

β	0.0001
γ	1/180
μ	1/70
ρ	0.95
$\mathbf{X}(t_{167})$	$S_0 = 0, E_0 = 1, I_0 = 52, R_0 = 347$

Table A.10: 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 be-

SIR model

Parameter	Prior distribution
R_0	Beta'(0.6, 0.7, 1, 4)
β	Gamma(0.6, 10000)
μ	Gamma(0.7, 100)
\mathbf{p}_{t_1}	Dirichlet(90, 0.5, 0.01)
ρ	Beta(10, 1)

SEIR model

Parameter	Prior distribution
R_0	Beta'(0.6, 0.7, 1, 4)
β	Gamma(0.6, 10000)
γ	Gamma(0.5, 100)
μ	Gamma(0.7, 100)
\mathbf{p}_{t_1}	Dirichlet(90, 0.5, 0.5, 0.01)
ρ	Beta(10, 1)

Table A.11: 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 β and μ . The per-contact infectivity rate is β , the rate at which an exposed individual becomes infectious is γ , the recovery rate is μ , the binomial sampling probability is ρ , and the initial state probabilities are \mathbf{p}_{t_1} .

A.4.2 Additional results

SIR model

Parameter	Posterior median (95% Credible interval)
R_0	4.05 (3.40, 4.81)
β	0.000035 (0.000030, 0.000040)
μ	0.0034 (0.0031, 0.0038)
ρ	0.95 (0.93, 0.97)

SEIR model

Parameter	Posterior median (95% Credible interval)
R_0	23.80 (15.10, 36.98)
β	0.00021 (0.00013, 0.00032)
γ	0.0047 (0.0038, 0.0061)
μ	0.0035 (0.0032, 0.0038)
ρ	0.95 (0.94, 0.97)

Table A.12: 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.

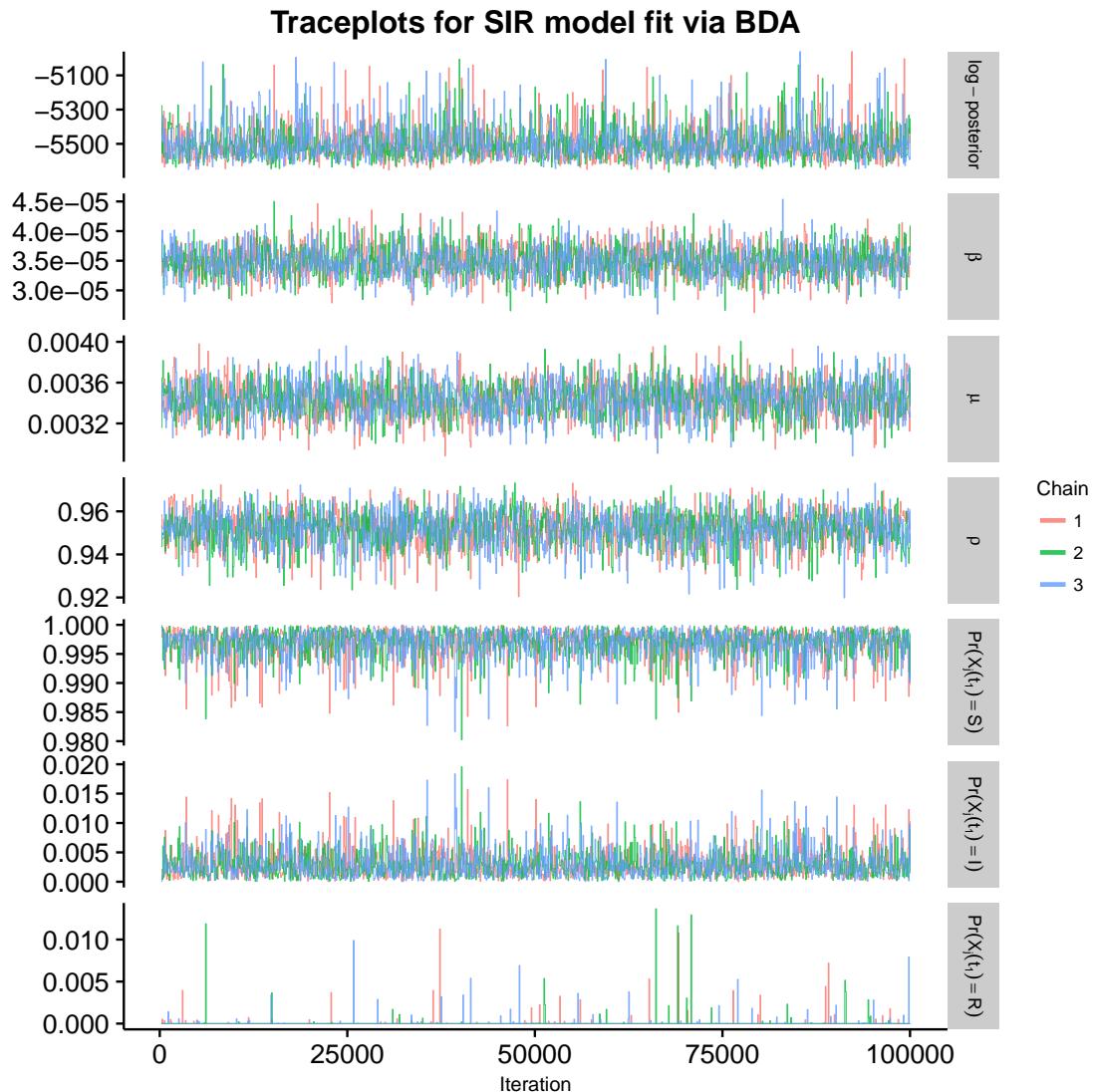


Figure A.12: Traceplots of the log-posterior and model parameters for the SIR model fit using BDA following an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

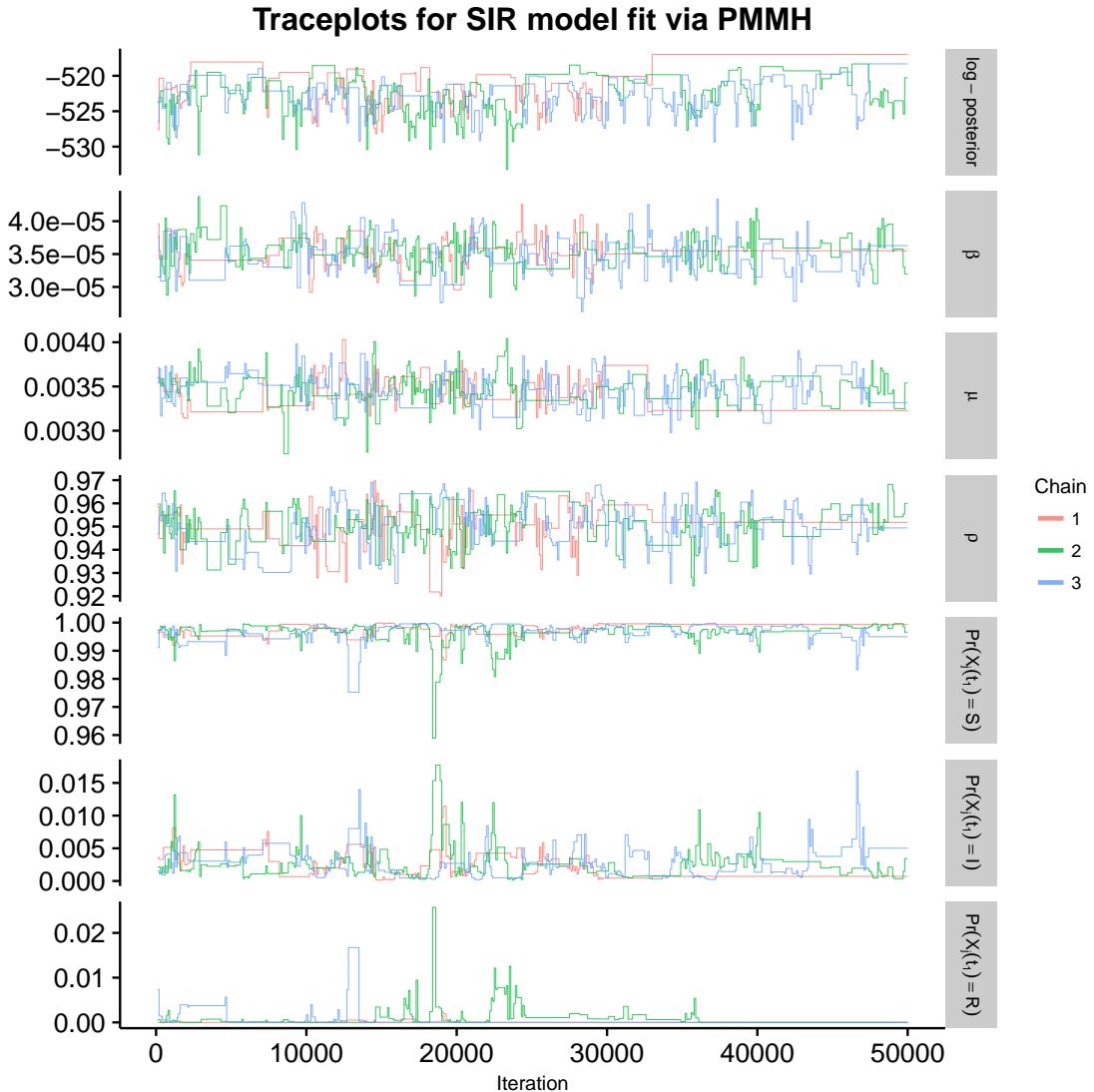


Figure A.13: Traceplots of the log-posterior and model parameters for the SIR model fit using PMMH with 2,500 particles, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH. β denotes the per-contact infectivity rate, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

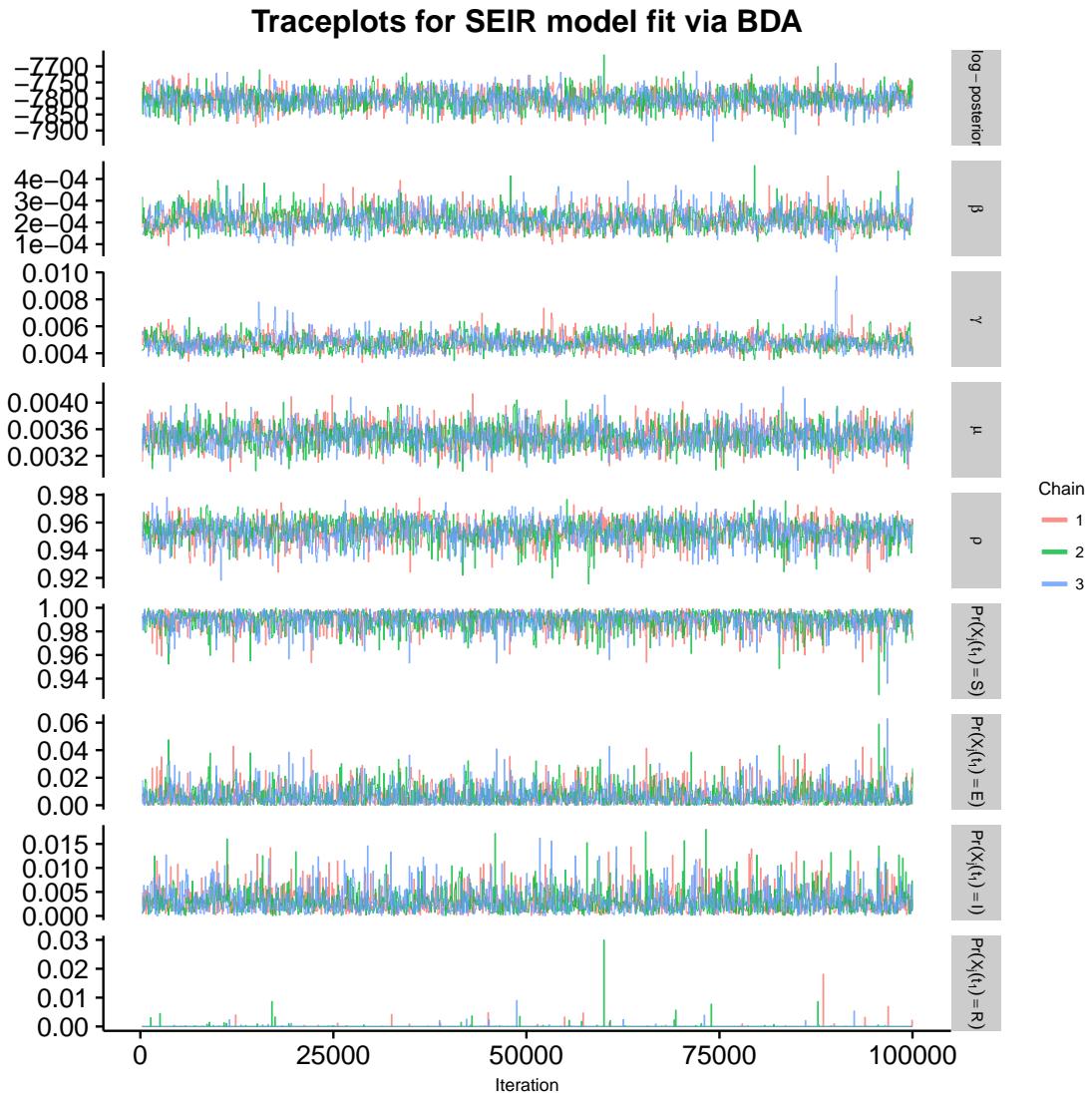


Figure A.14: Traceplots of the log-posterior and model parameters for the SEIR model fit using BDA following an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, γ is the rate at which an exposed individual becomes infectious, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

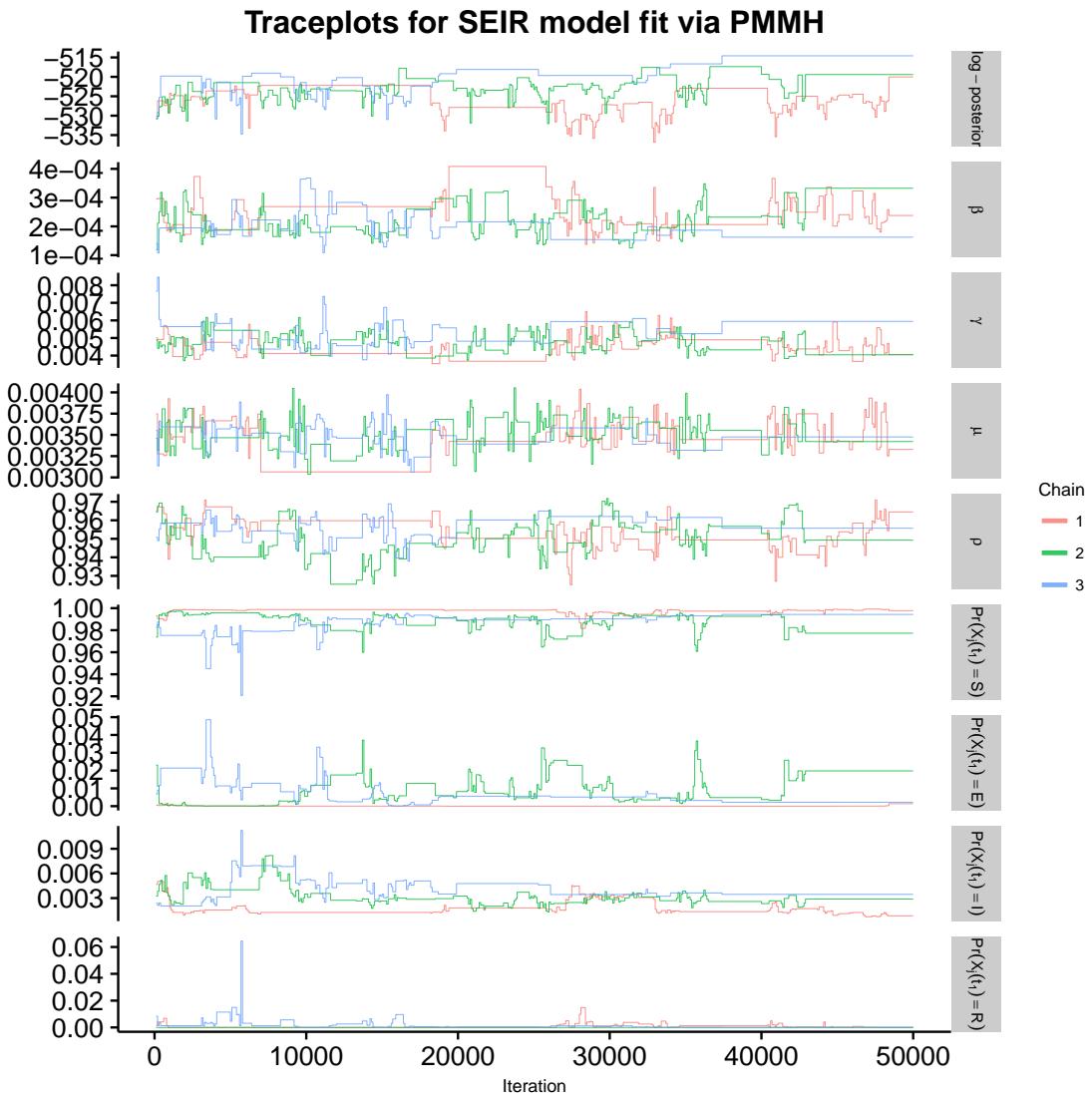


Figure A.15: Traceplots of the log-posterior and model parameters for the SEIR model fit using PMMH with 2,500 particles, following a tuning run of 5,000 iterations used to estimate the covariance matrix for the RWMH. β denotes the per-contact infectivity rate, γ is the rate at which an exposed individual becomes infectious, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

A.5 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 susceptible. 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 A.13).

Param.	Prior distribution
R_0	Beta'(0.00042 $\times \frac{1250}{N}$, 0.35, 1, 2 / N)
β	Gamma(0.00042 $\times \frac{1250}{N}$, 1)
μ	Gamma(0.35, 2)
\mathbf{p}_{t_1}	Dirichlet(100, 1, 5)
ρ	Beta(1,1)

Table A.13: Prior distributions for SIR model and measurement process parameters. The prior for R_0 is the induced prior implied by β and μ . The per-contact infectivity rate is β , the recovery rate is μ , the binomial sampling probability is ρ , and the initial state probabilities are \mathbf{p}_{t_1} . The prior for β was scaled in accordance with the assumed population size.

A.6 Simulation 4 — Effect of Prior Specification on Inference — Setup and Additional Results

A.6.1 Simulation details

We ran three MCMC chains for each of the SIR models fit under the prior regimes that are specified in Table A.14 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.

Parameter	Prior Distribution			
	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 A.14: 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 β and μ . 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 3.8×10^{-4} and 2.7×10^4 . In regimes one and two, 80% of the prior mass for ρ lay between 0.17 and 0.27, while in regimes three and four the prior mass for ρ 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.

A.6.2 Convergence diagnostics

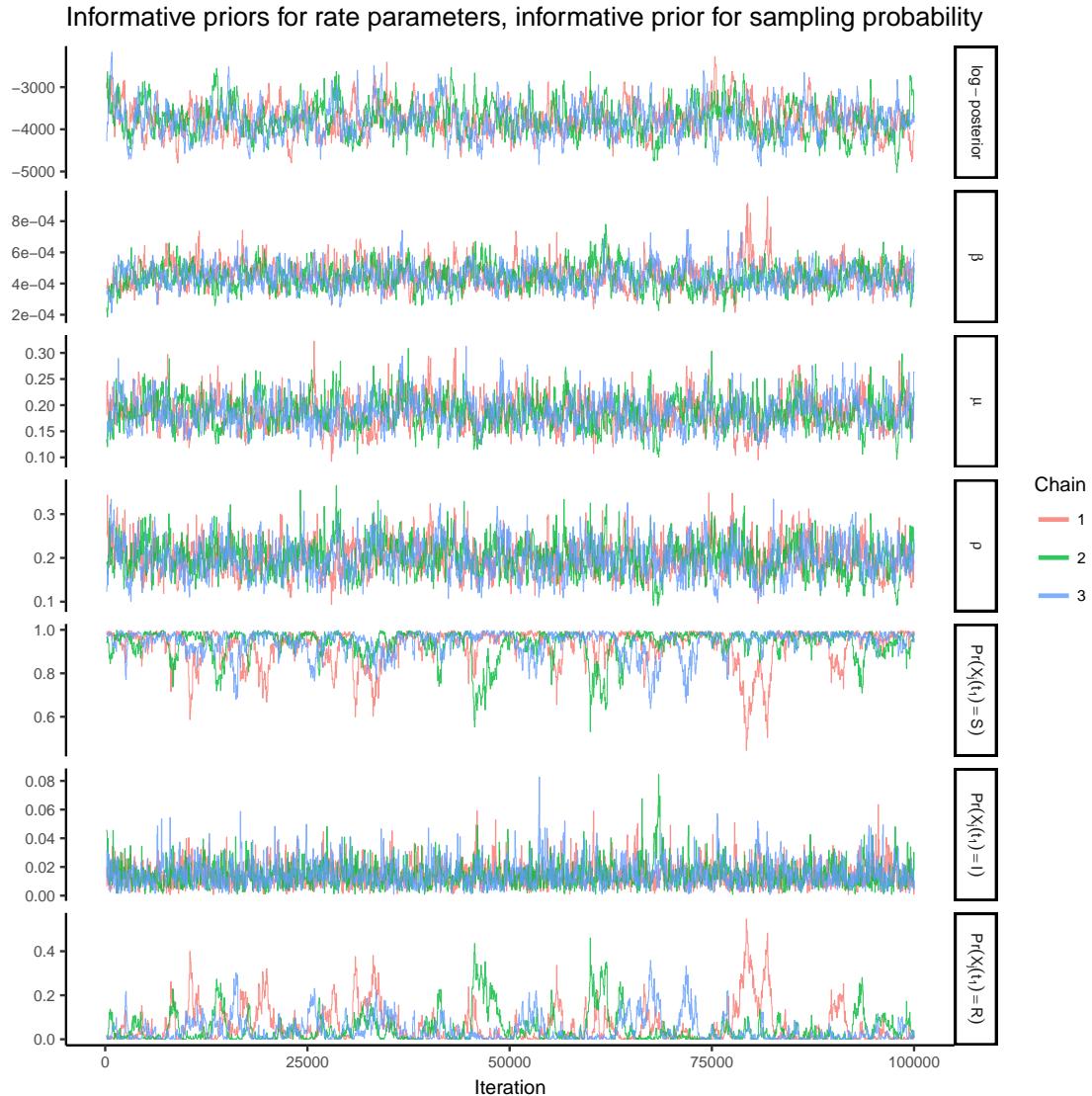


Figure A.16: Traceplots of the log-posterior and model parameters for the SIR model fit under informative priors for all model parameters. β denotes the per-contact infectivity rate, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

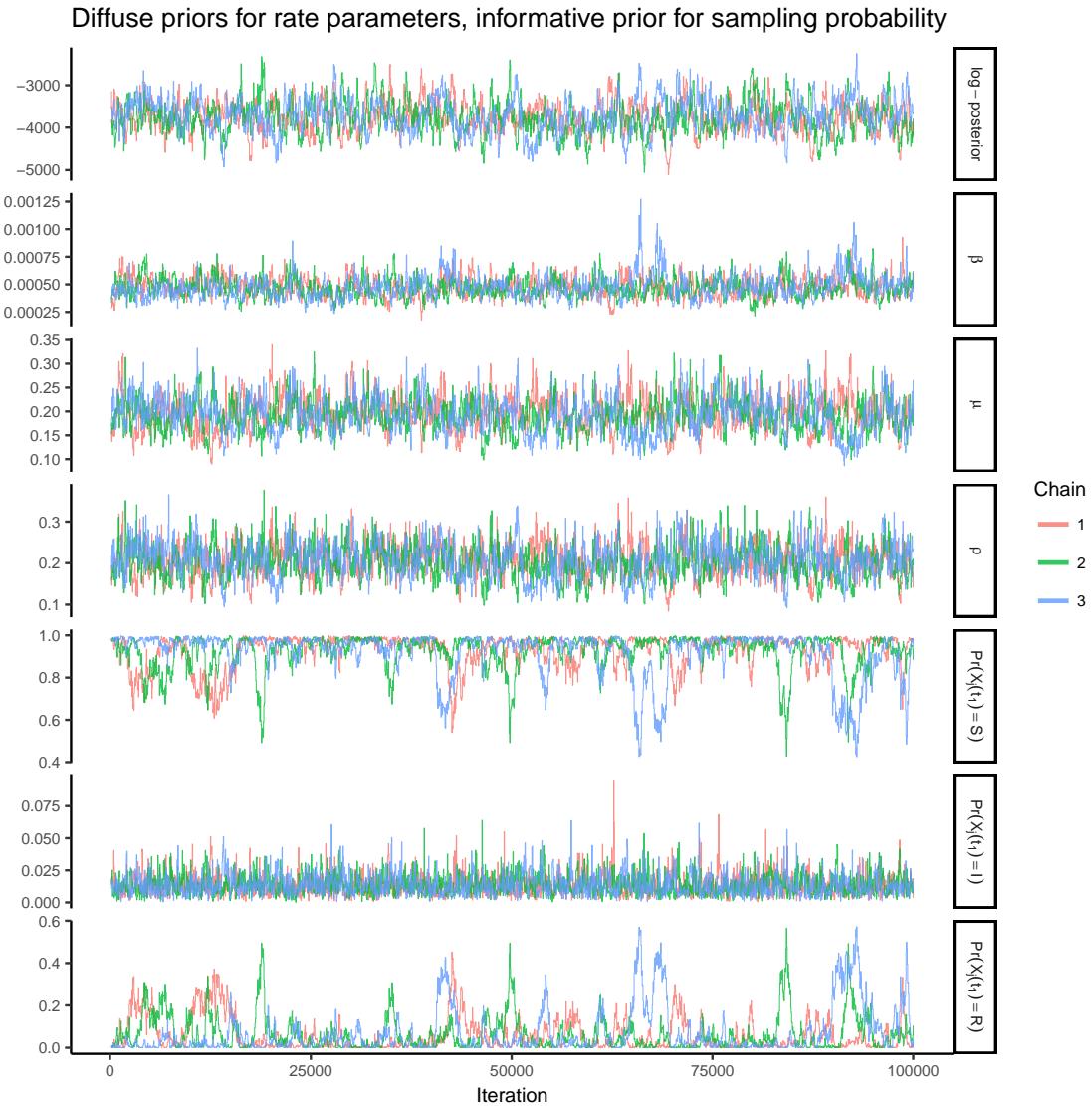


Figure A.17: Traceplots of the log–posterior and model parameters for the SIR model fit under diffuse priors for the rate parameters and an informative prior for the binomial sampling probability. β denotes the per–contact infectivity rate, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

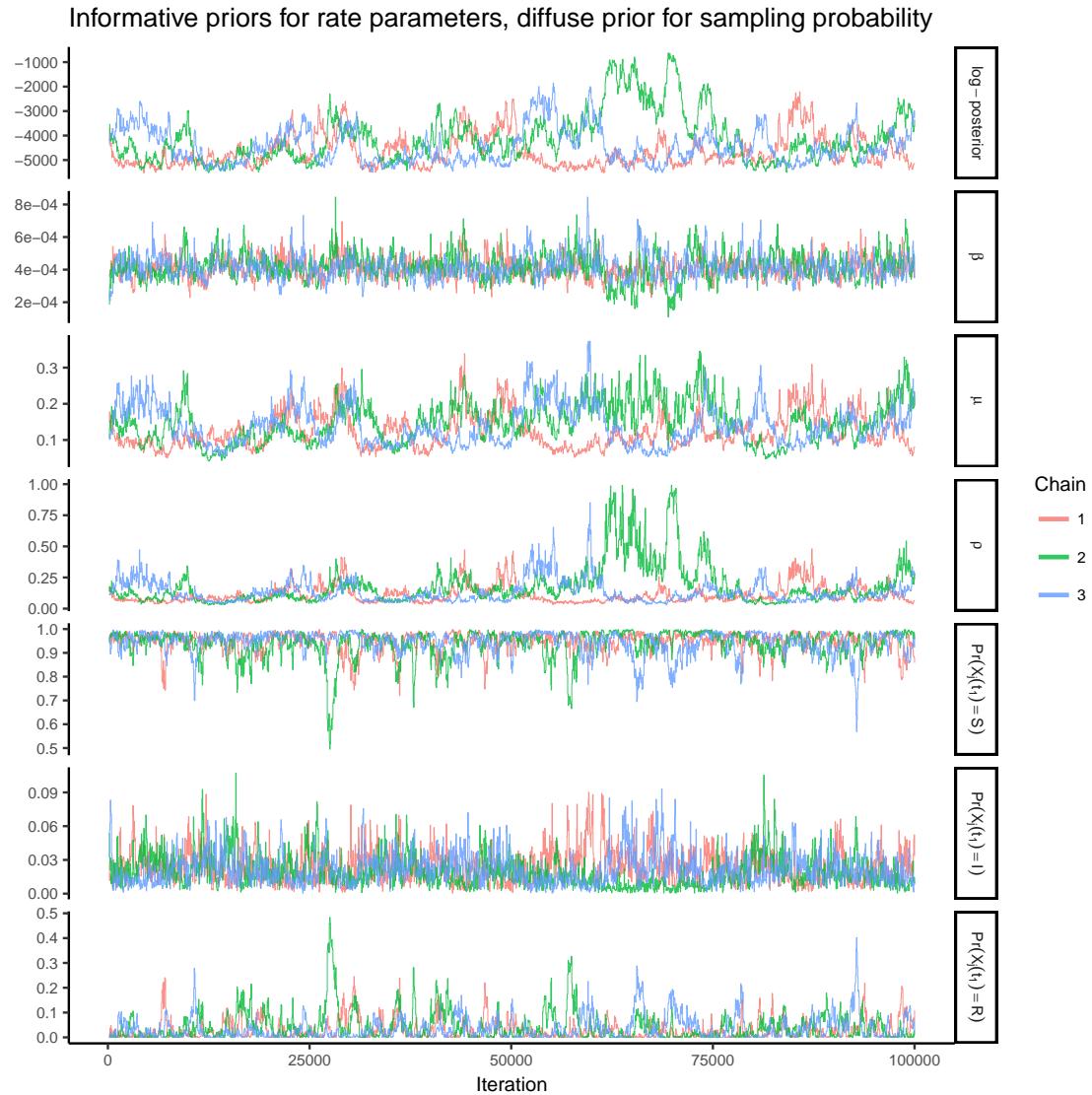


Figure A.18: Traceplots of the log-posterior and model parameters for the SIR model fit under informative priors for the rate parameters and a diffuse prior for the binomial sampling probability. β denotes the per-contact infectivity rate, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

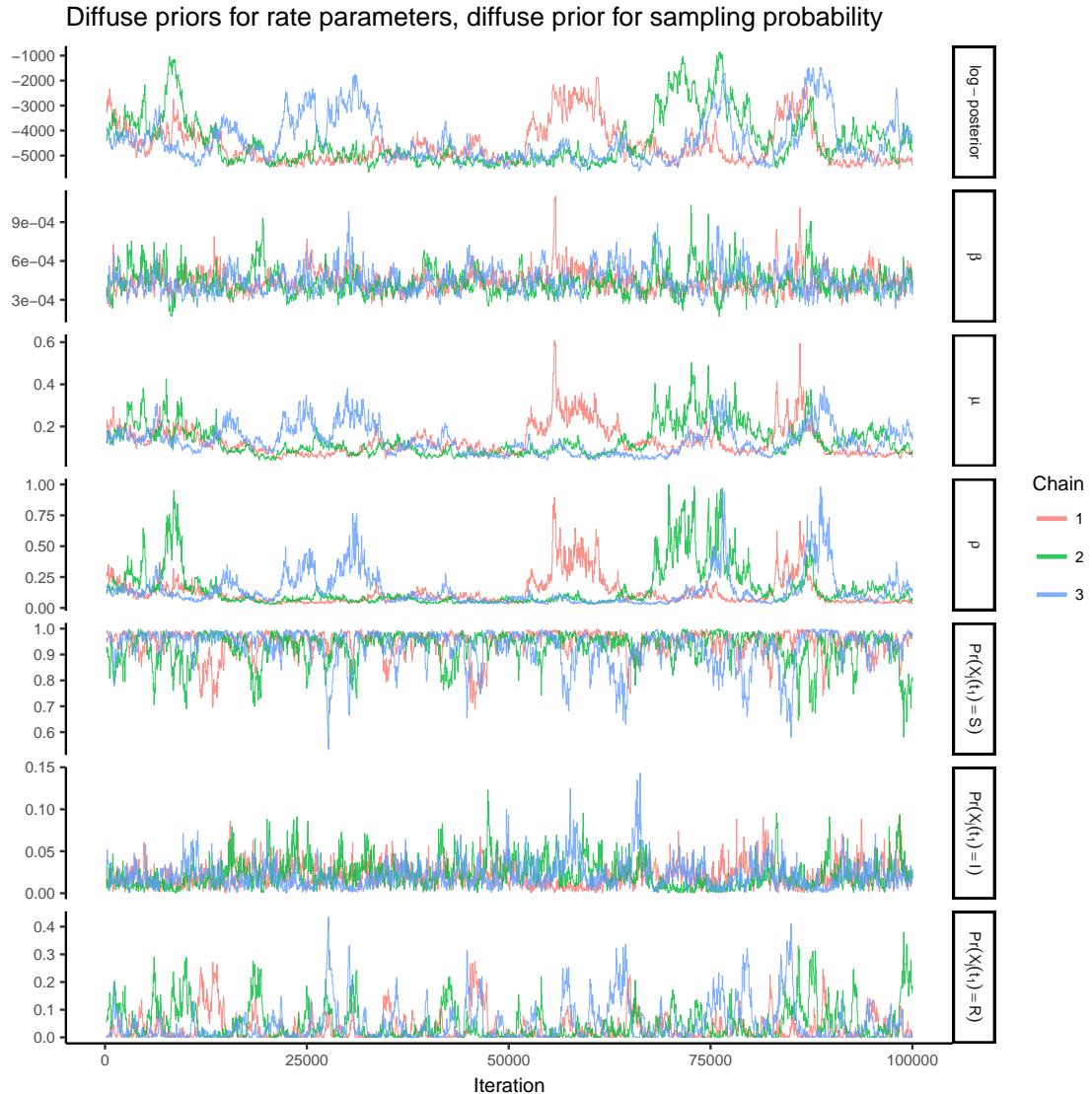


Figure A.19: Traceplots of the log-posterior and model parameters for the SIR model fit under diffuse priors for all model parameters. β denotes the per-contact infectivity rate, μ is the recovery rate, ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

A.7 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 A.15 and A.16. 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 τ -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).

A.7.1 Boarding school example — MCMC diagnostics

SIR model fit to boarding school data via BDA

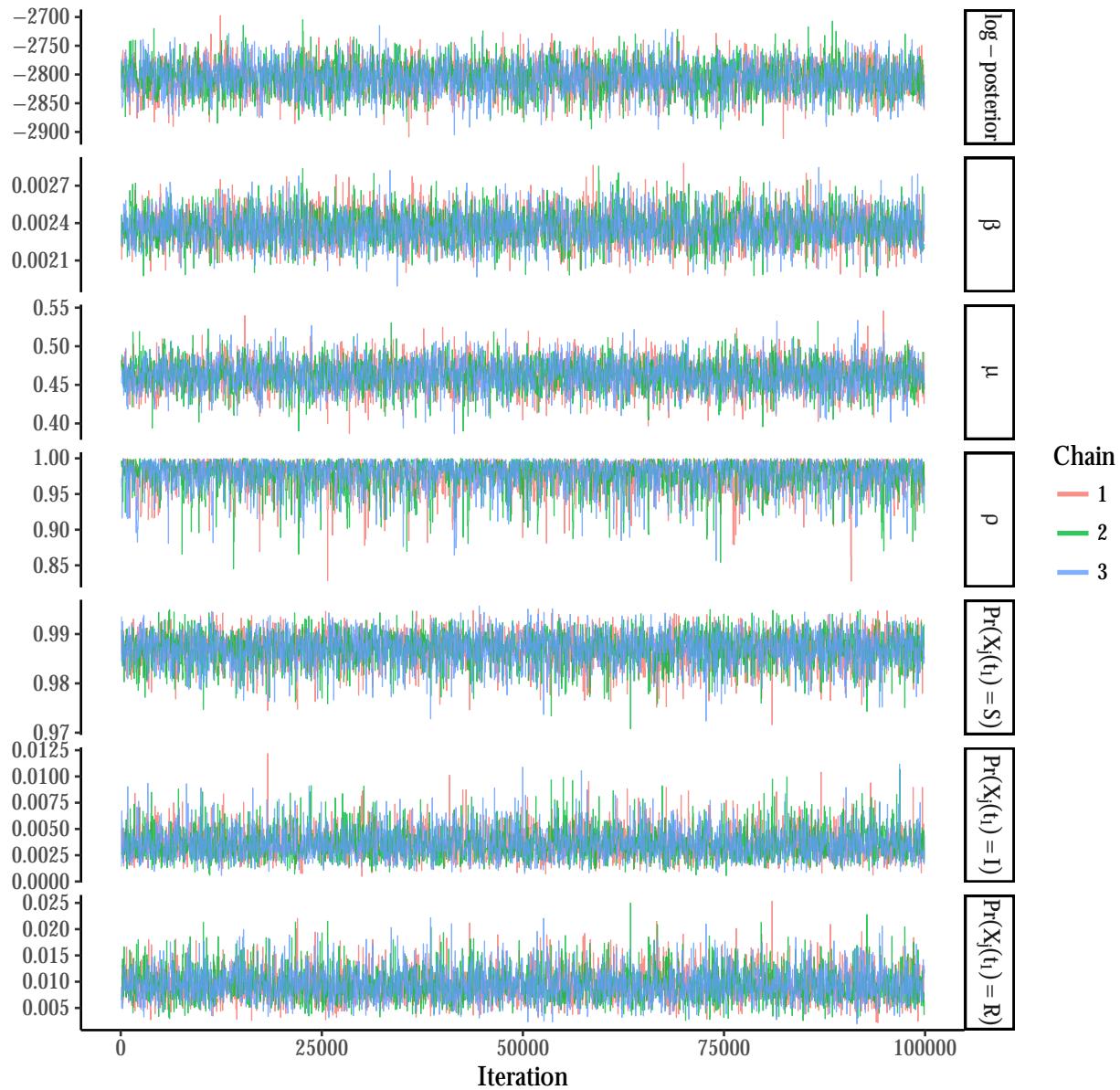


Figure A.20: Traceplots of the log-posterior and model parameters for the SIR model fit under binomial emissions using BDA following an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

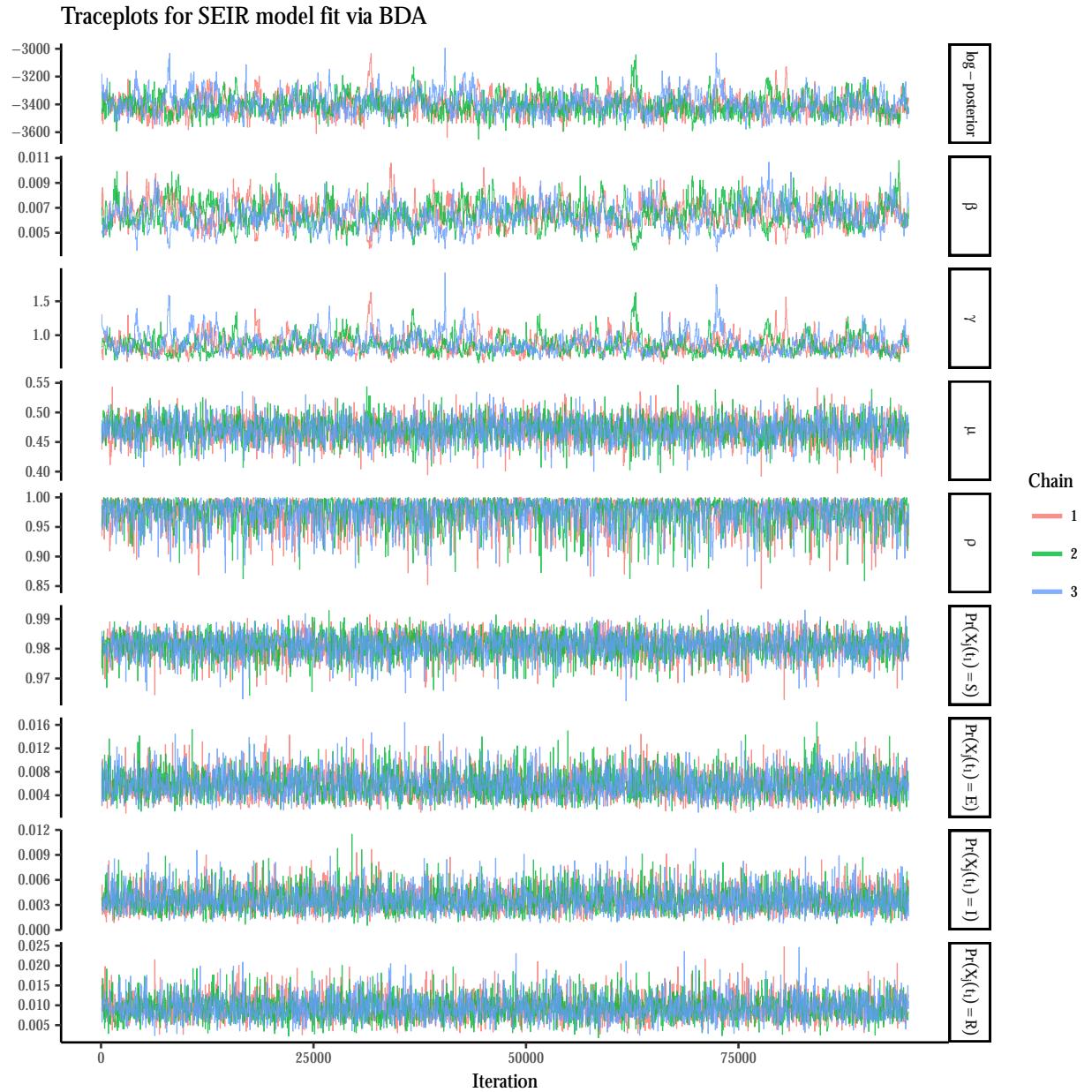


Figure A.21: Traceplots of the log–posterior and model parameters for the SEIR model fit under binomial emissions via BDA following an initial burn–in of 5,000 iterations. β denotes the per–contact infectivity rate, μ is the recovery rate, γ is the rate at which immunity is lost, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

Traceplots for SIR model fit via PMMH w/tau leaping

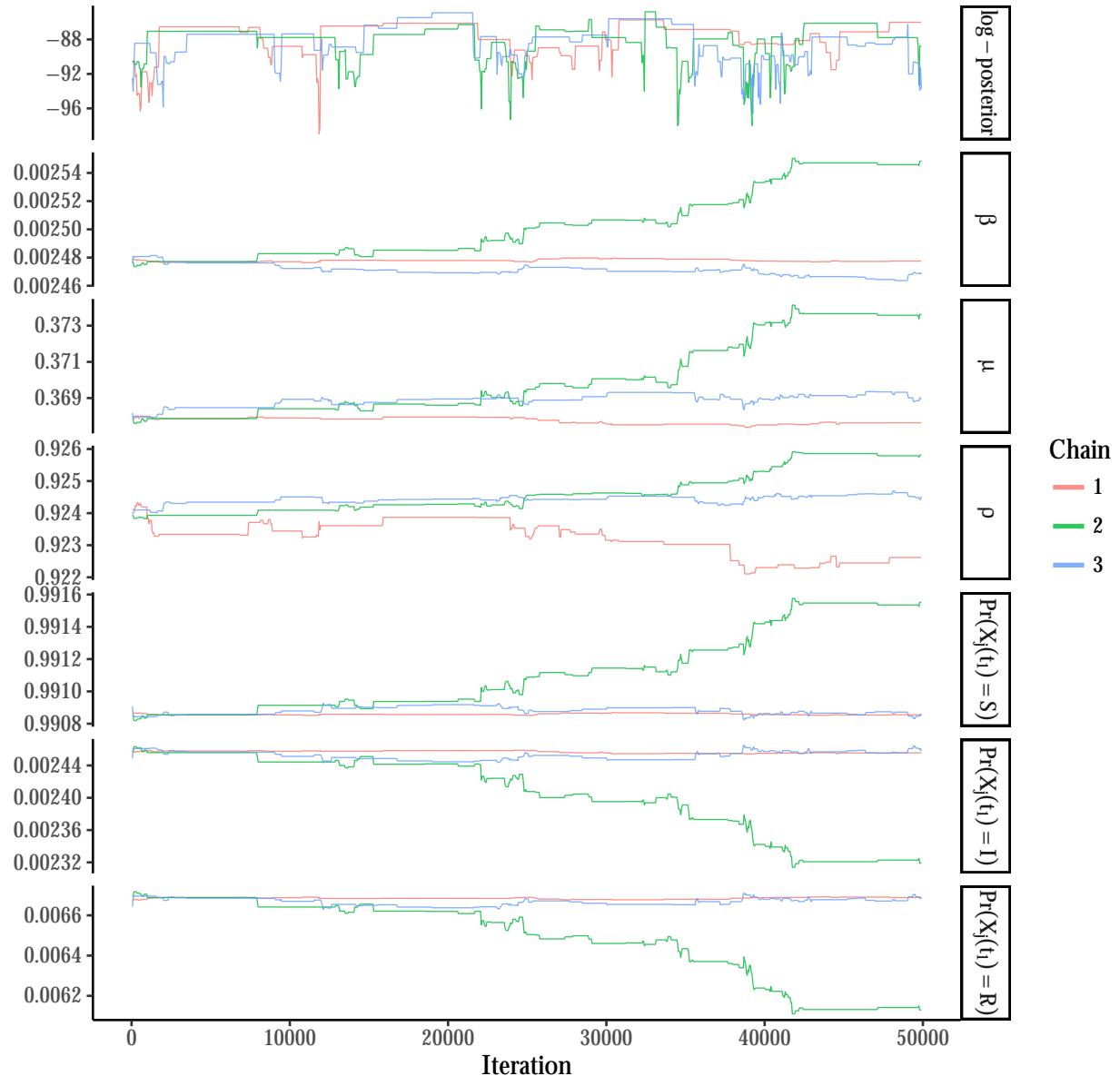


Figure A.22: Traceplots of the log-posterior and model parameters for the SIR model fit under binomial emissions using PMMH with 5,000 particles per chain and a time-step of 2 hours in the approximate τ -leaping algorithm, following a tuning run of 2,000 iterations to estimate the RWMH covariance matrix and in initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

Traceplots for SEIR model fit via PMMH w/tau leaping

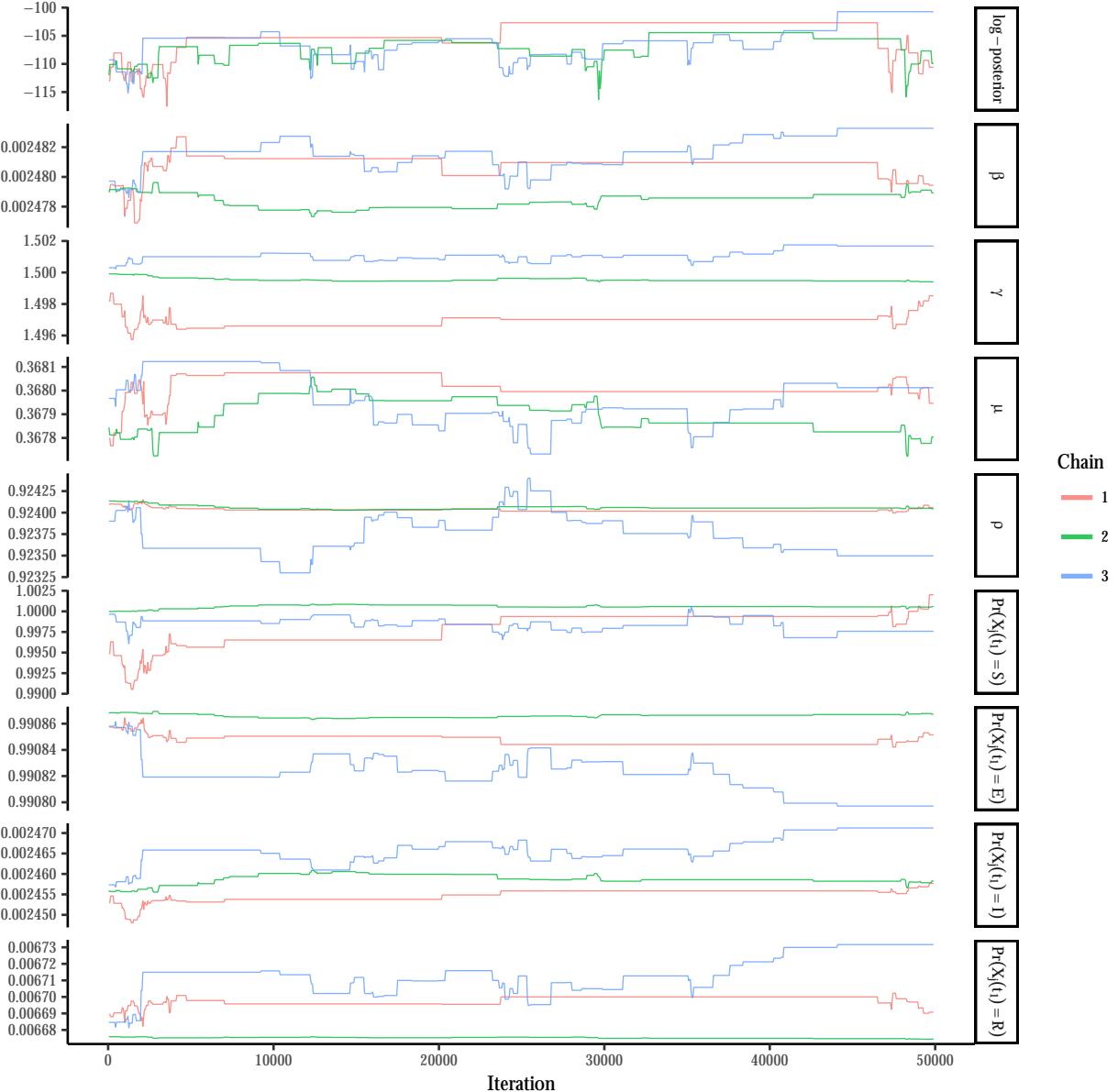


Figure A.23: Traceplots of the log-posterior and model parameters for the SEIR model fit under binomial emissions using PMMH with 5,000 particles per chain and a time-step of 2 hours in the approximate τ -leaping algorithm, following a tuning run of 2,000 iterations to estimate the RWMH covariance matrix and an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, γ is the rate at which immunity is lost, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

Parameter	Prior Distribution	Posterior Median (95% BCI)
R_0	Beta'(0.001, 1, 1, 1526)	3.89 (3.40, 4.47)
β	Gamma(0.001, 1)	0.0024 (0.0021, 0.0026)
μ	Gamma(1,2)	0.46 (0.42, 0.50)
ρ	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 A.15: 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 β , the recovery rate is μ , and the binomial sampling probability is ρ . The prior for R_0 is the implied prior induced by the priors for β and μ . Effective sample size were β : 11,304; μ : 16,238; ρ : 3,920; $p_{S_{t_1}}$: 26,989; $p_{I_{t_1}}$: 284,431; $p_{R_{t_1}}$: 22,761.

A.7.2 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 A.22 and A.23). 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 A.24). That even a model with this simple set of time-varying dynamics would undoubtedly still be misspecified with respect to the

Parameter	Prior Distribution	Posterior Median (95% BCI)
R_0	Beta'(0.001, 1, 1, 1526)	3.89 (3.40, 4.47)
β	Gamma(0.001, 1)	0.0064 (0.0046, 0.0086)
γ	Gamma(0.001, 1)	0.84 (0.66, 1.19)
μ	Gamma(1,2)	0.47 (0.43, 0.51)
ρ	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 A.16: 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 β , the rate at which an exposed individual becomes infectious is γ , the recovery rate is μ , and the binomial sampling probability is ρ . The prior for R_0 is the implied prior induced by the priors for β and μ . Effective sample size were β : 679; γ : 658; μ : 10,069; ρ : 3,244; $p_{S_{t_1}}$: 26,868; $p_{I_{t_1}}$: 26,168; $p_{R_{t_1}}$: 273,613.

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.

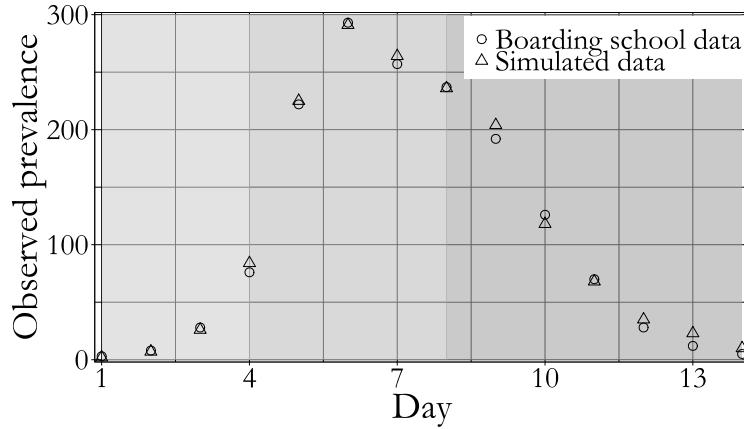


Figure A.24: British boardings school data and data simulated under SEIR dynamics with time-varying dynamics over three epochs (indicated by different shaded regions). The simulated dataset was generated using the following parameters: in the first epoch (days 1-4), $\beta = 0.0035$, $\gamma = 1.25$, $\mu = 0.3$. 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.

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 given in Tables A.15

and A.16, and the negative binomial overdispersion parameter, ϕ , was assigned a $\text{Gamma}(1, 0.1)$ 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 ρ and on the log scale for ϕ . 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 τ –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 A.25), and the credible intervals for the model parameters (Figure A.26). 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 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

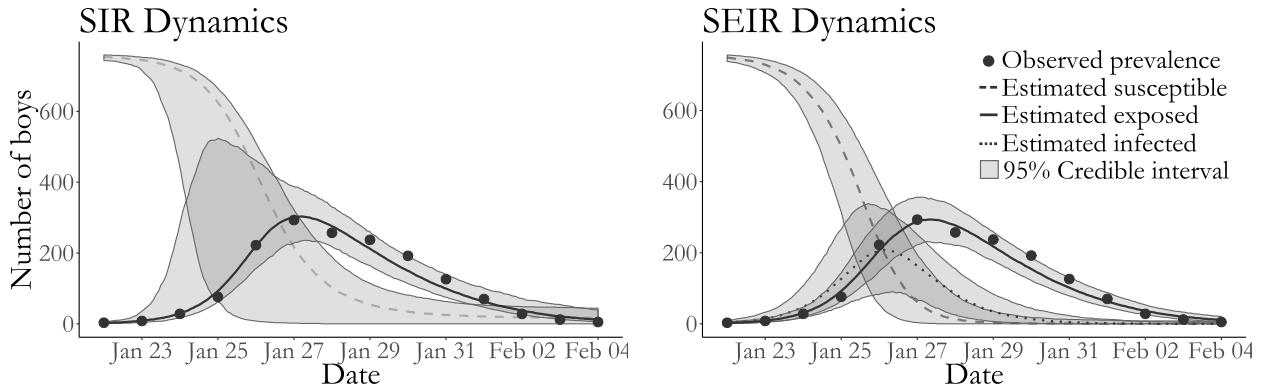


Figure A.25: Boarding school data, pointwise posterior median estimates and pointwise 95% credible intervals under negative binomial emissions (grey shaded areas) for the numbers of infected boys (solid line) and susceptible boys (dashed line). Posterior estimates based on a thinned sample, with every 250th configuration retained.

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 A.28).

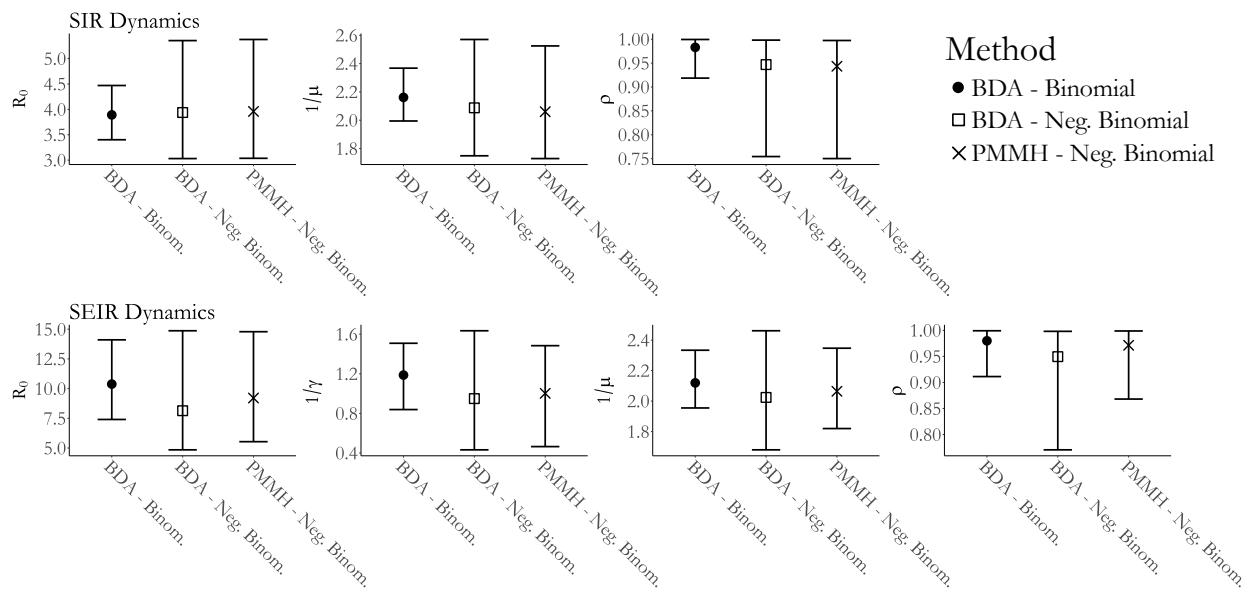


Figure A.26: Posterior medians and 95% credible intervals for SIR and SEIR models fit with BDA and PMMH to the British boarding school data under binomial and negative binomial emission distributions.

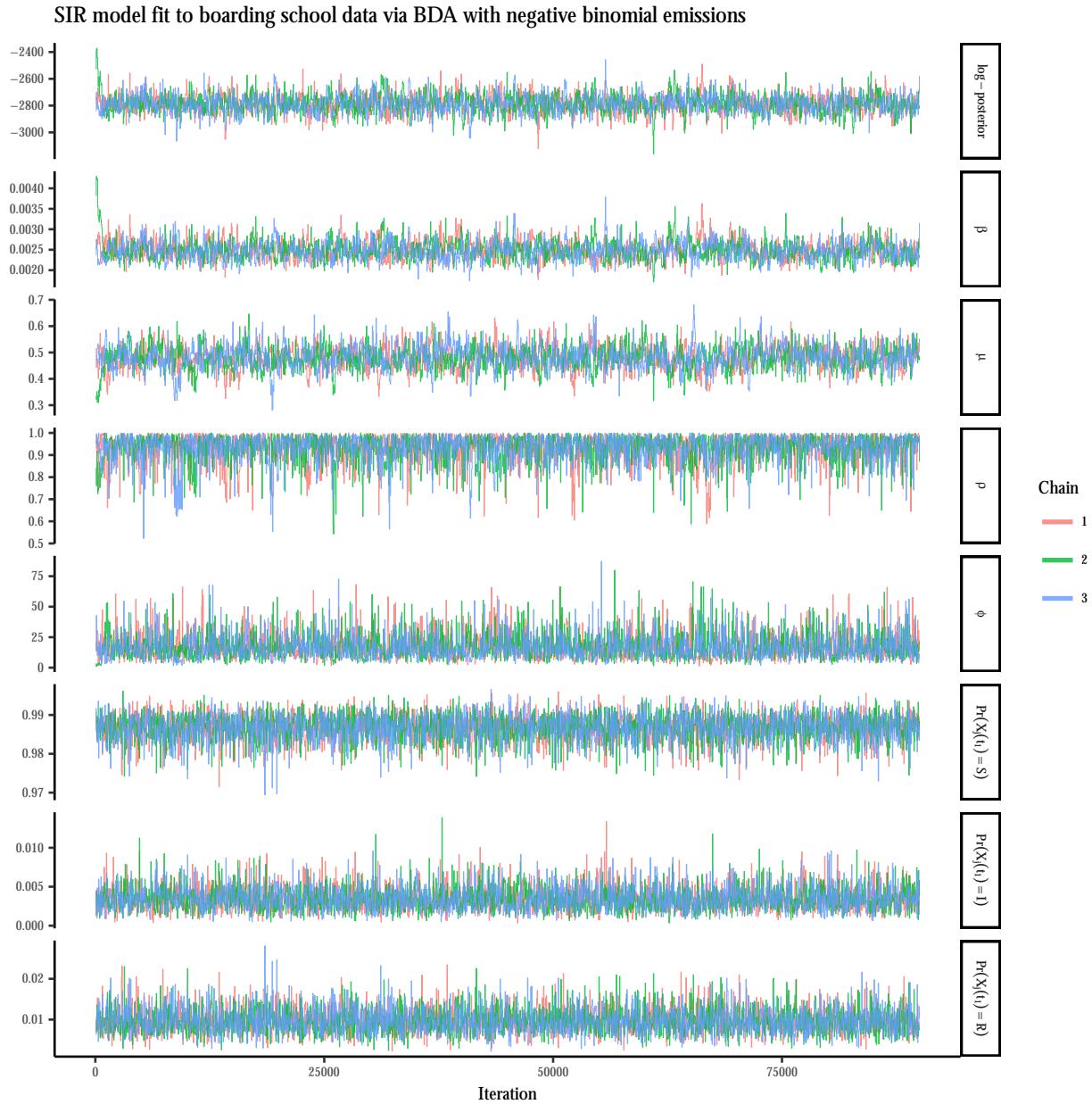


Figure A.27: Traceplots of the log-posterior and model parameters for the SIR model fit under negative binomial emissions using BDA following an initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

Traceplots for SEIR model fit via BDA with negative binomial emissions

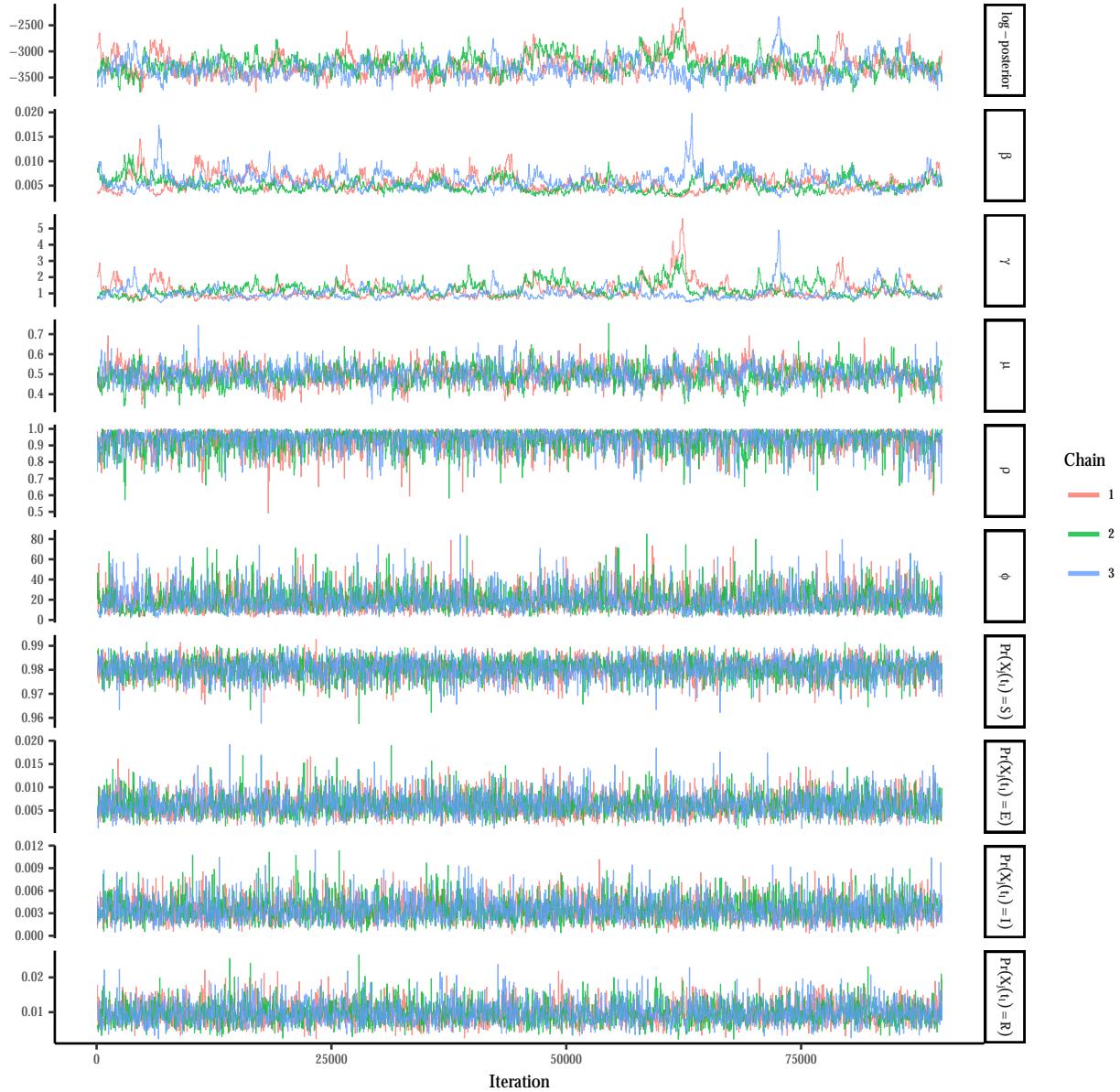


Figure A.28: Traceplots of the log-posterior and model parameters for the SEIR model fit under negative binomial emissions via BDA following an initial burn-in of 5,000 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, γ is the rate at which immunity is lost, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

Traceplots for SIR model fit via PMMH w/tau leaping with negative binomial emissions

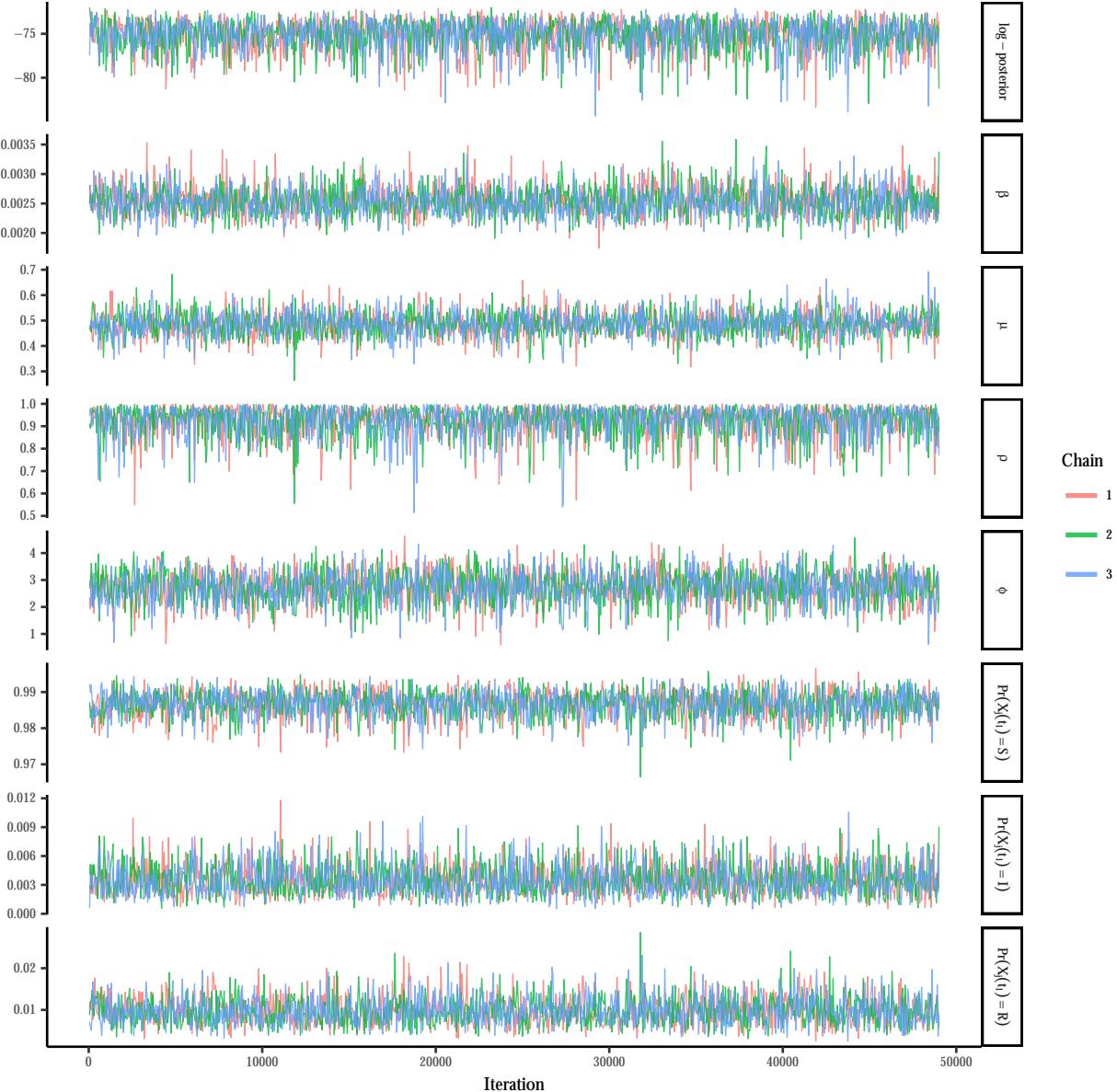


Figure A.29: Traceplots of the log-posterior and model parameters for the SIR model fit under negative binomial emissions using PMMH with 5,000 particles per chain and a time-step of 2 hours in the approximate τ -leaping algorithm, following a tuning run of 2,000 iterations to estimate the RWMH covariance matrix and in initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

Traceplots for SEIR model fit via PMMH w/tau leaping with negative binomial emissions

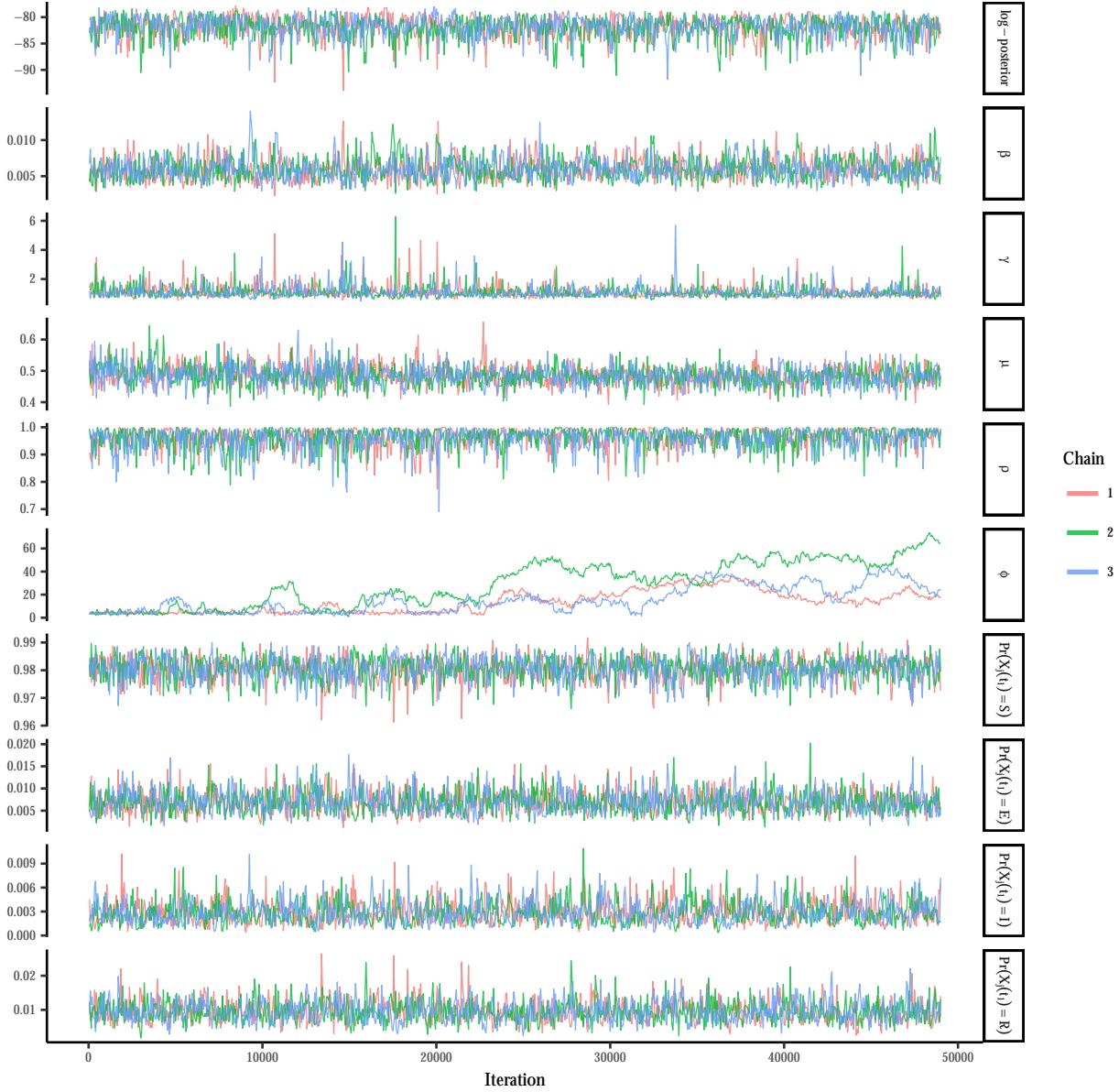


Figure A.30: Traceplots of the log-posterior and model parameters for the SEIR model fit under negative binomial emissions using PMMH with 5,000 particles per chain and a time-step of 2 hours in the approximate τ -leaping algorithm, following a tuning run of 2,000 iterations to estimate the RWMH covariance matrix and in initial burn-in of 100 iterations. β denotes the per-contact infectivity rate, μ is the recovery rate, γ is the rate at which immunity is lost, and ρ is the binomial sampling probability. Traceplots are thinned to display every 50th iteration.

Appendix B

APPENDIX TO CHAPTER 4

B.1 Tuning the Initial Elliptical Slice Sampling Bracket Width

When fitting SEMs with complex dynamics, e.g., when there are many strata or when the dynamics are time varying, we will be able improve the computational efficiency of our MCMC by initialing the 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 runtime of our MCMC.

In models where it is advantageous to shring the initial bracket width, we will typically set the initial bracket width to a constant times the standard deviation of the accepted angles in a tuning phase. Since we do not step out the 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. Figure B.1 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 joint Ebola model of Section 4.4. In this instance, we were able to substantially reduce the number of contractions, and hence likelihood evaluations, per ElliptSS update while leaving the distribution of accepted angles essentially unchanged. We like to call this a “free lunch”.

We mentioned in Section 4.2.6 that our ElliptSS algorithm was modified slightly from

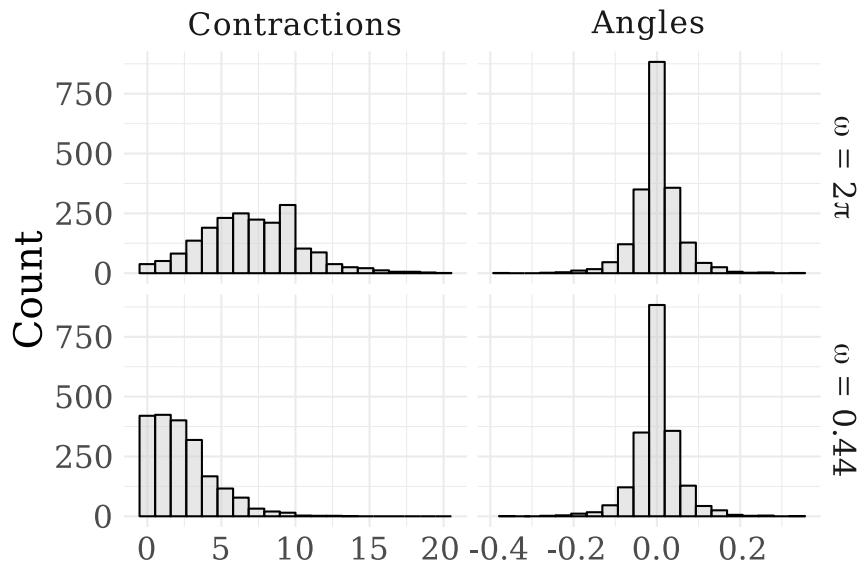


Figure B.1: Distributions of the numbers of contractions per ElliptSS update and the accepted angles for an MCMC chain for the joint Ebola model of Section 4.4 . An initial bracket width of 2π was used for the first 5,000 iterations (top row), after which the initial bracket width was set to $2\sqrt{2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

that presented in [159] in order to facilitate tuning of the initial ElliptSS bracket width. In both cases, the distribution of proposed states, \mathbf{Z}_{prop} , is centered at the current state \mathbf{Z}_{cur} . However, the distribution of angles of accepted states using our algorithm will be centered around 0, whereas the distribution of angles for accepted proposals using the algorithm in [159] will be bimodal with peaks at 0 and 2π . The algorithms are nevertheless equivalent due to the rotational symmetry of the proposals (a proposal made using an angle ϕ is equivalent to a proposal using $\phi + 2\pi$). It is more natural to compute the standard deviation of the accepted angles if the distribution of angles is symmetric about zero than if it is bimodal (which would require a rotation).

B.2 Inference for Initial Compartment Volumes

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

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

where \mathbf{p} is a vector of subject-level initial state probabilities, $\mathbf{P} = \text{diag}(\mathbf{p})$, N 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 N 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 N and probability vector \mathbf{p} if $\alpha = 1$, or of a dirichlet-multinomial distribution with parameters $\boldsymbol{\alpha} \implies \mathbf{p} = \boldsymbol{\alpha}/\boldsymbol{\alpha}^T(1)$, and over-dispersion $\alpha = (N + \boldsymbol{\alpha}^T\mathbf{1})/(1 + \boldsymbol{\alpha}^T\mathbf{1})$. 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 $\mathbf{m} = N\mathbf{p}$, $\mathbf{V} = \alpha N(\mathbf{P} - \mathbf{p}\mathbf{p}^T)$, and $\mathbf{V}^{1/2}$ be the matrix square root of \mathbf{V} , which we

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

Algorithm 10 Sampling LNA draws and initial volumes via elliptical slice sampling.

- 1: **procedure** `doELLIPTSS2`($\mathbf{Z}_{cur}^X, \mathbf{Z}_{cur}^{X_0}, \boldsymbol{\theta}, \mathbf{Y}, \mathcal{I}, \omega = 2\pi$)
- 2: Sample ellipse: $\mathbf{Z}_{prop}^X \sim N(\mathbf{0}, \mathbf{I}), \mathbf{Z}_{prop}^{X_0} \sim N(\mathbf{0}, \mathbf{I})$
- 3: Sample threshold: $u | \mathbf{x} \sim \text{Unif}(0, L(\mathbf{Y} | \text{doLNA}(\mathbf{Z}_{cur}, \boldsymbol{\theta}, \mathcal{I})))$
- 4: Position the bracket:

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

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

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

- 5: Make the initial proposal:

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

- 6: **if** $L(\mathbf{Y} | \text{doLNA}(\mathbf{Z}', \boldsymbol{\theta}', \mathcal{I})) > u$ **then** accept $\mathbf{Z}^{X'}, \mathbf{Z}^{X'_0}$
 - 7: **return** \mathbf{Z}'
 - 8: **else**
 - 9: Shrink bracket and try a new angle:
 - 10: **If:** $\phi < 0$ **then:** $L_\phi \leftarrow \phi$ **else:** $R_\phi \leftarrow \phi$
 - 11: $\phi \sim \text{Unif}(L_\phi, R_\phi)$
 - 12: **GoTo:** 5
-

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

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

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

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

As an example, consider the single country SEIR model fit to the incidence data from Sierra Leone in Section 4.4. This model includes parameters for the external force of infection and the effective population size, which add complexity to the usual formulation of the SEIR dynamics as being entirely driven by endogenous contacts within a closed homogeneously mixing population. The effective population size is roughly the size of the initially susceptible population. The model parameters on their natural scales are provided in Table B.1. Each of the model parameters has a clear marginal interpretation, but upon examining the pairwise scatterplots of the posterior (Figure B.2) it becomes obvious that the parameters interact

in highly non-linear ways. We would encounter a variety of pathological computational problems if we were to naively parameterize the MCMC estimation scale without considering the ways in which the parameters interact to affect the dynamics. For example, it would be extremely difficult for any sampler that does not account for the curvature in the posterior, e.g., Hamiltonian Monte Carlo (HMC), to explore the parameter space. (An aside: we experimented with implementing the LNA in `Stan` and using HMC to sample the posterior, but repeatedly integrating the LNA ODEs along with their augmented sensitivity equations was prohibitively slow for even simple models).

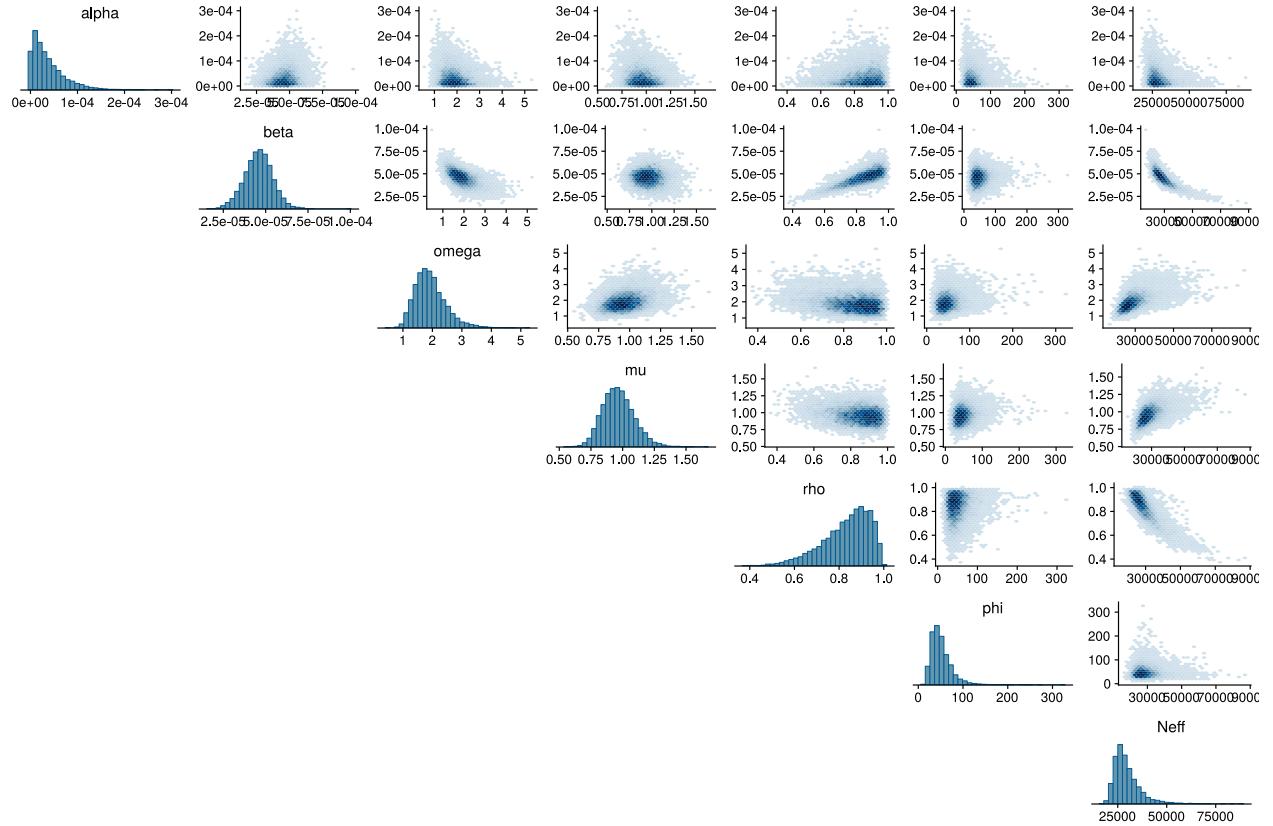


Figure B.2: Marginal histograms and pairwise scatterplots of posterior samples for parameters for the SEIR model fit to the Sierra Leone Ebola dataset using the estimation scale in Table B.3. The parameters on the estimation scales in this figure and their interpretations are provided in Table B.1.

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 B.2 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.

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

Parameter	Interpretation	Domain
$\log(R_{adj}^{ext}) = \log(\alpha N_{eff}/\mu)$	Log adjusted reproductive number given an infected outside the population	$(-\infty, \infty)$
$\log(R_{adj} - 1) = \log(\beta N_{eff}/\mu - 1)$	Log basic reproductive number given an infected inside the population 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	$(-\infty, \infty)$
$\log(N_{eff})$	Log effective population size	$(-\infty, \log(N))$

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 B.2 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 [152] relates the fraction of the population that eventually becomes

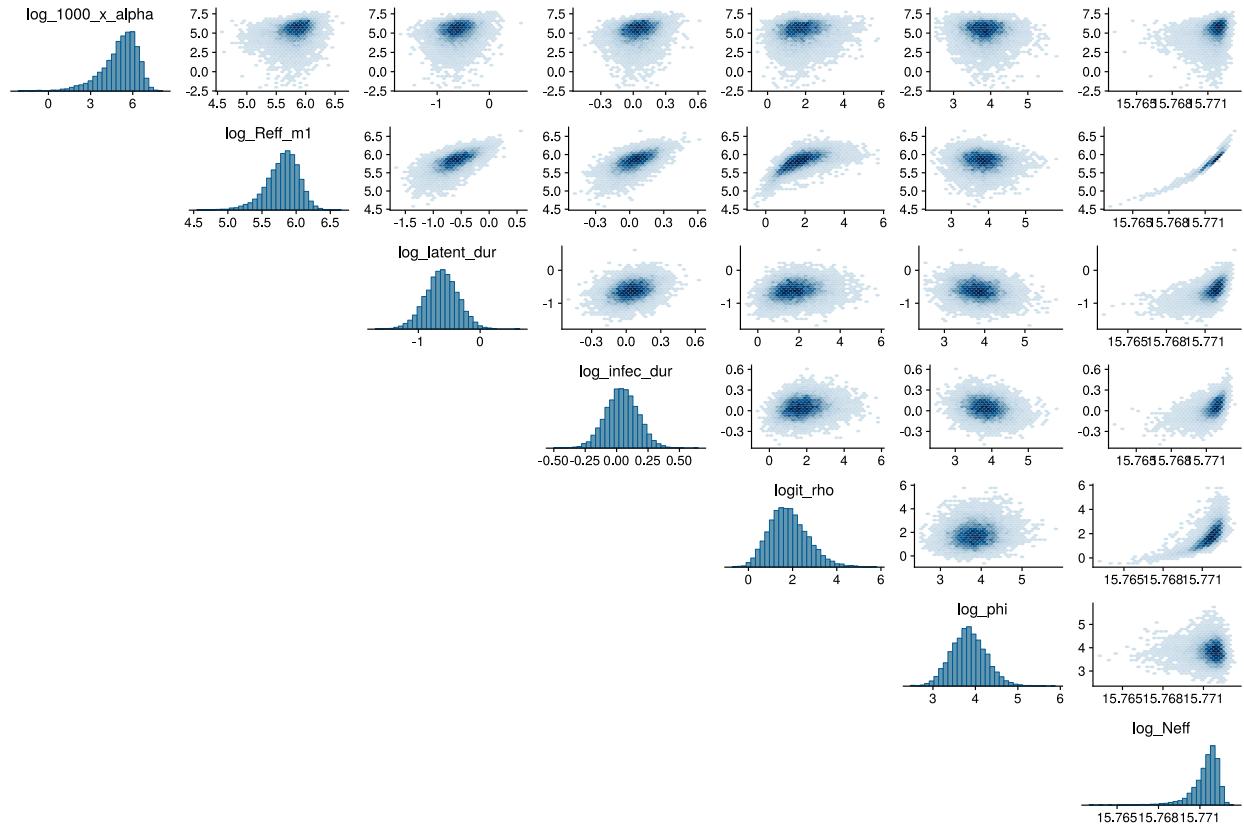


Figure B.3: Marginal histograms and pairwise scatterplots of posterior samples for parameters for the SEIR model fit to the Sierra Leone Ebola dataset using the estimation scale in Table B.3. The parameters on the estimation scales in this figure and their interpretations are provided in Table B.2.

infected π , with the basic reproductive number:

$$\pi = 1 - e^{-R0\pi}. \quad (\text{B.2})$$

As $R0$ increases, a larger fraction of the population becomes infected. As the effective population size, N_{eff} , increases, so too does the absolute outbreak size, $\pi \times N_{eff}$. The expected observed outbreak size is the absolute outbreak size multiplied by the mean case detection probability, $\pi \times \rho \times N_{eff}$, which should be concordant with the total number of observed cases. We can interpret the product, $\rho \times N_{eff}$, as a measure of the case detection rate. This is discussed further in the next section. Taken together, the combination of parameters that we would expect to jointly act on the model, *a posteriori*, is the adjusted reproduction number offset by the case detection rate, $R0 \times \rho \times N_{eff}$, which will enter our estimation scale as $\log(R0 \times \rho \times N_{eff})$. The other modification of the estimation scale in Table B.2 replaces the log latent period with the log ratio of infectious to latent period durations. The new estimation scale is given in Table B.3. On this estimation scale, the posterior for Sierra Leone is much better behaved, with weaker pairwise correlations and little in the way of non-linear relationships between the model parameters.

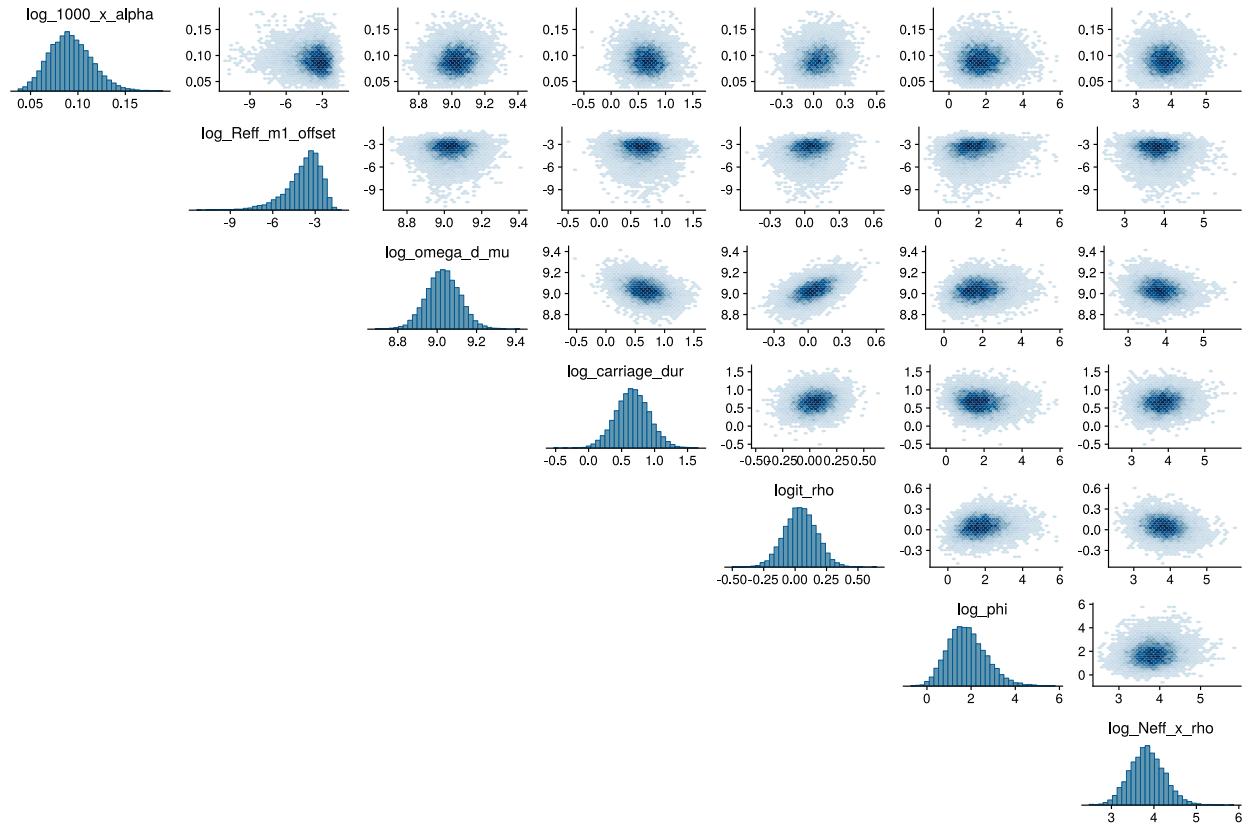


Figure B.4: Marginal histograms and pairwise scatterplots of posterior samples for parameters for the SEIR model fit to the Sierra Leone Ebola dataset using the estimation scale in Table B.3. The parameters on the estimation scales in this figure and their interpretations are provided in Table B.3.

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

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

B.4 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. One alternative is to split the population into two (or more) sub-populations, one that participates in the outbreak and another that is separated from the outbreak. For example, in the case of the SIR model, we might have two susceptible compartments, S^R and S^E , where S^R is a susceptible subpopulation that is effectively removed since they experience no infectious contact, and S^E is a susceptible population that may become exposed. In this model, the *effective population size* is $N_{eff} = S^E + I + R$. The model is otherwise constructed in the same way as the 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 N_{eff} enter into the complete data likelihood, (4.19), 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 [128, 222, 68]. Thus, the emission densities have means of the form, $\rho (N'_{S\bar{S}I}(t_\ell) - N'_{S\bar{S}I}(t_{\ell-1})) N_{eff}$, where $\mathbf{N}' = \mathbf{N}/N_{eff}$ is the equivalent representation of the LNA in terms of compartment concentrations.

In principle, absent prior information about the scales of ρ and N_{eff} , we will have difficulty estimating the case detection probability and effective population size. In practice, there are a few reasons that ρ and N_{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 mean case reporting rate, $\rho \times N_{eff}$, 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 R_0 and recovery rate for an SIR model, are known to be independent of the population size [61]. 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, $\rho N_{eff} \mathbf{N}'$. 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 can check that $\rho \times N_{eff}$ 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 B.4, 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 Section B.5.1). The results are summarized in Table B.5. 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 N_{eff} vis-a-vis their prior intervals. This suggests that the data are informative about $\rho \times N_{eff}$. The priors for ρ and N_{eff} are not completely flat, and the scale of the observed counts is itself informative about N_{eff} . Therefore, we still expect, and observe, some shrinkage for ρ and N_{eff} , individually.

Table B.4: 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)
$R_0 - 1$	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)	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)
N_{eff}	Effective population size	Unif(5000, 50000)	27500 (6125, 48875)

Table B.5: 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 $\rho \times N_{eff}$ are computed with respect to the induced prior resulting from the marginal priors for ρ and N_{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)
N_{eff}	0.96	800 (-18900, 16100)	32200 (14600, 41100)	0.75 (0.34, 0.96)
$\rho \times N_{eff}$	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)

B.5 Simulation Details and Additional Results for Section 4.3.1

B.5.1 Simulation Setup and MCMC Details

In this simulation, repeated for each of the three different regimes of population size and initial conditions given in Table 4.1, we simulated 500 datasets as follows:

1. Draw $\log(R_0 - 1)$, $1/\mu$, $\text{logit}(\rho)$, $\text{log}(\phi)$ from the priors given in Table 4.1.
2. Simulate an outbreak, $\mathbf{N}|\boldsymbol{\theta}$, under SIR dynamics from the MJP via Gillespie's direct algorithm [87]. 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., $Y_\ell \sim \text{Neg.Binomial}(\rho(N_{SI}(t_\ell) - N_{SI}(t_{\ell-1})), \phi)$. 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 4.1. Five MCMC chains per model were initialized at random values near the true parameters and run for 35,000 iterations per chain. The first 10,000 iterations used to warm up each chain and adaptively estimate the empirical covariance matrix to be used in the multivariate Gaussian random walk Metropolis–Hastings proposals for parameters. The empirical covariance matrix was initialized as 0.01 times an identity matrix. After the warm-up period, the empirical covariance matrix was frozen and the final 25,000 iterations from each chain were combined to form the final MCMC sample. Convergence was assessed using potential scale reduction factors (PSRFs) [35], computed via the `coda` R package [173]. PSRFs were less than 1.05 in all cases.

For models fit via the LNA and ODE approximations, the covariance matrix was adapted as in algorithm 4 of [11]. 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 [73].

Inference via the MMTL approximation within PMMH were fit using the `pomp` R package [125]. 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 degenerated during the adaptive phase of the MCMC. If this was the case, the MCMC was restarted at a different set of random initial conditions. The posterior samples from all five

MCMC chains were combined after discarding the initial samples from the adaptation phase.

B.5.2 Additional Coverage Simulation Results

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

Table B.6: Median (2.5%, 97.5%) quantiles of run times, in minutes, for MCMC chains in the coverage simulation presented in Section 4.3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	$\log(R_0)$	0.93	0 (-0.51, 0.55)	1.04 (0.83, 1.36)	1340 (274, 4029)	0.63 (0.13, 3.4)
LNA	$\log(\mu)$	0.95	-0.01 (-0.5, 0.48)	0.93 (0.74, 1.13)	1024 (199, 3988)	0.48 (0.1, 2.91)
LNA	$\text{logit}(\rho)$	0.93	-0.05 (-1.44, 0.56)	1.18 (0.54, 2.86)	1225 (316, 3941)	0.62 (0.17, 4.05)
LNA	$\log(\phi)$	0.96	0 (-0.66, 0.71)	1.35 (0.92, 2.22)	2021 (687, 4535)	1 (0.31, 3.53)
MMTL	$\log(R_0)$	0.95	0.03 (-0.48, 0.55)	1.09 (0.88, 1.39)	7483 (5453, 9197)	—
MMTL	$\log(\mu)$	0.95	-0.03 (-0.51, 0.45)	0.92 (0.75, 1.09)	7481 (5442, 9197)	—
MMTL	$\text{logit}(\rho)$	0.94	-0.03 (-1, 0.61)	1.17 (0.51, 2.98)	6725 (4328, 8506)	—
MMTL	$\log(\phi)$	0.96	-0.02 (-0.69, 0.71)	1.33 (0.91, 2.21)	7486 (5405, 9215)	—
ODE	$\log(R_0)$	0.89	-0.03 (-0.76, 0.56)	1.03 (0.65, 1.38)	6438 (4956, 7650)	185 (105, 340)
ODE	$\log(\mu)$	0.86	0.05 (-0.51, 0.79)	0.91 (0.53, 1.2)	6420 (4908, 7663)	184 (103, 344)
ODE	$\text{logit}(\rho)$	0.72	0.12 (-1.03, 1.37)	1.25 (0.48, 2.94)	6237 (4346, 7556)	203 (111, 401)
ODE	$\log(\phi)$	0.75	-0.25 (-1.64, 0.54)	1.25 (0.88, 2.11)	6529 (5459, 7778)	189 (116, 340)

Table B.7: Detailed small population ($N = 10,000$) regime results for the coverage simulation presented in Section 4.3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over-dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	$\log(R_0)$	0.93	-0.03 (-0.5, 0.49)	0.93 (0.66, 1.28)	2158 (397, 4981)	0.67 (0.16, 3)
LNA	$\log(\mu)$	0.93	0.01 (-0.43, 0.41)	0.83 (0.58, 1.07)	1919 (307, 5276)	0.63 (0.12, 3.05)
LNA	$\text{logit}(\rho)$	0.95	-0.01 (-1, 0.5)	1.13 (0.47, 2.81)	1704 (439, 4636)	0.61 (0.18, 3.49)
LNA	$\log(\phi)$	0.94	0.01 (-0.57, 0.7)	1.1 (0.79, 1.86)	2916 (1179, 5442)	1.02 (0.38, 3.1)
MMTL	$\log(R_0)$	0.95	-0.01 (-0.46, 0.51)	0.98 (0.67, 1.32)	7284 (4713, 9107)	—
MMTL	$\log(\mu)$	0.93	-0.01 (-0.47, 0.37)	0.82 (0.58, 1.05)	7167 (4488, 8912)	—
MMTL	$\text{logit}(\rho)$	0.95	-0.03 (-0.98, 0.56)	1.09 (0.43, 2.96)	6452 (3999, 8318)	—
MMTL	$\log(\phi)$	0.94	-0.02 (-0.59, 0.65)	1.1 (0.79, 1.84)	7096 (4640, 8999)	—
ODE	$\log(R_0)$	0.86	-0.02 (-0.56, 0.59)	0.84 (0.52, 1.27)	6592 (5213, 7771)	187 (88, 354)
ODE	$\log(\mu)$	0.82	-0.01 (-0.66, 0.47)	0.73 (0.44, 1.07)	6558 (5129, 7664)	190 (87, 359)
ODE	$\text{logit}(\rho)$	0.75	-0.03 (-1.13, 0.85)	0.88 (0.37, 2.75)	6421 (5017, 7643)	209 (107, 410)
ODE	$\log(\phi)$	0.82	-0.17 (-1.36, 0.57)	1.04 (0.78, 1.7)	6637 (5452, 7755)	193 (102, 365)

Table B.8: Detailed medium population ($N = 50,000$) regime results for the coverage simulation presented in Section 4.3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over-dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	$\log(R_0)$	0.95	-0.01 (-0.38, 0.51)	0.77 (0.47, 1.24)	3248 (638, 6127)	1.13 (0.22, 3.99)
LNA	$\log(\mu)$	0.95	0.01 (-0.44, 0.33)	0.67 (0.4, 1.02)	3134 (499, 6038)	1.12 (0.18, 3.95)
LNA	$\text{logit}(\rho)$	0.93	0.01 (-0.8, 0.57)	0.99 (0.35, 2.66)	2514 (572, 5905)	1.03 (0.24, 4.71)
LNA	$\log(\phi)$	0.95	0 (-0.44, 0.54)	0.94 (0.67, 1.5)	3987 (2201, 6184)	1.63 (0.71, 3.95)
MMTL	$\log(R_0)$	0.93	0.03 (-0.38, 1.04)	0.8 (0.31, 1.27)	7030 (3789, 8915)	—
MMTL	$\log(\mu)$	0.94	-0.02 (-0.46, 0.32)	0.65 (0.38, 0.98)	6858 (3602, 8871)	—
MMTL	$\log(\rho)$	0.92	0.01 (-0.7, 0.65)	0.95 (0.34, 2.64)	6166 (3282, 7906)	—
MMTL	$\log(\phi)$	0.95	-0.03 (-0.51, 0.52)	0.94 (0.66, 1.5)	6266 (3735, 8960)	—
ODE	$\log(R_0)$	0.89	-0.01 (-0.37, 0.55)	0.65 (0.36, 1.22)	6828 (5566, 8025)	193 (104, 417)
ODE	$\log(\mu)$	0.89	0 (-0.49, 0.34)	0.55 (0.3, 1)	6815 (5520, 7877)	198 (105, 428)
ODE	$\text{logit}(\rho)$	0.84	-0.01 (-0.78, 0.68)	0.74 (0.27, 2.62)	6575 (5219, 7838)	210 (118, 486)
ODE	$\log(\phi)$	0.91	-0.08 (-0.59, 0.46)	0.9 (0.64, 1.44)	6721 (5571, 7683)	209 (117, 413)

Table B.9: Detailed large population ($N = 250,000$) regime results for the coverage simulation presented in Section 4.3.1. Models were fit via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over-dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

B.6 Supplementary Coverage Simulations with Fixed Parameters

B.6.1 Simulation Setup

The simulations presented in this section supplement the results of Section 4.3.1 in assessing the statistical and computation performance of the LNA approximation vis-a-vis the ODE and MMTL approximations. In contrast to the previous coverage simulation, here we will fix the model parameters to one of four regimes, presented in Table B.10, that are characterized by either fast or moderate outbreak dynamics, and high or low detection probability. In each setting, we simulated 500 outbreaks from a MJP with SIR dynamics in a population of 50,000 individuals, five of whom were initially infected and the rest of whom were susceptible. The observed incidence in each epoch was a negative binomial sample of the true incidence. Outbreaks for which the number of observed cases was less than 25 were re-simulated. MCMC chains were tuned and SIR models were fit via the LNA, ODE, and MMTL approximations as described in Section B.5.1. The initial compartment volumes were fixed at the true values. The models were fit using diffuse priors, also presented in Table B.10. We caution that the priors used in this exercise are perhaps unreasonably diffuse, particularly in the context of epidemic modeling where prior information is often available, and that we expect credible intervals will be overly wide as a result (we would still expect nominal coverage to be incorrect, even under tighter priors, since the datasets were simulated under fixed parameter regimes).

B.6.2 Fixed Parameter Coverage Results

Coverage for credible intervals of ODE models tended to fall below nominal levels in spite of the bias towards wide intervals due to the diffusivity of the priors. This was particularly the case in parameter regimes 1 and 3, where the basic reproductive number was lower (and hence the simulated outbreak trajectories further from their thermodynamic limits). In these parameter regimes, coverage was particularly poor due as estimates of the outbreak dynamics tended to be farther from their true values and credible intervals were too tight and did not properly account for uncertainty about the parameter estimates, particularly those governing the measurement process. Coverage levels for models fit via the LNA and MMTL approximations exceeded their nominal levels as expected.

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

Table B.10: Parameter regimes under which datasets were simulated and priors used to fit SIR models. Five hundred datasets were simulated for each of the parameter regimes from a MJP with SIR dynamics. $R_0 = \beta N / \mu$ is the basic reproductive number and μ is the recovery rate. The observed incidence was a negative binomial sample of the true incidence in each inter-observation interval with case detection probability ρ and overdispersion parameter ϕ .

	Regime 1	Regime 2	Regime 3	Regime 4
	Low R0/Low ρ	High R0/Low ρ	Low R0/High ρ	High R0/High ρ
R0	1.75	3.25	1.75	3.25
ρ	0.25	0.25	0.75	0.75
μ	1	0.4	1	0.4
ϕ	5	5	5	5

Parameter	Interpretation	Prior	Median (95% Interval)
$R_0 - 1$	Basic reproduction # - 1	LogCauchy(0.4, 1)	$\implies R_0 = 2.50$ (1.00, 4.9×10^5)
$1/\mu$	Mean infectious period	LogCauchy(-0.7, 1)	1.43 (4.3×10^{-6} , 4.7×10^5)
ρ	Mean case detection prob.	Unif(0, 1)	0.5 (0.025, 0.975)
ϕ	Neg.Binom. overdispersion	LogCauchy(1.5, 1)	4.48 (1.4×10^{-5} , 1.5×10^6)

B.6.3 Fixed Parameter Coverage Simulation Results

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

Table B.11: Median (2.5%, 97.5%) quantiles of run times, in minutes, for MCMC chains in fixed parameter coverage simulations. Models were fit via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). True parameter values are presented in Table B.10. Parameter regimes 1 and 3 had slower outbreak dynamics ($R_0 = 1.75$, vs. $R_0 = 3.25$). Parameter regimes 3 and 4 had higher case detection rates ($\rho = 0.75$, vs $\rho = 0.25$).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	$\log(R_0)$	0.99	0.23 (-0.33, 0.85)	1.9 (1.2, 3.49)	1200 (251, 2508)	19.1 (2.7, 76.4)
LNA	$\log(\mu)$	0.98	-0.21 (-0.81, 0.28)	1.73 (1.04, 3.42)	1086 (228, 2295)	27.9 (3.7, 100.5)
LNA	$\text{logit}(\rho)$	0.98	-0.06 (-0.46, 0.4)	1.07 (0.73, 1.67)	1233 (360, 2180)	8.22 (2.03, 21.4)
LNA	$\log(\phi)$	0.98	0.01 (-0.51, 0.74)	1.32 (1.17, 1.71)	2950 (1145, 4334)	10.8 (2.1, 30.0)
MMTL	$\log(R_0)$	0.99	0.22 (-0.86, 0.87)	4.4 (1.92, 53.26)	232 (110, 445)	—
MMTL	$\log(\mu)$	1.00	-0.22 (-0.79, 0.52)	2.02 (1.26, 3.63)	125 (58, 283)	—
MMTL	$\text{logit}(\rho)$	0.99	-0.06 (-0.47, 0.48)	1.21 (0.83, 1.72)	492 (258, 874)	—
MMTL	$\log(\phi)$	0.97	-0.01 (-0.54, 0.71)	1.32 (1.16, 1.72)	942 (541, 1791)	—
ODE	$\log(R_0)$	0.84	0.3 (-0.77, 2.43)	1.78 (1.12, 5.93)	3364 (252, 6513)	3894 (169, 13949)
ODE	$\log(\mu)$	0.78	-0.24 (-2.57, 0.7)	1.56 (0.93, 5.81)	3307 (243, 6489)	6454 (277.52, 15693)
ODE	$\text{logit}(\rho)$	0.78	-0.08 (-0.67, 0.79)	0.77 (0.51, 1.52)	5225 (2398, 6933)	2500 (869, 5719)
ODE	$\log(\phi)$	0.94	-0.1 (-0.73, 0.57)	1.26 (1.14, 1.54)	5702 (2768, 6991)	1451 (532, 3117)

Table B.12: Detailed results for the fixed parameter simulation in which outbreaks and datasets were simulated under parameter regime 1, characterized by slow outbreak dynamics ($R_0 = 1.75$) and low mean case detection probability ($\rho = 0.25$). Models were fit via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the negative binomial case detection probability, ϕ is the negative binomial over-dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	$\log(R_0)$	0.99	-0.35 (-0.87, -0.05)	2.01 (1.42, 3.03)	1551 (399, 3065)	64.2 (10.3, 265.1)
LNA	$\log(\mu)$	0.99	0.34 (0.09, 0.8)	1.9 (1.34, 2.92)	1431 (368, 2865)	23.8 (5.0, 101.7)
LNA	$\text{logit}(\rho)$	0.94	0.12 (-0.23, 0.55)	1.02 (0.73, 1.71)	1354 (509, 2357)	13.1 (3.2, 40.6)
LNA	$\log(\phi)$	0.96	0.02 (-0.53, 0.81)	1.38 (1.25, 1.76)	3286 (1756, 4442)	19.5 (6.3, 50.5)
MMTL	$\log(R_0)$	0.99	-0.41 (-1.11, -0.08)	13.94 (2.95, 55.97)	132 (31, 314)	—
MMTL	$\log(\mu)$	0.99	0.4 (0.12, 0.99)	2.74 (2.13, 3.69)	213 (90, 1944)	—
MMTL	$\text{logit}(\rho)$	0.90	0.18 (-0.18, 0.64)	1.78 (1.1, 2.73)	375 (171, 737)	—
MMTL	$\log(\phi)$	0.97	-0.03 (-0.59, 0.78)	1.42 (1.25, 2)	586 (315, 993)	—
ODE	$\log(R_0)$	0.99	-0.22 (-0.84, 0.65)	1.85 (1.31, 3.58)	3979 (1875, 5747)	9412 (2368, 34873)
ODE	$\log(\mu)$	0.98	0.21 (-0.78, 0.8)	1.72 (1.21, 3.53)	3976 (1851, 5752)	3824 (9812, 12367)
ODE	$\text{logit}(\rho)$	0.85	0.05 (-0.34, 0.52)	0.71 (0.5, 1.08)	5024 (2846, 6390)	2753 (1205, 6858)
ODE	$\log(\phi)$	0.95	-0.03 (-0.64, 0.67)	1.33 (1.23, 1.6)	5634 (4377, 6661)	2027 (995, 4479)

Table B.13: Detailed results for the fixed parameter simulation in which outbreaks and datasets were simulated under parameter regime 2, characterized by fast outbreak dynamics ($R_0 = 3.25$) and low mean case detection probability ($\rho = 0.25$). Models were fit via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the negative binomial case detection probability, ϕ is the negative binomial over-dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	$\log(R_0)$	0.96	0.26 (-0.22, 0.93)	1.68 (1.01, 3.21)	1354 (198, 3054)	16.0 (1.6, 70.2)
LNA	$\log(\mu)$	0.96	-0.23 (-0.85, 0.16)	1.49 (0.84, 3.23)	1210 (191, 2931)	25.1 (2.8, 83.6)
LNA	$\text{logit}(\rho)$	0.99	-0.2 (-0.96, 0.84)	3.1 (1.61, 4.6)	935 (393, 1893)	6.9 (2.4, 18.7)
LNA	$\log(\phi)$	0.97	0.02 (-0.52, 0.67)	1.25 (1.12, 1.52)	2851 (1425, 4283)	13.4 (4.3, 31.6)
MMTL	$\log(R_0)$	0.99	0.24 (-0.48, 0.96)	2.49 (1.44, 36.95)	313 (139, 634)	—
MMTL	$\log(\mu)$	0.99	-0.24 (-0.87, 0.24)	1.75 (1, 3.44)	141 (57, 359)	—
MMTL	$\text{logit}(\rho)$	1.00	-0.2 (-0.93, 0.86)	3.48 (2.04, 4.99)	439 (244, 711)	—
MMTL	$\log(\phi)$	0.97	-0.02 (-0.53, 0.64)	1.25 (1.12, 1.56)	723 (372, 1376)	—
ODE	$\log(R_0)$	0.79	0.34 (-0.41, 2.6)	1.52 (0.95, 5.73)	3688 (285, 6453)	2951 (127, 10756.39)
ODE	$\log(\mu)$	0.77	-0.28 (-2.73, 0.36)	1.31 (0.8, 5.69)	3650 (263, 6457)	5349 (177, 14237)
ODE	$\text{logit}(\rho)$	0.80	-0.19 (-1.41, 1.57)	2.27 (0.9, 4.66)	3418 (1896, 5721)	1767 (704, 4279)
ODE	$\log(\phi)$	0.91	-0.15 (-0.75, 0.48)	1.16 (1.08, 1.34)	5787 (2679, 7035)	1754 (640, 4124)

Table B.14: Detailed results for the fixed parameter simulation in which outbreaks and datasets were simulated under parameter regime 3, characterized by slow outbreak dynamics ($R_0 = 1.75$) and high mean case detection probability ($\rho = 0.75$). Models were fit via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the recovery rate, ρ is the negative binomial case detection probability, ϕ is the negative binomial over-dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

Method	Parameter	Coverage	PMD	95% CIW	ESS	Rel. GM ESS/CPU time
LNA	$\log(R_0)$	1.00	-0.27 (-0.68, 0.04)	1.7 (1.24, 2.64)	2217 (504, 4142)	68.6 (10.0, 342.8)
LNA	$\log(\mu)$	1.00	0.25 (-0.01, 0.62)	1.62 (1.18, 2.63)	2014 (493, 3894)	26.4 (4.7, 85.4)
LNA	$\text{logit}(\rho)$	0.98	0.25 (-0.58, 1.35)	3.49 (2.06, 4.63)	1290 (563, 2572)	14.8 (4.9, 59.9)
LNA	$\log(\phi)$	0.96	0.03 (-0.52, 0.76)	1.31 (1.2, 1.62)	3506 (1910, 4689)	27.0 (11.2, 66.1)
MMTL	$\log(R_0)$	1.00	-0.35 (-0.93, -0.04)	13.37 (1.83, 46.16)	181 (48, 409)	—
MMTL	$\log(\mu)$	1.00	0.33 (0.06, 0.83)	2.65 (1.61, 3.72)	272 (128, 1994)	—
MMTL	$\text{logit}(\rho)$	0.96	0.4 (-0.49, 1.57)	4.49 (3.15, 7.21)	309 (133, 545)	—
MMTL	$\log(\phi)$	0.97	-0.06 (-0.63, 0.64)	1.62 (1.26, 2.61)	457 (213, 848)	—
ODE	$\log(R_0)$	0.99	-0.18 (-0.78, 0.84)	1.67 (1.14, 3.73)	4463 (1987, 6336)	7879 (1915, 29567)
ODE	$\log(\mu)$	0.99	0.18 (-0.99, 0.73)	1.56 (1.07, 3.67)	4424 (1952, 6312)	3254 (867, 8316)
ODE	$\text{logit}(\rho)$	0.89	0.13 (-0.86, 1.71)	2.59 (1.1, 4.57)	3378 (1971, 5315)	2271 (916, 7206)
ODE	$\log(\phi)$	0.94	-0.07 (-0.66, 0.62)	1.26 (1.17, 1.48)	5697 (4518, 6929)	2490 (1211, 6558)

Table B.15: Detailed results for the fixed parameter simulation in which outbreaks and datasets were simulated under parameter regime 4, characterized by fast outbreak dynamics ($R_0 = 3.25$) and high mean case detection probability ($\rho = 0.75$). Models were fit via the linear noise approximation (LNA), multinomial modified τ -leaping (MMTL) within particle marginal Metropolis–Hastings, and deterministic ordinary differential equations (ODE). R_0 is the basic reproductive number of an outbreak, μ is the negative binomial case detection probability, ϕ is the negative binomial over-dispersion parameter. We report the coverage rates of 95% Bayesian credible intervals along with 50% (2.5%, 97.5%) quantiles of posterior median deviations (PMD), 95% credible interval widths (CIW), effective sample size (ESS), and relative geometric mean effective sample size per CPU time (Rel. GM ESS/CPU time).

B.7 Synthetic Ebola Outbreak — Simulation Settings and Additional Results

B.7.1 Simulation Details

We simulated an outbreak for a stratified SEIR model, diagrammed in Figure B.5, for the spread of Ebola in Guinea, Liberia, and Sierra Leone, via Gillespie's direct algorithm [87]. Transmission was driven by contact with infected individuals within the country, and migration of infected individuals from other countries. The countries differed in their transmission dynamics, case detection probabilities, and effective population sizes. In addition, the times at which transmission commenced in each country were staggered. The outbreak began in Guinea with transmission beginning in Liberia at the start of week 10, and in Sierra Leone at the start of week 19. The rates of between country transmission, along with the other rates of state transition, for Liberia and Sierra Leone were set to zero prior to those times. The constants and parameters used in simulating the outbreak and data are given in Tables B.16 and B.16. The priors used in fitting the models are given in Tables B.17 and B.18.

Table B.16: True and effective population sizes, and initial conditions for a simulated outbreak in Guinea, Liberia, and Sierra Leone. Week zero is one week before the first observation is accrued and the time at which within-country transmission was assumed to begin.

	Guinea	Liberia	Sierra Leone
True population size (N)	11.8 million	4.4 million	7.1 million
Effective population size (N_{eff})	15,000	35,000	25,000
Week zero (t_0)	0	10	19
Initial state at t_0 (S_0, E_0, I_0, R_0)	$(11.8 \times 10^6 - 7, 4, 3, 0)$	$(4.4 \times 10^6 - 5, 3, 2, 0)$	$(7.1 \times 10^6 - 5, 3, 2, 0)$

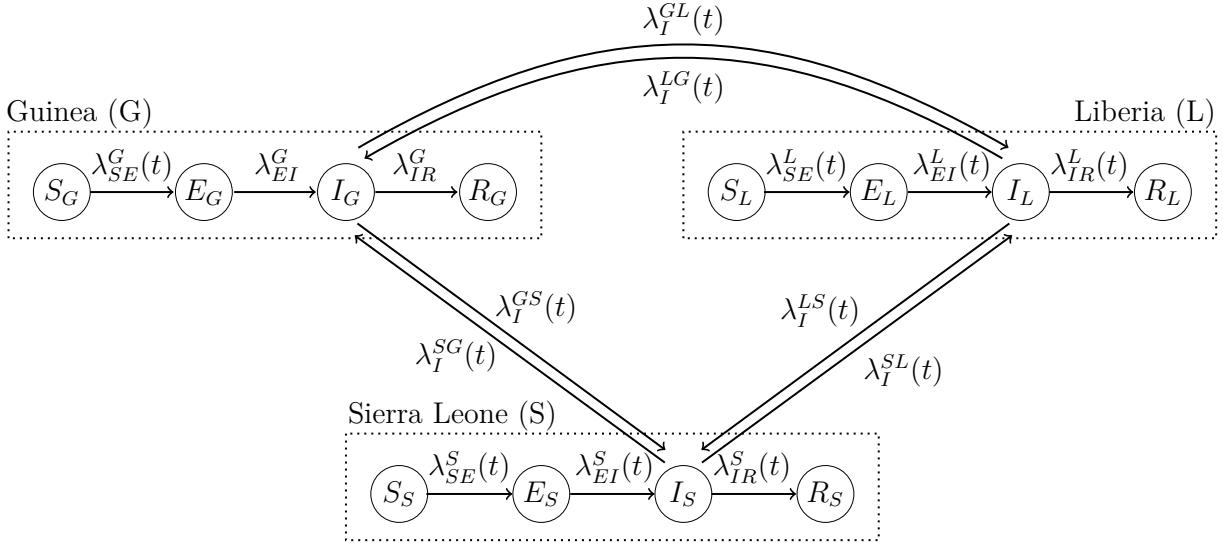


Figure B.5: Diagram of state transitions for a stratified SEIR model used in simulating an outbreak in Guinea, Liberia, and Sierra Leone. Dashed boxes denote countries, nodes in circles denote the model compartments: susceptible (S), exposed (E), infectious (I), recovered (R). Compartments are subscripted 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_{SE}^L is the rate at which susceptible individuals become exposed in Liberia, λ_I^{SG} is the rate at which an infected individual in Sierra Leone migrates to Guinea. The rate at which susceptible individuals in Guinea become infected is time varying with one changepoint at 33 weeks. Transmission in Liberia and Sierra Leone commenced at 10 and 19 weeks, respectively. Full expressions for the rates are given in Table B.19.

Table B.17: Parameters used to simulate an Ebola outbreak in West Africa using a country-stratified SEIR model, prior distributions, and 95% prior intervals. Subscripts, G, L, S , indicate specific countries, or generic countries A, B if a prior is shared. Adjusted reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$.

Parameter	Truth	Interpretation	Prior	Median (95% Interval)
$R_{eff,G}^{(1)}$, $t < 33$	1.3	Adjusted reproduction #	$R_{eff,G}^{(2)} / R_G^{(1)} \sim \text{LogNormal}(0, 0.5^2)$	$R_{eff,G}^{(2)} / R_{eff,G}^{(1)} = 1.00$ (0.38, 2.66)
$R_{eff,G}^{(2)}$, $t \geq 33$	1.2	Adjusted reproduction #	$R_{eff,G}^{(2)} - 1 \sim \text{LogNormal}(\log(0.5), 1.1)$	$\Rightarrow R_{eff,G}^{(2)} = 1.50$ (1.06, 5.32)
$R_{eff,L}$	2.0	Adjusted reproduction #	$R_{eff,L} \sim \text{LogNormal}(\log(0.5), 1.1)$	$\Rightarrow R_{eff,L} = 1.50$ (1.06, 5.32)
$R_{eff,S}$	1.55	Adjusted reproduction #	$R_{eff,S} \sim \text{LogNormal}(\log(0.5), 1.1)$	$\Rightarrow R_{eff,S} = 1.50$ (1.06, 5.32)
$\alpha_{GS}, \alpha_{GL}, \alpha_{LG}, \alpha_{LS}, \alpha_{SG}, \alpha_{SL}$	0.003	Infectious migration rate from country A to B	$1000\alpha_{AB} \sim \text{Exponential}(0.25)$	# migrations per 1000 infected = 2.78 (0.10, 14.76)
ω_G	1.1	Rate from $E_G \rightarrow I_G$	$(1/\mu_G)/(1/\omega_G) \sim \text{LogNormal}(0.0, 0.3^2)$	$(1/\mu_G)/(1/\omega_G) = 1.00$ (0.56, 1.80)
ω_L	0.85	Rate from $E_L \rightarrow I_L$	$(1/\mu_L)/(1/\omega_L) \sim \text{LogNormal}(0.0, 0.3^2)$	$(1/\mu_L)/(1/\omega_L) = 1.00$ (0.56, 1.80)
ω_S	1	Rate from $E_S \rightarrow I_S$	$(1/\mu_S)/(1/\omega_S) \sim \text{LogNormal}(0.0, 0.3^2)$	$(1/\mu_S)/(1/\omega_S) = 1.00$ (0.56, 1.80)
μ_G	1	Rate from $E_G \rightarrow I_G$	$1/\mu_G \sim \text{LogNormal}(0, 0.25^2)$	$1/\mu_G = 7.0$ (4.29, 11.43)
μ_L	0.8	Rate from $E_L \rightarrow I_L$	$1/\mu_L \sim \text{LogNormal}(0, 0.25^2)$	$1/\mu_L = 7.0$ (4.29, 11.43)
μ_S	1.1	Rate from $E_S \rightarrow I_S$	$1/\mu_S \sim \text{LogNormal}(0, 0.25^2)$	$1/\mu_S = 7.0$ (4.29, 11.43)
$N_{eff,G}$	15,000	Effective population size	$N_{eff,G} \sim \text{Unif}(\sum_{\ell=1}^L Y_\ell^G, N_G/4)$	$N_{eff,G} \in (4032, 2.95 \times 10^6)$
$N_{eff,L}$	35,000	Effective population size	$N_{eff,L} \sim \text{Unif}(\sum_{\ell=1}^L Y_\ell^L, N_L/4)$	$N_{eff,L} \in (5945, 1.1 \times 10^6)$
$N_{eff,S}$	25,000	Effective population size	$N_{eff,S} \sim \text{Unif}(\sum_{\ell=1}^L Y_\ell^S, N_S/4)$	$N_{eff,S} \in (9531, 1.775 \times 10^6)$
ρ_G	0.85	Mean case detection prob.	$\text{logit}(\rho_G) \sim \text{LogNormal}(0, 1.4^2)$	$\Rightarrow \rho_G = 0.5, (0.06, 0.94)$
ρ_L	0.25	Mean case detection prob.	$\text{logit}(\rho_L) \sim \text{LogNormal}(0, 1.4^2)$	$\Rightarrow \rho_L = 0.5, (0.06, 0.94)$
ρ_S	0.55	Mean case detection prob.	$\text{logit}(\rho_S) \sim \text{LogNormal}(0, 1.4^2)$	$\Rightarrow \rho_S = 0.5, (0.06, 0.94)$
ϕ_G, ϕ_L, ϕ_S	50	Neg. Binomial overdispersion	$\log(\phi) \sim \text{LogNormal}(0, 5^2)$	$\Rightarrow \rho_G = 1, (5.5 \times 10^{-5}, 1.8 \times 10^4)$

Table B.18: Priors for the initial compartment volumes at the times when transmission commenced in each country for a simulated Ebola outbreak in West Africa. The initial compartment volumes for each country are assigned independent truncated multivariate normal priors that respect the boundary conditions on the state space of compartment volumes for that country (explained in Section B.2). 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	$(11.8 \times 10^6 - 7.1, 4, 3, 0.1)$
Liberia	$(4.4 \times 10^6 - 5.1, 3, 2, 0.1)$
Sierra Leone	$(7.1 \times 10^6 - 5.1, 3, 2, 0.1)$

B.7.2 MCMC Details

We fit two models to the simulated data, the true model where migration was modeled explicitly and the approximated model, described in Section 4.3.3, where cross-border transmission was incorporated into the rate of infection via virtual migration. The model fitting procedure and priors were the same for both models. We ran five chains for each model, initialized at random parameter values, for 200,000 iterations per chain. The model parameters, not including the initial compartment volumes, were jointly updated via MVNSS (see Section 2.2.1). 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 with covariances for parameters belonging to different countries set to zero. Thus, country specific parameters were proposed jointly but were independent in the proposal. Migration parameters were blocked with parameters corresponding to the destination country. The initial compartment volumes were updated jointly with the LNA

Table B.19: Rates of transitions between compartments for the full stratified SEIR model for the spread of Ebola in Guinea, Liberia, and Sierra Leone with cross-country migration of infected individuals across country borders. These rates were used in simulating the outbreak and data, and in the supplementary model presented in Section B.7.3. The model is diagrammed in Figure B.5. Subscripts in the rates denote model compartments, superscripts denote countries. Subscripts for parameters denote countries. 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. The rate at which susceptibles in Guine become infected is time varying with one changepoint at 33 weeks. Transmission in Liberia and Sierra Leone prior to weeks 10 and 19 is expressed in the initial compartment volumes at those times. Thus, all respective transition rates are set to zero prior to the times when transmission commenced in those countries.

From	To	Country	Rate
Susceptible	Exposed	Guinea	$\lambda_{SE}^G(t) = (\beta_G^{(1)} \mathbb{1}_{\{t < 33\}} + \beta_G^{(2)} \mathbb{1}_{\{t \geq 33\}}) S_G I_G$
Susceptible	Exposed	Liberia	$\lambda_{SE}^L(t) = \beta_L S_L I_L \mathbb{1}_{\{t \geq 10\}}$
Susceptible	Exposed	Sierra Leone	$\lambda_{SE}^S(t) = \beta_S S_S I_S \mathbb{1}_{\{t \geq 19\}}$
Exposed	Infected	Guinea	$\lambda_{EI}^G = \omega_G E_G$
Exposed	Infected	Liberia	$\lambda_{EI}^L(t) = \omega_L E_L \mathbb{1}_{\{t \geq 10\}}$
Exposed	Infected	Sierra Leone	$\lambda_{EI}^S(t) = \omega_S E_S \mathbb{1}_{\{t \geq 19\}}$
Infected	Recovered	Guinea	$\lambda_{IR}^G = \mu_G I_G$
Infected	Recovered	Liberia	$\lambda_{IR}^L(t) = \mu_L I_L \mathbb{1}_{\{t \geq 10\}}$
Infected	Recovered	Sierra Leone	$\lambda_{IR}^S(t) = \mu_S I_S \mathbb{1}_{\{t \geq 19\}}$
Infected (A)	Infected (B)	Countries A,B	$\lambda_I^{AB}(t) = \alpha_{AB} I_A \mathbb{1}_{\{t \geq t_0^A\}} \mathbb{1}_{\{t \geq t_0^B\}}$

Table B.20: Approximate rates of transitions between compartments for the stratified SEIR model for the spread of Ebola in Guinea, Liberia, and Sierra Leone with cross-country transmission via virtual migration of infected individuals. These rates were used in the primary model presented in Section 4.3.3. The model is diagrammed in Figure B.5. Subscripts in the rates denote model compartments, superscripts denote countries. Subscripts for parameters denote countries. 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. The rate at which susceptibles in Guine become infected is time varying with one changepoint at 33 weeks. Transmission in Liberia and Sierra Leone prior to weeks 10 and 19 is expressed in the initial compartment volumes at those times. Thus, all respective transition rates are set to zero prior to the times when transmission commenced in those countries.

From	To	Country	Rate
Susceptible	Exposed	Guinea	$\lambda_{SE}^G(t) = (\beta_G^{(1)} \mathbf{1}_{\{t < 33\}} + \beta_G^{(2)} \mathbf{1}_{\{t \geq 33\}}) \times (I_G + \alpha_{LG} I_L \mathbf{1}_{\{t \geq 10\}} + \alpha_{SG} I_S \mathbf{1}_{\{t \geq 19\}}) S_G$
Susceptible	Exposed	Liberia	$\lambda_{SE}^L(t) = \beta_L (I_L + \alpha_{GL} I_G + \alpha_{SL} I_S \mathbf{1}_{\{t \geq 19\}}) S_L \mathbf{1}_{\{t \geq 10\}}$
Susceptible	Exposed	Sierra Leone	$\lambda_{SE}^S(t) = \beta_S (I_S + \alpha_{GS} I_G + \alpha_{LS} I_L) S_S \mathbf{1}_{\{t \geq 19\}}$
Exposed	Infected	Guinea	$\lambda_{EI}^G = \omega_G E_G$
Exposed	Infected	Liberia	$\lambda_{EI}^L(t) = \omega_L E_L \mathbf{1}_{\{t \geq 10\}}$
Exposed	Infected	Sierra Leone	$\lambda_{EI}^S(t) = \omega_S E_S \mathbf{1}_{\{t \geq 19\}}$
Infected	Recovered	Guinea	$\lambda_{IR}^G = \mu_G I_G$
Infected	Recovered	Liberia	$\lambda_{IR}^L(t) = \mu_L I_L \mathbf{1}_{\{t \geq 10\}}$
Infected	Recovered	Sierra Leone	$\lambda_{IR}^S(t) = \mu_S I_S \mathbf{1}_{\{t \geq 19\}}$

paths in an EllipSS step. The MCMC alternated between three ElliptSS and two 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 B.3 with the one addition being the log ratio of adjusted reproductive numbers, subtract one, for Guinea, $\log\left((R_{eff,G}^{(2)} - 1)/(R_{eff,G}^{(1)} - 1)\right)$. Convergence was assessed visually by inspection of traceplots of posterior samples, and via potential scale reduction factors (PSRFs) [35] computed via the `coda` R package [173]. The full model took approximately 9.6 days to run, while the approximate model took about 2.8 days.

B.7.3 Additional Results

Table B.21: Posterior estimates of initial numbers of exposed, infected, and recovered individuals for a stratified approximate SEIR model with virtual migration of infecteds fit to simulated Ebola data. The effective number of susceptibles is equal to the effective population size, reported in Table B.22, minus the numbers of exposed, infected, and recovered individuals, but is not reported since the effective population size is not marginally identifiable.

Parameter	Truth	Estimate
$E_{0,G}$	4.00	4.99 (1.39, 8.66)
$E_{0,G}$	3.00	3.4 (0.6, 6.76)
$E_{0,G}$	3.00	3.16 (0.45, 6.55)
$I_{0,G}$	3.00	3.7 (0.82, 6.91)
$I_{0,L}$	2.00	2.67 (0.45, 5.23)
$I_{0,S}$	2.00	2.46 (0.4, 4.95)
$R_{0,G}$	0.00	0.12 (0.07, 0.16)
$R_{0,L}$	0.00	0.12 (0.07, 0.17)
$R_{0,S}$	0.00	0.11 (0.06, 0.17)

Table B.22: Posterior medians (95% Bayesian credible intervals) for parameters of the full (Figure B.5) and approximate (Figure 4.9) stratified SEIR models fit to a simulated Ebola outbreak. The full model explicitly models cross–border transmission via migration of infected individuals, while the approximate model incorporates cross–border transmission through virtual migrations. Subscripts, G, L, S , indicate specific countries, or generic countries A, B . Adjusted reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$. Parameter interpretations and priors are given in Table B.17.

Parameter	Truth	Posterior median (95% BCI)	
		<i>Approximate Model</i>	<i>Full Model</i>
$R_{eff,G}^{(1)}$	1.30	1.24 (1.12, 1.44)	1.23 (1.13, 1.43)
$R_{eff,G}^{(2)}$	1.20	1.28 (1.1, 1.63)	1.27 (1.13, 1.52)
$R_{eff,L}$	2.00	1.69 (1.37, 2.22)	2.07 (1.56, 3.06)
$R_{eff,S}$	1.55	1.54 (1.31, 1.97)	1.52 (1.31, 1.91)
$7/\omega_G$	6.36	5.55 (2.92, 10.44)	3.78 (1.89, 8.21)
$7/\omega_L$	8.24	6 (2.78, 11.2)	8.28 (3.87, 15.19)
$7/\omega_S$	7.00	6.14 (3.26, 11.22)	5.67 (2.98, 10.58)
$7/\mu_G$	7.00	6.19 (4.06, 9.65)	5.83 (3.46, 9.85)
$7/\mu_L$	8.75	6.41 (4.05, 9.95)	8.96 (5.3, 14.7)
$7/\mu_S$	6.36	6.46 (4.17, 10.04)	6.55 (4.25, 9.89)
$1000\alpha_{GL}$	3.00	3.46 (0.13, 17.22)	16.22 (6.32, 35.08)
$1000\alpha_{GS}$	3.00	3.75 (0.14, 18.93)	20.8 (8.14, 43.62)
$1000\alpha_{LG}$	3.00	2.19 (0.12, 8.33)	8.72 (3.29, 19.35)
$1000\alpha_{LS}$	3.00	3.03 (0.22, 11.3)	10.15 (3.44, 22.94)
$1000\alpha_{SG}$	3.00	2.96 (0.16, 13.67)	10.05 (3.46, 22.77)
$1000\alpha_{SL}$	3.00	2.69 (0.12, 11.51)	5.58 (2.02, 14.79)
$\rho_G N_{eff,G}$	12750	11570 (7750, 19320)	11590 (7500, 19560)
$\rho_L N_{eff,L}$	8750	10030 (8070, 14180)	8560 (7220, 11420)
$\rho_S N_{eff,S}$	15000	14860 (11510, 20580)	15420 (11740, 21230)
ϕ_G	50.00	59.78 (27.36, 211.88)	42.87 (22.15, 96.72)
ϕ_L	50.00	46.61 (20.82, 105.31)	46.62 (20.77, 101.46)
ϕ_S	50.00	33.13 (17.76, 60.48)	31.02 (16.93, 55.48)

Table B.23: Effective sample sizes and potential scale reduction factors for the stratified SEIR model with virtual migration of infecteds fit to a simulated Ebola outbreak.

Parameter	ESS	PSRF
$\log \left(R_{eff,G}^{(2)} / R_{eff,G}^{(1)} \right)$	461	1.01
$\log \left(R_{eff,G}^{(2)} \right)$	367	1.01
$\log (R_{eff,L})$	505	1.06
$\log (R_{eff,S})$	759	1.02
$\log (\omega_G / \mu_G)$	1057	1.00
$\log (\omega_L / \mu_L)$	664	1.05
$\log (\omega_S / \mu_S)$	851	1.01
$\log (1 / \mu_G)$	719	1.01
$\log (1 / \mu_L)$	692	1.04
$\log (1 / \mu_S)$	1038	1.01
$\log (1000\alpha_{GL})$	712	1.03
$\log (1000\alpha_{GS})$	813	1.04
$\log (1000\alpha_{LG})$	344	1.13
$\log (1000\alpha_{LS})$	539	1.09
$\log (1000\alpha_{SG})$	494	1.03
$\log (1000\alpha_{SL})$	641	1.10
$\log (\rho_G N_{eff,G})$	525	1.00
$\log (\rho_L N_{eff,L})$	603	1.08
$\log (\rho_S N_{eff,S})$	1020	1.01
$\text{logit}(\rho_G)$	445	1.01
$\text{logit}(\rho_L)$	354	1.04
$\text{logit}(\rho_S)$	540	1.03
$\log(\phi_G)$	479	1.08
$\log(\phi_L)$	999	1.00
$\log(\phi_S)$	1012	1.01

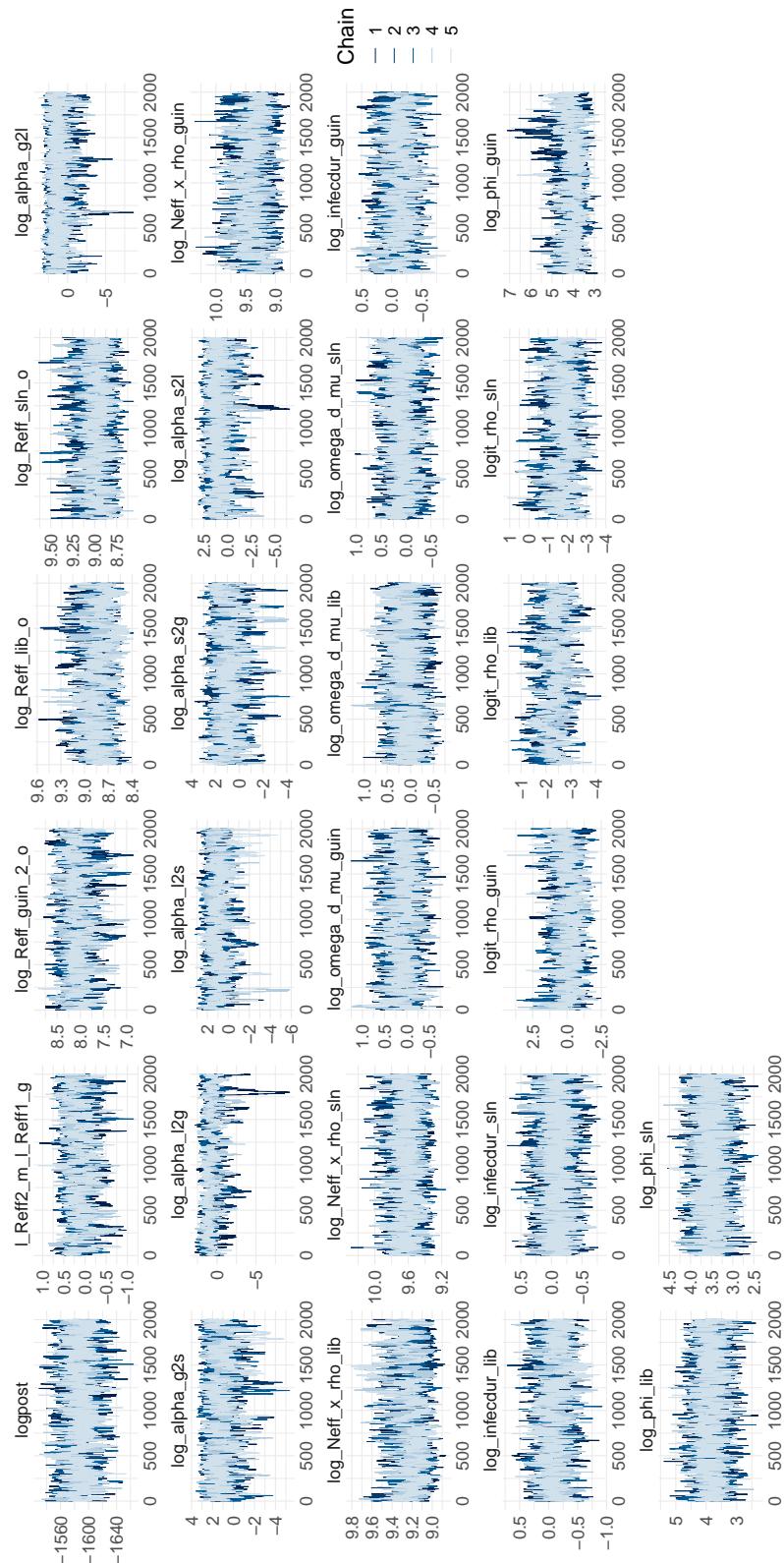


Figure B.6: Posterior traceplots for a stratified SEIR model fit to a simulated Ebola outbreak.

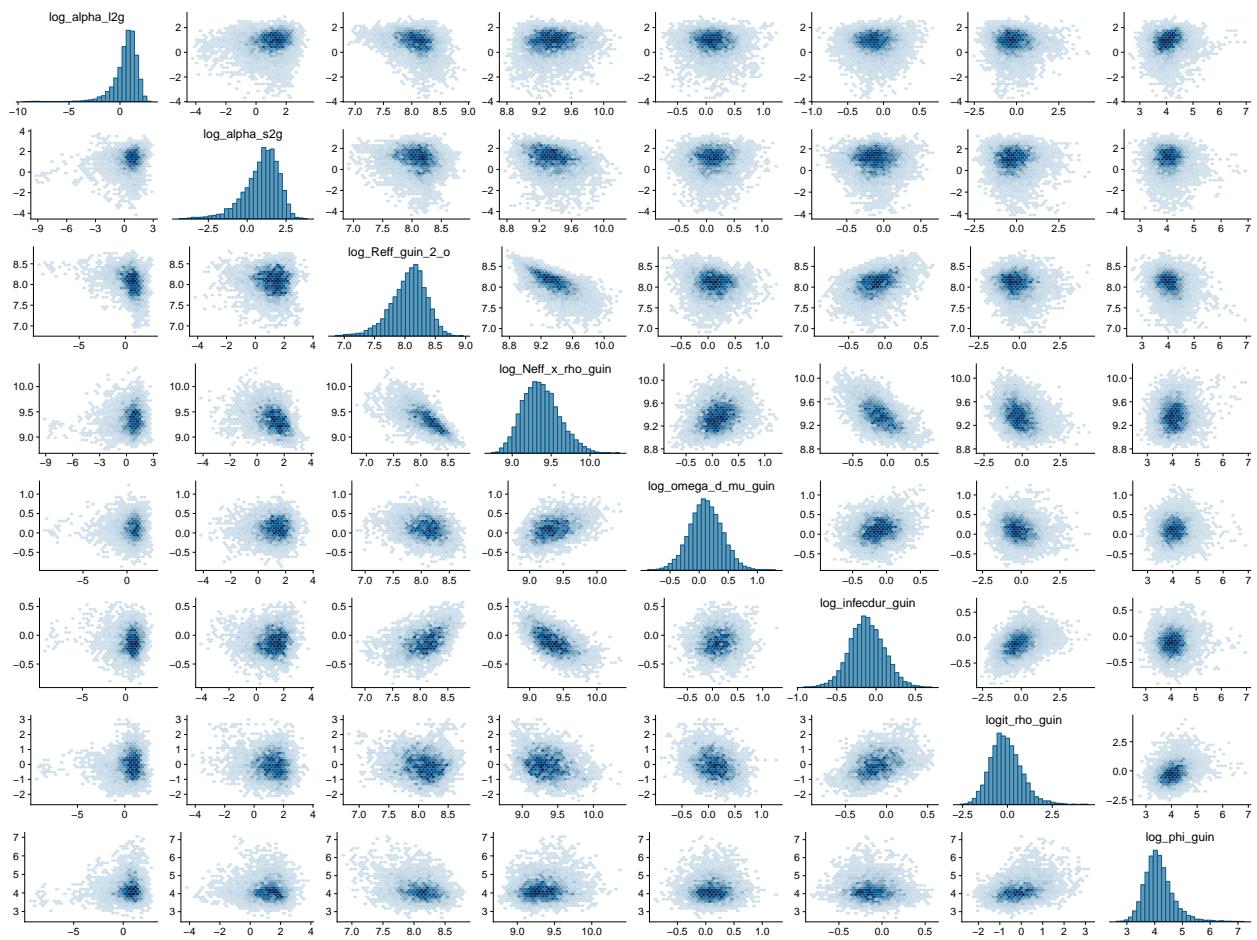


Figure B.7: Scatterplots of Guinea-specific parameters in a stratified SEIR model for a simulated Ebola outbreak.

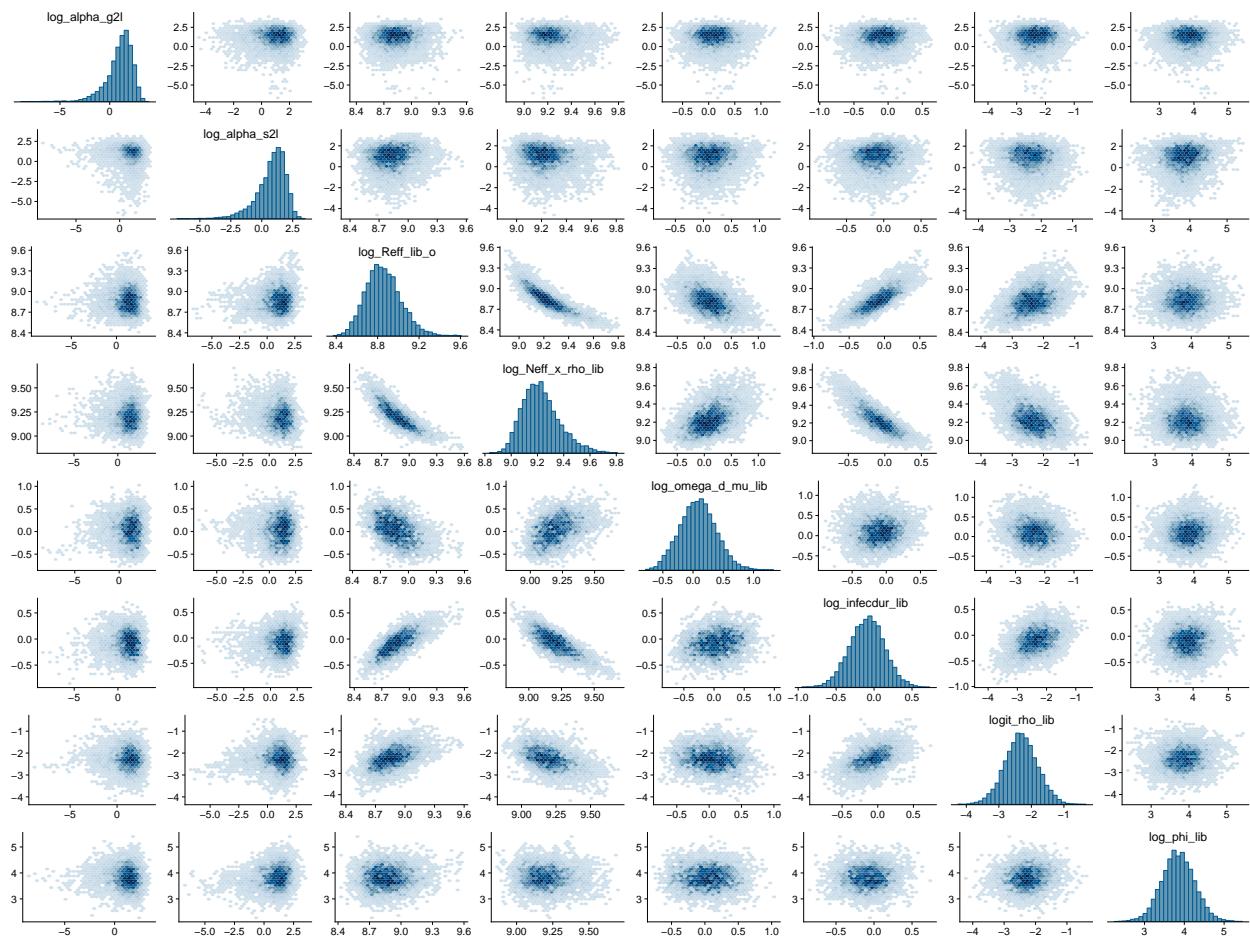


Figure B.8: Scatterplots of Liberia-specific parameters in a stratified SEIR model for a simulated Ebola outbreak.

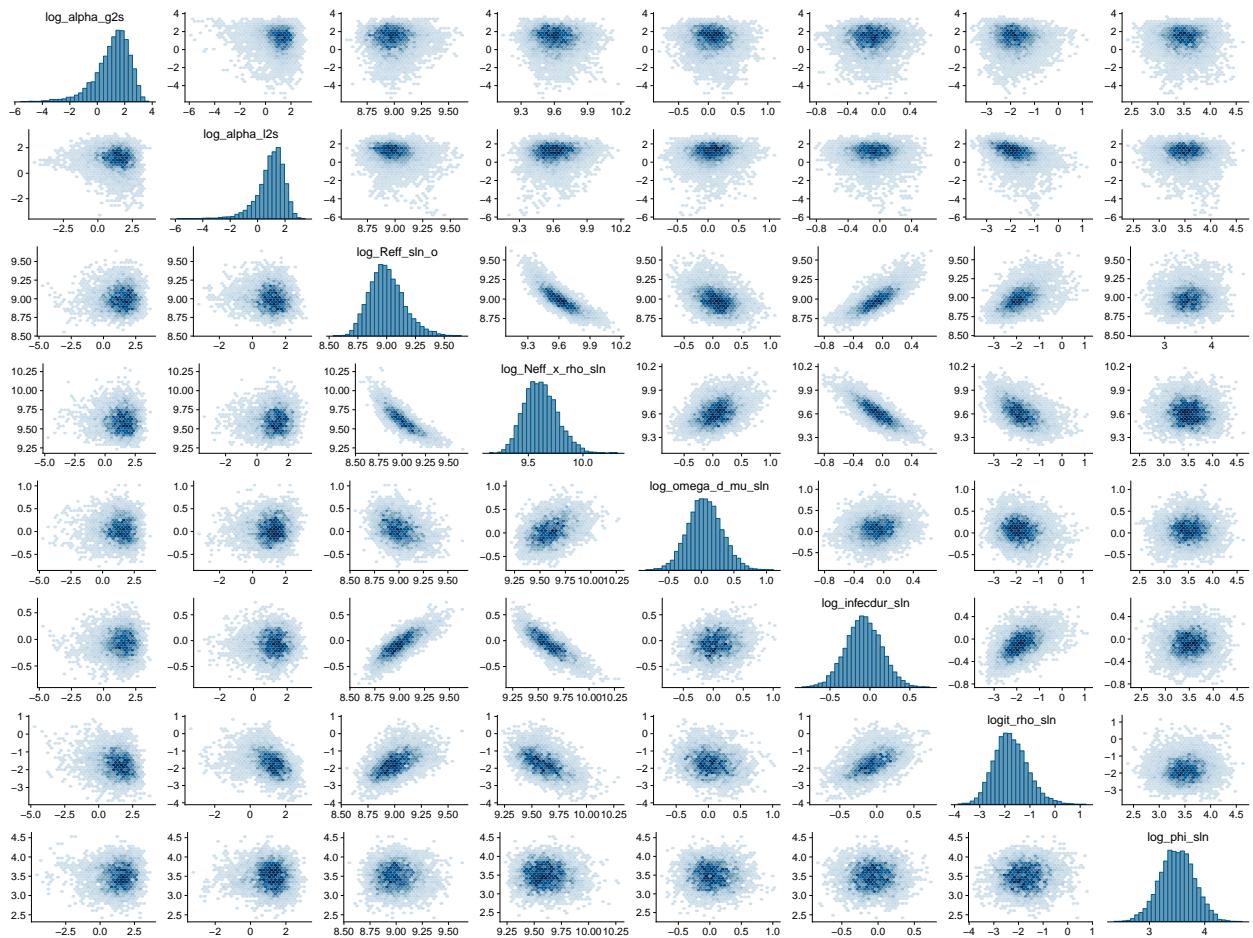


Figure B.9: Scatterplots of Sierra Leone-specific parameters in a stratified SEIR model for a simulated Ebola outbreak.

B.8 Modeling Ebola in West Africa — MCMC Details and Supplementary Results

B.8.1 Details and Supplementary Results for Single-Country Models

Table B.24: Rates of transitions between compartments of single country SEIR models, diagrammed in Figure 4.15, for the spread of Ebola in Guinea, Liberia, and Sierra Leone with constant rate of infectious contact from outside the population. Subscripts in the rates denote model compartments, superscripts denote countries. Subscripts for parameters denote countries. 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. The rate at which susceptibles in Guine become infected is time varying with one changepoint at 33 weeks. Transmission in Liberia and Sierra Leone prior to weeks 10 and 19 is expressed in the initial compartment volumes at those times, but is not modeled explicitly.

From	To	Country	Rate
Susceptible	Exposed	Guinea	$\lambda_{SE}^G(t) = \left(\alpha_G + (\beta_G^{(1)} \mathbb{1}_{\{t < 33\}} + \beta_G^{(2)} \mathbb{1}_{\{t \geq 33\}}) I_G \right) S_G$
Susceptible	Exposed	Liberia	$\lambda_{SE}^L = (\alpha_L + \beta_L I_L) S_L$
Susceptible	Exposed	Sierra Leone	$\lambda_{SE}^S = (\alpha_S + \beta_S I_S) S_S$
Exposed	Infected	Guinea	$\lambda_{EI}^G = \omega_G E_G$
Exposed	Infected	Liberia	$\lambda_{EI}^L(t) = \omega_L E_L$
Exposed	Infected	Sierra Leone	$\lambda_{EI}^S = \omega_S E_S$
Infected	Recovered	Guinea	$\lambda_{IR}^G = \mu_G I_G$
Infected	Recovered	Liberia	$\lambda_{IR}^L = \mu_L I_L$
Infected	Recovered	Sierra Leone	$\lambda_{IR}^S = \mu_S I_S$

Priors and MCMC details

We fit country specific SEIR models via the LNA and ODE to Ebola incidence data from Guinea, Liberia, and Sierra Leone under two prior regimes given in Tables B.26 and B.27. The model fitting procedure and priors were the same for all models. We ran five chains for each model, initialized at random parameter values, for 200,000 iterations per chain. The

model parameters, not including the initial compartment volumes, were jointly updated via MVNSS (see Section 2.2.1). 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 initial compartment volumes were updated jointly with the LNA paths in an EllipSS step. The MCMC alternated between two ElliptSS and one MVNSS update 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 B.3 with the one addition being the log ratio of adjusted reproductive numbers, subtract one, for Guinea, $\log\left((R_{eff,G}^{(2)} - 1)/(R_{eff,G}^{(1)} - 1)\right)$. Convergence was assessed visually by inspection of traceplots of posterior samples, and via potential scale reduction factors (PSRFs) [35] computed via the `coda` R package [173]. The models took between 8 hours and 34 hours to run, depending on the speed of the cluster cores.

Table B.25: Priors for the initial compartment volumes at the times when transmission commenced in each country for single country models fit to data from the Ebola outbreak in West Africa. The initial compartment volumes for each country are assigned independent truncated multivariate normal priors that respect the boundary conditions on the state space of compartment volumes for that country (explained in Section B.2). 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	$(11.8 \times 10^6 - 7.1, 4, 3, 0.1)$
Liberia	$(4.4 \times 10^6 - 5.1, 3, 2, 0.1)$
Sierra Leone	$(7.1 \times 10^6 - 5.1, 3, 2, 0.1)$

Table B.26: Parameters for a single country SEIR models fit to the West Africa Ebola outbreak, prior distributions used in the main analysis, 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 shared. Adjusted reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$.

Parameter	Interpretation	Prior	Median (95% Interval)	References
$R_{eff,G}/R_{eff,G}^{(1)}$	Ratio of adjusted reproduction #s	LogNormal(0, 0.5 ²)	$R_{eff,G}/R_{eff,G}^{(1)} = 1.00$ (0.38, 2.66)	[49, 51, 52, 123]
$R_{eff,G}^{(2)}(t) - 1$	Adjusted reproduction # - 1, $t \geq 33$	LogNormal(log(0.5), 0.75 ²)	$\Rightarrow R_{eff,G}^{(2)} = 1.50(1.11, 3.17)$	[49, 51, 52, 123]
$R_{eff,L} - 1$	Adjusted reproduction # - 1	LogNormal(log(0.5), 0.75 ²)	$\Rightarrow R_{eff,L} = 1.50(1.11, 3.17)$	[49, 51, 52, 123]
$R_{eff,S} - 1$	Adjusted reproduction # - 1	LogNormal(log(0.5), 0.75 ²)	$\Rightarrow R_{eff,S} = 1.50(1.11, 3.17)$	[49, 51, 52, 123]
$\alpha_G, \alpha_L, \alpha_S$	Rate of infectious contact from outside the country	1000 $\alpha_A \sim \text{Exponential}(0.25)$	# migrations per 1000 infected = 2.78 (0.10, 14.76)	[63]
ω_G/ω_G	Rate from $E_G \rightarrow I_G$	LogNormal(0.0, 0.3 ²)	$(1/\mu_G)/(1/\omega_G) = 1.00$ (0.56, 1.80)	[49, 212, 91]
ω_L/ω_L	Rate from $E_L \rightarrow I_L$	LogNormal(0.0, 0.3 ²)	$(1/\mu_L)/(1/\omega_L) = 1.00$ (0.56, 1.80)	[49, 212, 91]
ω_S/ω_S	Rate from $E_S \rightarrow I_S$	LogNormal(0.0, 0.3 ²)	$(1/\mu_S)/(1/\omega_S) = 1.00$ (0.56, 1.80)	[49, 212, 91]
$1/\mu_G$	Rate from $E_G \rightarrow I_G$	LogNormal(0.3, 0.2 ²)	$7/\mu_G = 9.45$ (6.38, 13.98)	[49, 212, 91]
$1/\mu_L$	Rate from $E_L \rightarrow I_L$	LogNormal(0.3, 0.2 ²)	$7/\mu_L = 9.45$ (6.38, 13.98)	[49, 212, 91]
$1/\mu_S$	Rate from $E_S \rightarrow I_S$	LogNormal(0.3, 0.2 ²)	$7/\mu_S = 9.45$ (6.38, 13.98)	[49, 212, 91]
$N_{eff,G}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^G, N_G/4$)	$N_{eff,G} \in (3627, 2.95 \times 10^6)$	Scale of counts
$N_{eff,L}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^L, N_L/4$)	$N_{eff,L} \in (4994, 1.1 \times 10^6)$	Scale of counts
$N_{eff,S}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^S, N_S/4$)	$N_{eff,S} \in (11317, 1.775 \times 10^6)$	Scale of counts
ρ_G	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_G = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ρ_L	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_L = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ρ_S	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_S = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ϕ_G, ϕ_L, ϕ_S	NegBinomial overdispersion	LogNormal(0, 5 ²)	= 1, $(5.5 \times 10^{-5}, 1.8 \times 10^4)$	Diffuse.

Table B.27: Parameters for a country-specific SEIR models fit to the West Africa Ebola outbreak, prior distributions used in the supplementary analysis, 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 shared. Adjusted reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$. Prior distributions for parameters controlling the outbreak dynamics for the supplementary models were more diffuse vis-a-vis the priors in the main analysis.

Parameter	Interpretation	Prior	Median (95% Interval)	References
$R_{eff,G}/R_{eff,G}^{(1)}$	Ratio of adjusted reproduction #s	LogNormal(0, 0.5 ²)	$R_{eff,G}/R_{eff,G}^{(1)} = 1.00 (0.38, 2.66)$	[49, 51, 52, 123]
$R_{eff,G}^{(2)}(t) - 1$	Adjusted reproduction # - 1, $t \geq 33$	LogNormal(log(0.5), 1.5 ²)	$\Rightarrow R_{eff,G}^{(2)} = 1.50(1.02, 10.46)$	[49, 51, 52, 123]
$R_{eff,L} - 1$	Adjusted reproduction # - 1	LogNormal(log(0.5), 1.5 ²)	$\Rightarrow R_{eff,L} = 1.50(1.02, 10.46)$	[49, 51, 52, 123]
$R_{eff,S} - 1$	Adjusted reproduction # - 1	LogNormal(log(0.5), 1.5 ²)	$\Rightarrow R_{eff,S} = 1.50(1.02, 10.46)$	[49, 51, 52, 123]
$1000\alpha_A$	Rate of infectious contact from outside country A	Exponential(0.125)	# migrations per 1000 infected = 5.55 (0.20, 29.51)	[63]
ω_G/ω_G	Rate from $E_G \rightarrow I_G$	LogNormal(0.0, 0.3 ²)	$(1/\mu_G)/(1/\omega_G) = 1.00 (0.56, 1.80)$	[49, 212, 91]
ω_L/ω_L	Rate from $E_L \rightarrow I_L$	LogNormal(0.0, 0.3 ²)	$(1/\mu_L)/(1/\omega_L) = 1.00 (0.56, 1.80)$	[49, 212, 91]
ω_S/ω_S	Rate from $E_S \rightarrow I_S$	LogNormal(0.0, 0.3 ²)	$(1/\mu_S)/(1/\omega_S) = 1.00 (0.56, 1.80)$	[49, 212, 91]
$1/\mu_G$	Rate from $E_G \rightarrow I_G$	LogNormal(0.3, 0.35 ²)	$7/\mu_G = 9.45 (4.76, 18.76)$	[49, 212, 91]
$1/\mu_L$	Rate from $E_L \rightarrow I_L$	LogNormal(0.3, 0.35 ²)	$7/\mu_L = 9.45 (4.76, 18.76)$	[49, 212, 91]
$1/\mu_S$	Rate from $E_S \rightarrow I_S$	LogNormal(0.3, 0.35 ²)	$7/\mu_S = 9.45 (4.76, 18.76)$	[49, 212, 91]
$N_{eff,G}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^G, NG/4$)	$N_{eff,G} \in (3627, 2.95 \times 10^6)$	Scale of counts
$N_{eff,L}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^L, NL/4$)	$N_{eff,L} \in (4994, 1.1 \times 10^6)$	Scale of counts
$N_{eff,S}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^S, NS/4$)	$N_{eff,S} \in (11317, 1.775 \times 10^6)$	Scale of counts
ρ_G	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_G = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ρ_L	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_L = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ρ_S	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_S = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ϕ_G, ϕ_L, ϕ_S	Neg.Binomial overdispersion	LogNormal(0, 5 ²)	$\Rightarrow = 1, (5.5 \times 10^{-5}, 1.8 \times 10^4)$	Diffuse.

Additional LNA results

Table B.28: Posterior estimates of initial numbers of exposed, infected, and recovered individuals for the main country specific SEIR LNA models fit to Ebola data from the West Africa outbreak under tight priors. The effective number of susceptibles is equal to the effective population size, reported in Table B.29, minus the numbers of exposed, infected, and recovered individuals, but is not reported since the effective population size is not marginally identifiable.

Parameter	Estimate
$E_{0,G}$	3.92 (0.7, 7.71)
$I_{0,G}$	3.01 (0.39, 6.26)
$R_{0,G}$	0.1 (0.06, 0.15)
$E_{0,L}$	3.12 (0.44, 6.41)
$I_{0,L}$	2.24 (0.29, 4.81)
$R_{0,L}$	0.11 (0.06, 0.16)
$E_{0,S}$	1.99 (0.17, 4.73)
$I_{0,S}$	1.98 (0.22, 4.5)
$R_{0,S}$	0.09 (0.05, 0.14)

Table B.29: Posterior medians (95% Bayesian credible intervals) for parameters of the country-specific SEIR LNA models (Figure 4.15) fit to Ebola outbreak data. Subscripts, G, L, S , indicate specific countries, or generic countries A, B . Adjusted reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$. Parameter interpretations and priors are given in Tables B.26 and B.27. The more informative of the two prior regimes was presented as the main model in Section 4.4.1.

Parameter	Posterior median (95% BCI)	
	Tight Priors	Loose Priors
$R_{eff,G}^{(1)}$	1.2 (1.1, 1.4)	1.1 (1.0, 1.4)
$R_{eff,G}^{(2)}$	1.4 (1.2, 1.8)	1.2 (1.0, 1.9)
$7/\omega_G$	8.4 (4.0, 17.2)	5.6 (2.2, 18.3)
$7/\mu_G$	9.0 (5.9, 13.4)	6.4 (3.1, 16.0)
$1000\alpha_G$	0.2 (0.0, 0.6)	0.1 (0.0, 0.7)
$\rho_G N_{eff,G}$	7200 (5100, 11900)	9800 (4900, 34600)
ϕ_G	4.6 (2.8, 9.4)	5.8 (3.0, 17.6)
$R_{eff,L}$	2.2 (1.7, 3.3)	2.7 (1.6, 4.3)
$7/\omega_L$	10.5 (4.8, 19.6)	13.7 (4.6, 23.1)
$7/\mu_L$	9.7 (6.6, 14.1)	11.7 (5.7, 20.3)
$1000\alpha_L$	0.0 (0.0, 0.06)	0.0 (0.0, 0.1)
$\rho_L N_{eff,L}$	5700 (4600, 7300)	5300 (4400, 7800)
ϕ_L	8.1 (3.7, 17.8)	8.6 (4.0, 19.2)
$R_{eff,S}$	1.3 (1.2, 1.5)	1.2 (1.1, 1.4)
$7/\omega_S$	4.5 (2.6, 7.4)	3.5 (1.9, 6.0)
$7/\mu_S$	6.2 (4.5, 8.5)	4.3 (2.6, 6.7)
$1000\alpha_S$	0.0 (0, 0.1)	0.0 (0, 0.1)
$\rho_S N_{eff,S}$	24300 (19100, 32800)	30700 (22000, 48700)
ϕ_S	46 (22, 112)	55 (26, 150)

Table B.30: Effective sample sizes and potential scale reduction factors for the main stratified SEIR LNA model with virtual migration of infecteds fit to data from the West Africa Ebola outbreak under tight priors.

Parameter	ESS	PSRF
$\log(1000\alpha_G)$	1512	1.01
$\log(R_{adj,G}^{(2)}/R_{adj,G}^{(1)})$	2992	1.00
$\log(R_{adj,G}^{(2)})$	2573	1.00
$\log(\omega_G/\mu_G)$	4953	1.00
$\log(1/mu_G)$	3323	1.00
$\text{logit}(\rho_G)$	2564	1.00
$\log(\phi_G)$	1865	1.01
$\log(N_{eff,G}\rho_G)$	1335	1.01
$\log(1000\alpha_L)$	1812	1.00
$\log(R_{adj,L}^{(2)})$	3595	1.00
$\log(\omega_L/\mu_L)$	3061	1.00
$\log(1/mu_L)$	4278	1.00
$\text{logit}(\rho_L)$	3495	1.00
$\log(\phi_L)$	1534	1.00
$\log(N_{eff,L}\rho_L)$	4333	1.00
$\log(1000\alpha_S)$	1503	1.00
$\log(R_{adj,S}^{(2)})$	949	1.01
$\log(\omega_S/\mu_S)$	2268	1.00
$\log(1/mu_S)$	1432	1.00
$\text{logit}(\rho_S)$	958	1.01
$\log(\phi_S)$	2124	1.00
$\log(N_{eff,S}\rho_S)$	969	1.00

Posterior distributions of single country Ebola SEIR model parameters

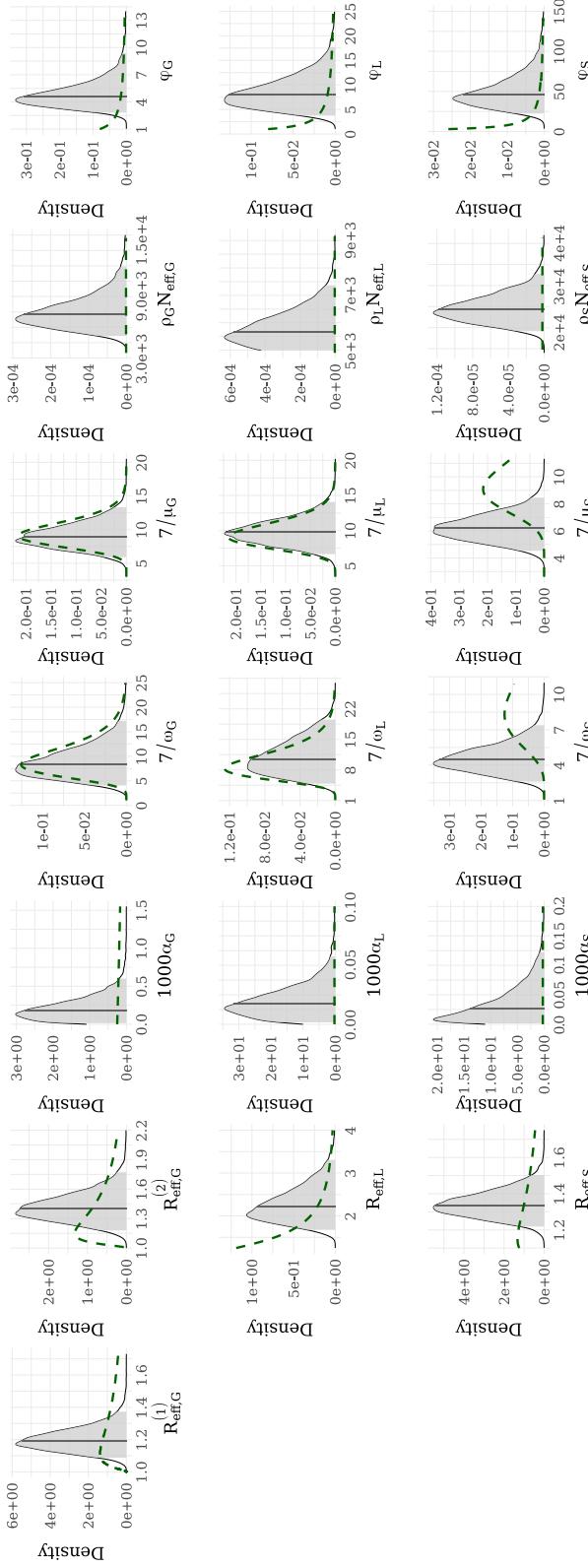


Figure B.10: Posterior distributions of parameters for the main country-specific SEIR LNA models fit to data from the West Africa Ebola outbreak. We show posterior medians (solid gray lines), 95% Bayesian credible intervals (light gray areas under the posterior densities), prior densities (induced priors for the reporting rate and latent period durations) over the posterior ranges (dashed green curves). $R_{adj} = \beta N_{eff}/\mu$ is the adjusted reproductive number, where β is the per-contact infection rate, N_{eff} is the effective population size, and μ is the recovery rate. The latent and infectious period durations, $7/\omega$ and $7/\mu$, respectively, are given in days. The scaled effective number of baseline infectious contacts in country A is $1000\alpha_A$. The effective reporting rate is ρN_{eff} , and ϕ is the negative binomial over-dispersion parameter. Subscripts indicate countries.

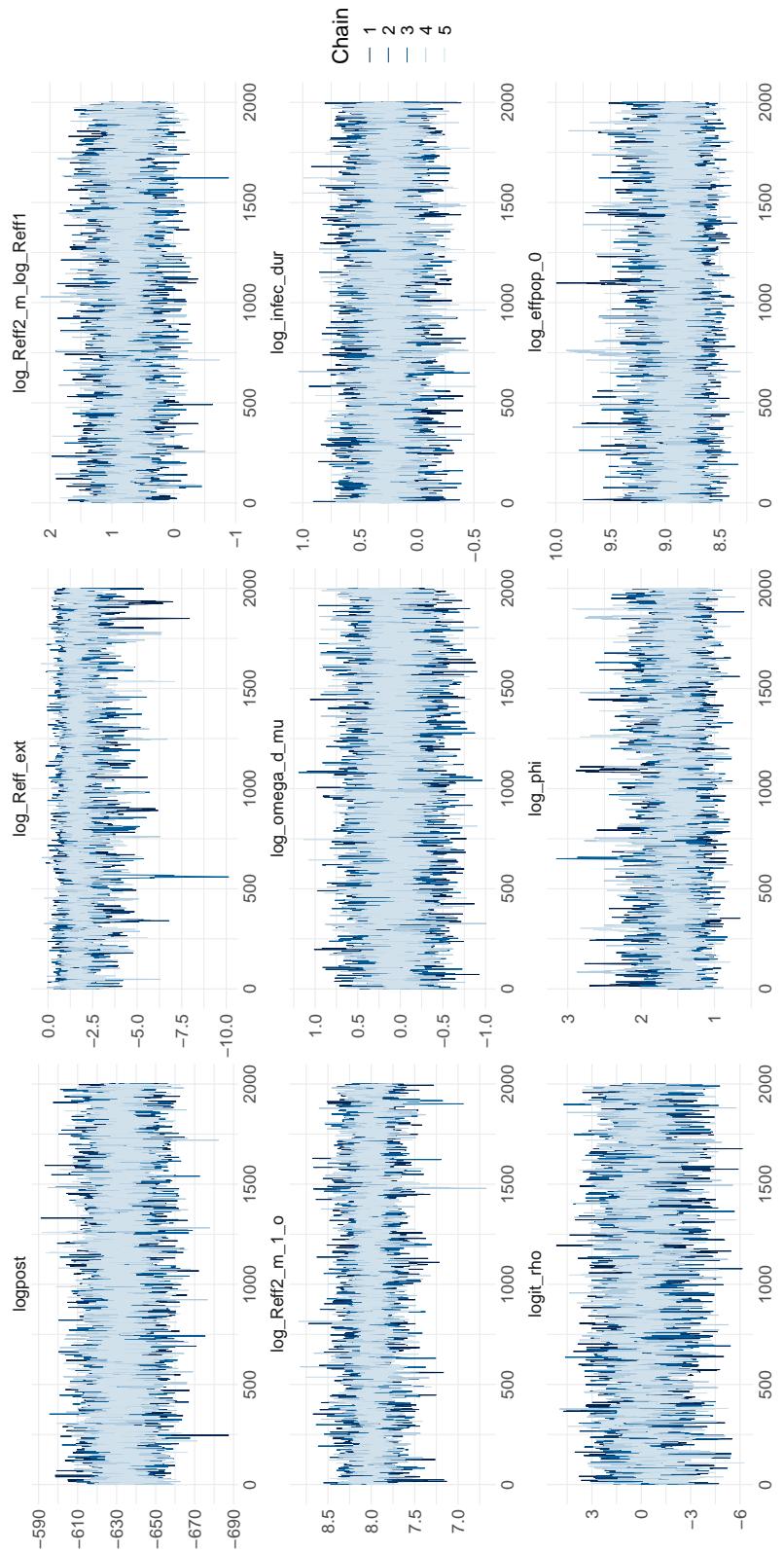


Figure B.11: Posterior traceplots for the SEIR LNA model fit to data from the Ebola outbreak in Guinea.

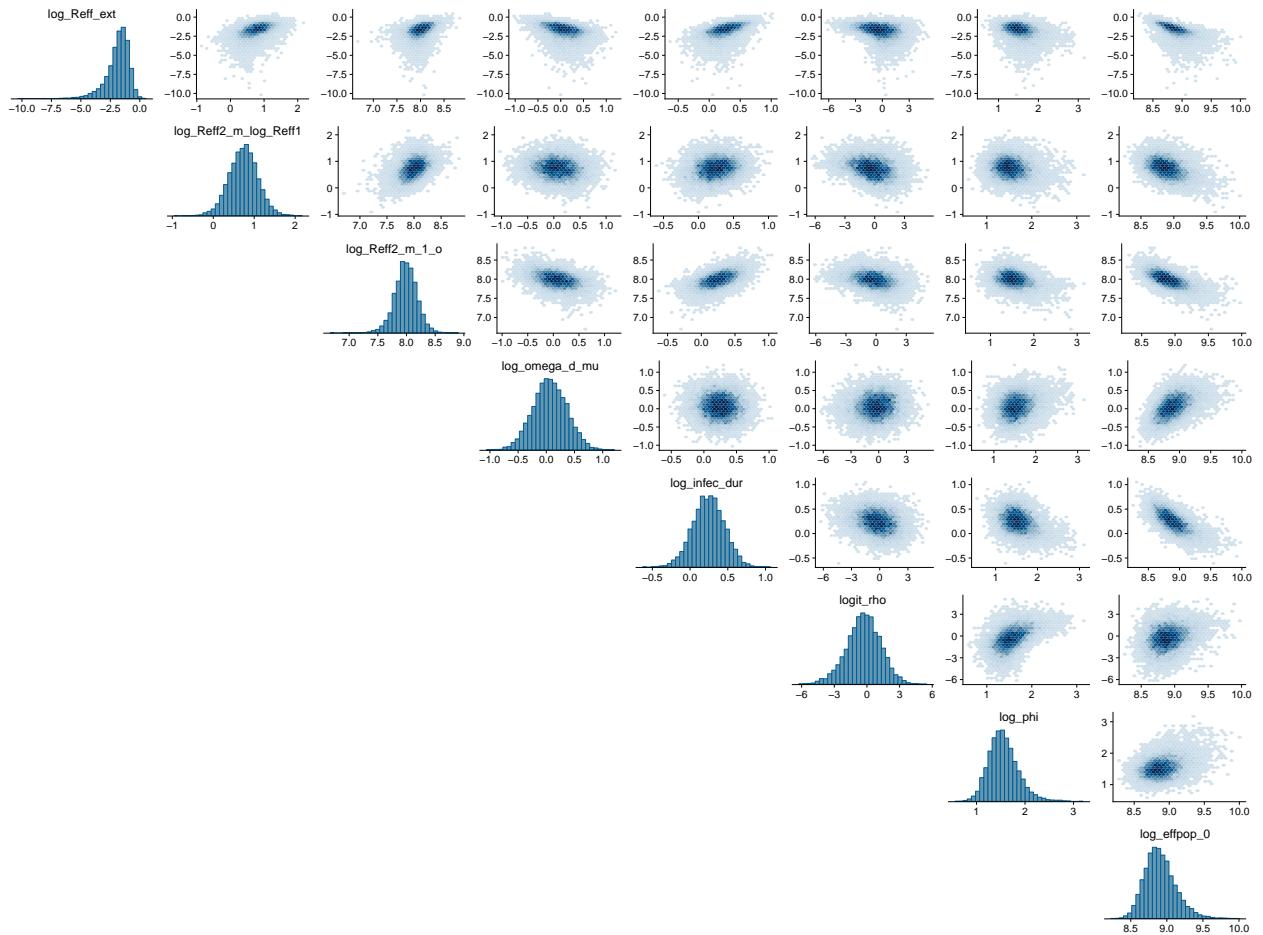


Figure B.12: Scatterplots of parameters for the SEIR LNA model fit to data from the Ebola outbreak in Guinea.

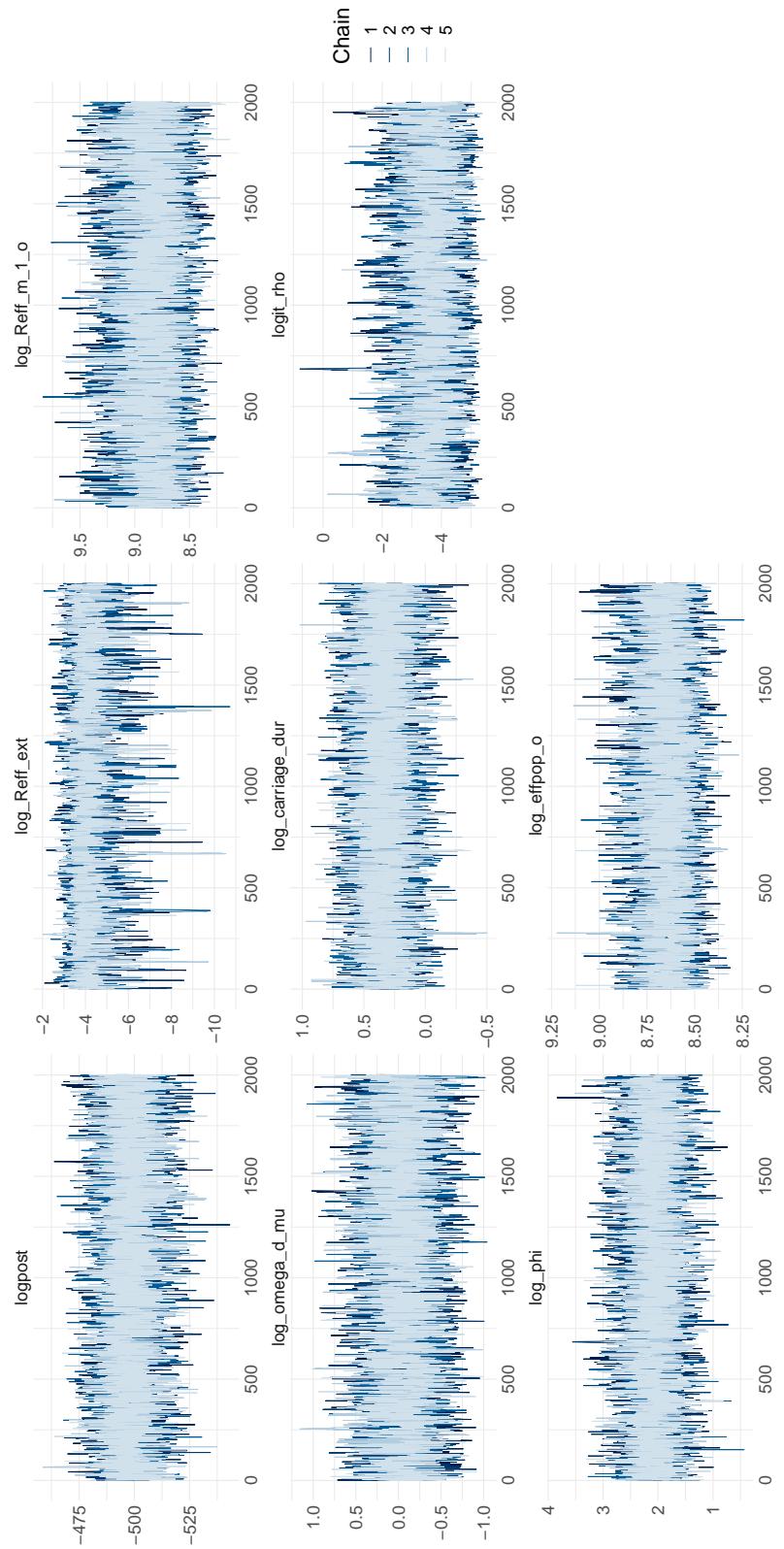


Figure B.13: Posterior traceplots for the SEIR LNA model fit to data from the Ebola outbreak in Liberia.

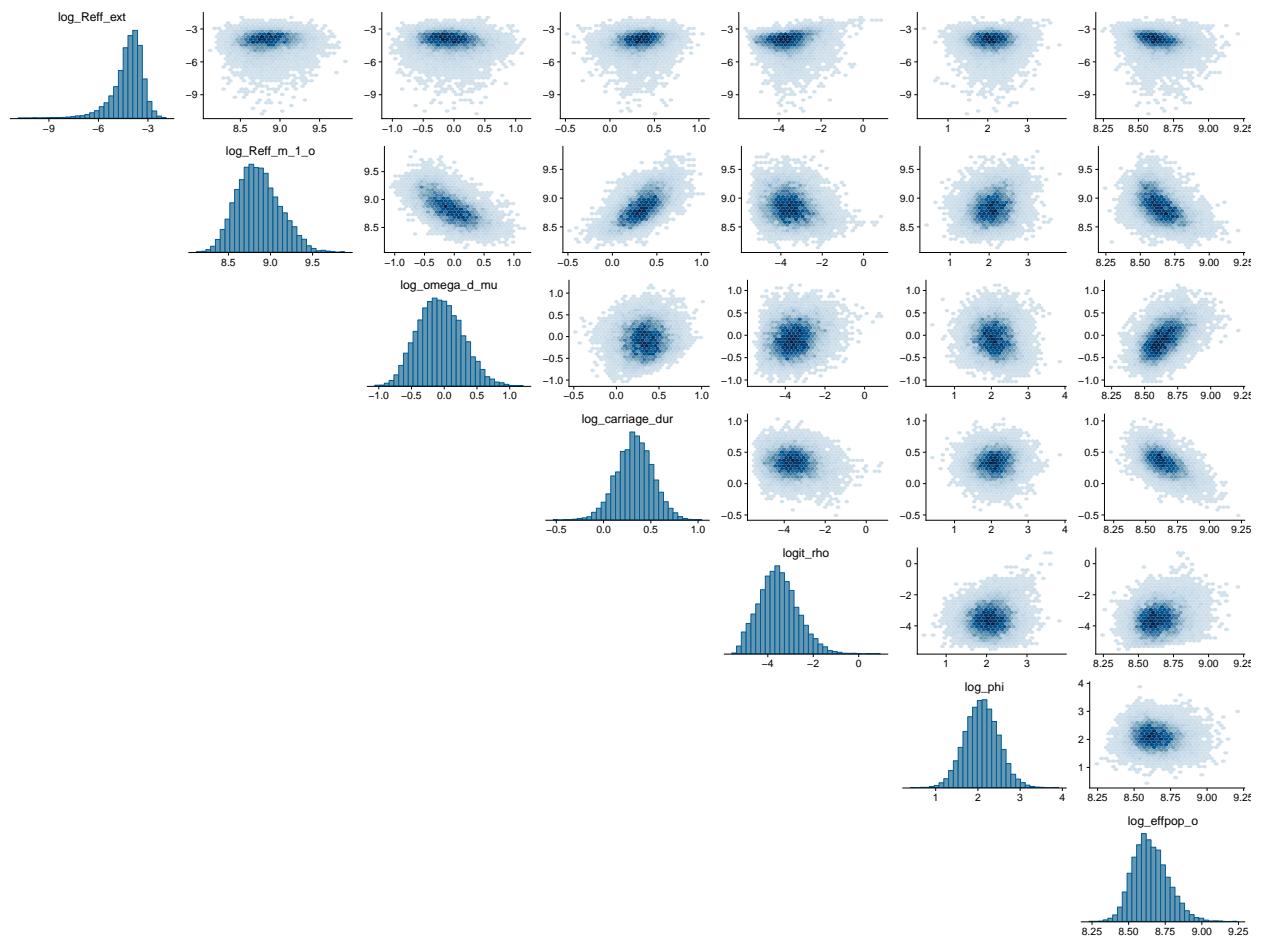


Figure B.14: Scatterplots of parameters for the SEIR LNA model fit to data from the Ebola outbreak in Liberia.

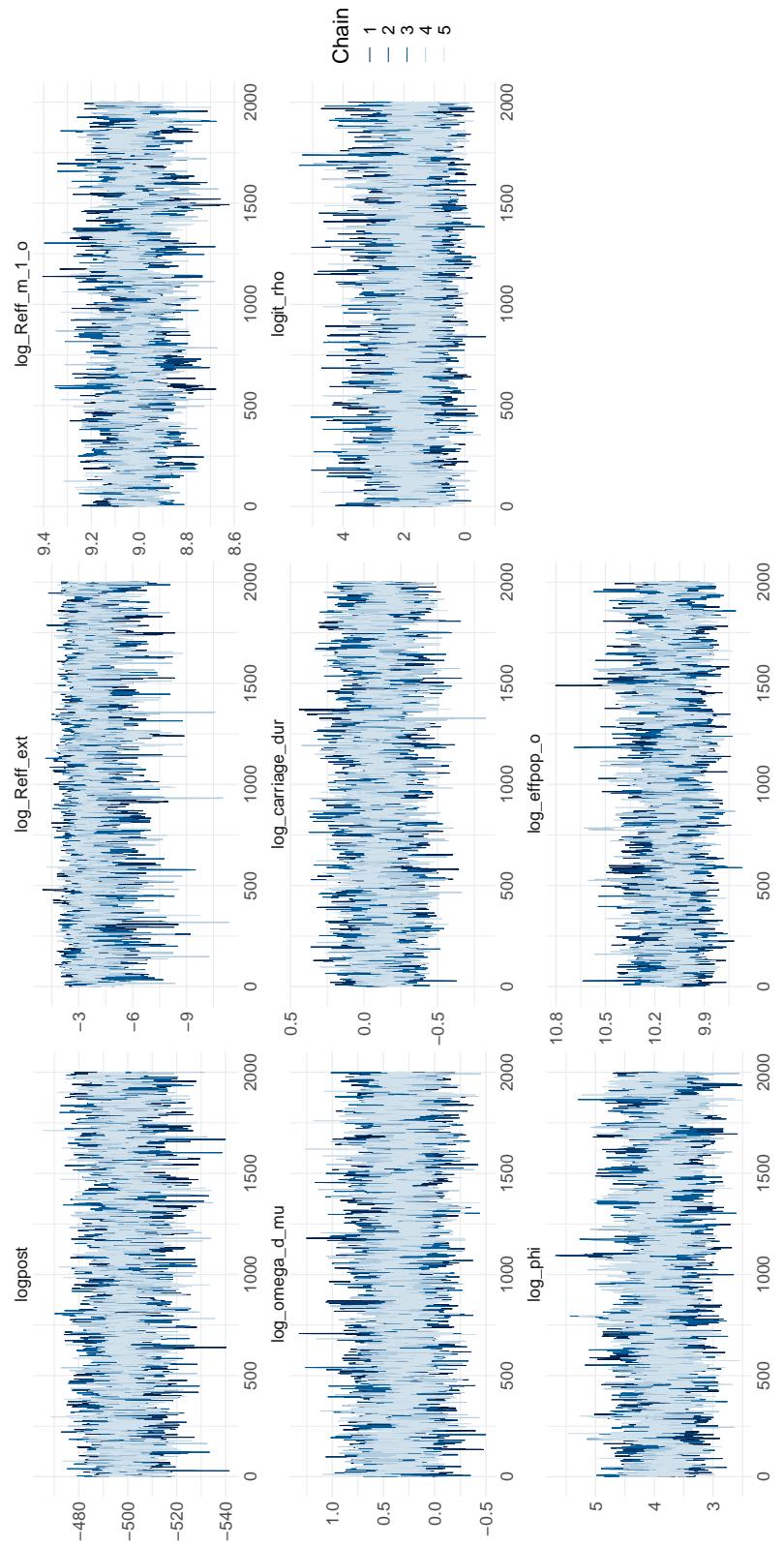


Figure B.15: Posterior traceplots for the SEIR model LNA fit to data from the Ebola outbreak in Sierra Leone.

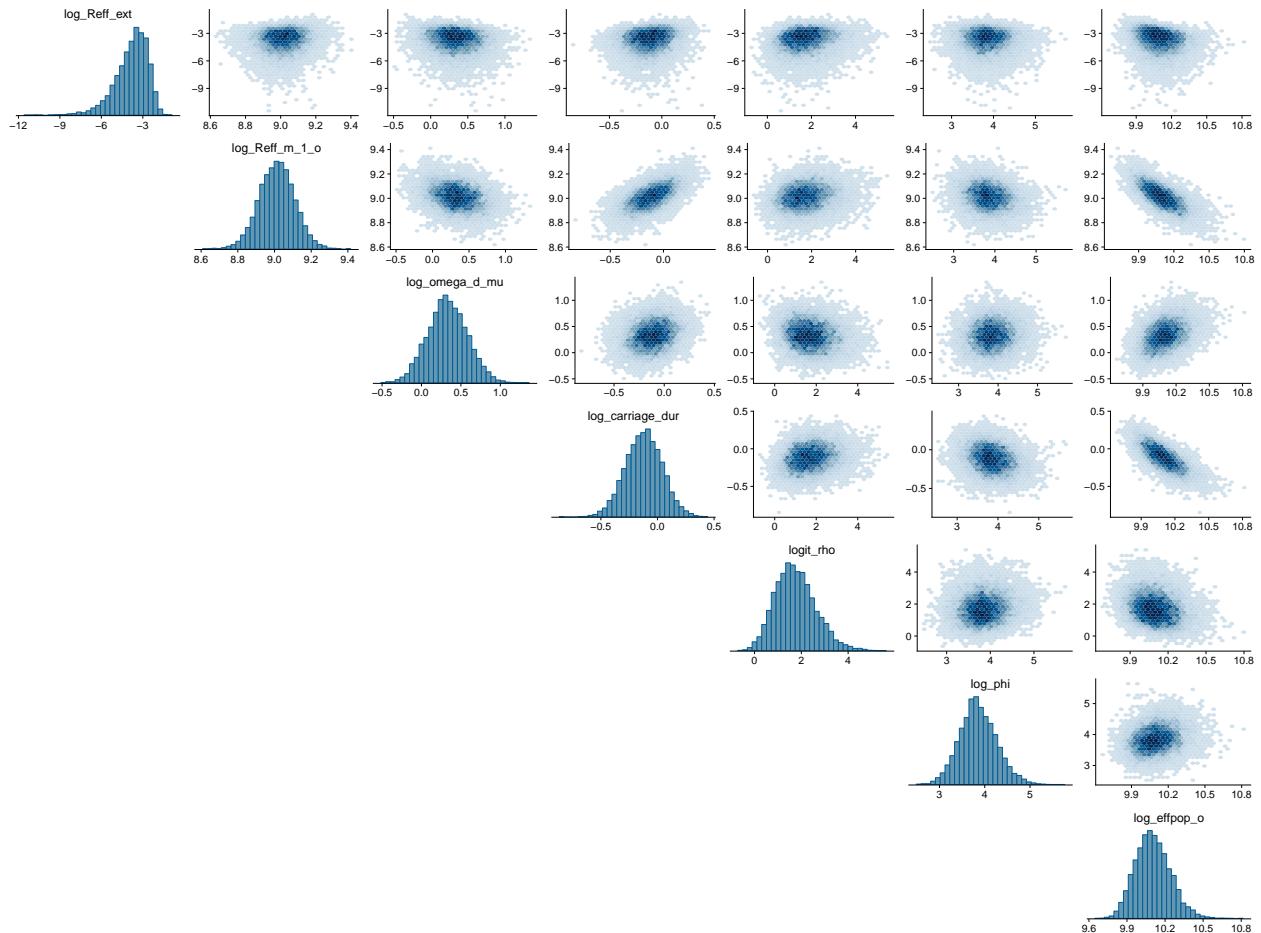


Figure B.16: Scatterplots of parameters for the SEIR LNA model fit to data from the Ebola outbreak in Sierra Leone.

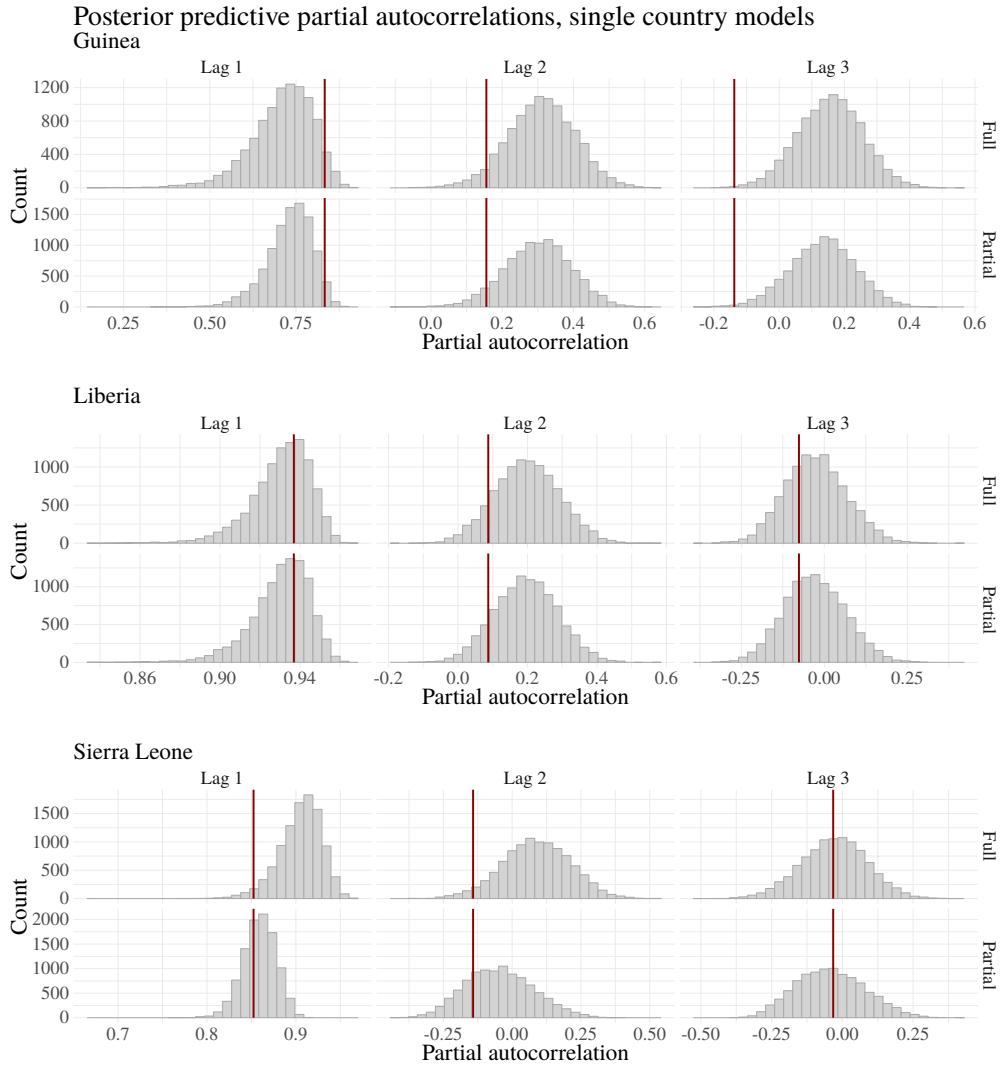


Figure B.17: Distributions of partial autocorrelations at lags 1, 2, and 3 for datasets generated under the full and partial posterior predictive distributions for the main country-specific SEIR LNA models fit to data from the West Africa Ebola outbreak. Vertical red lines are the partial autocorrelations for the observed incidence data at the respective lags.

Posterior distributions of LNA draws for single country Ebola models

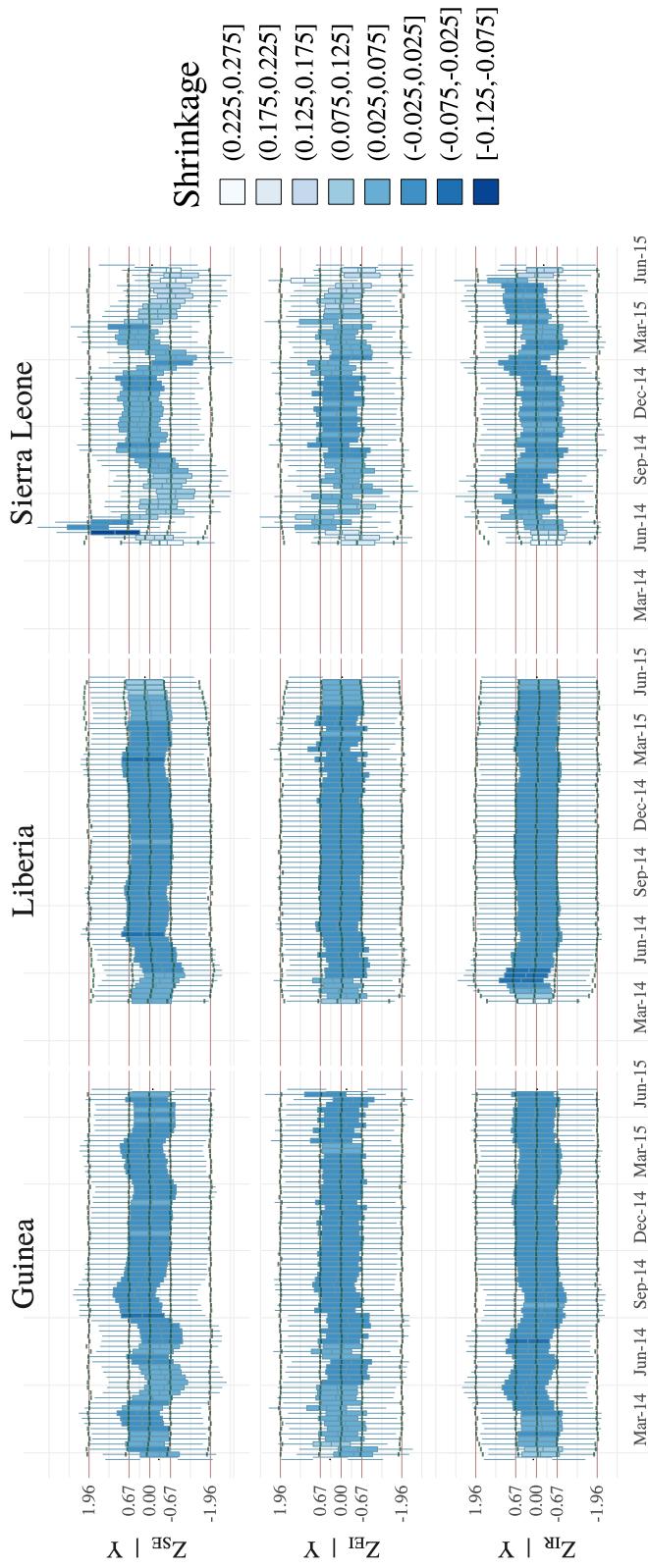


Figure B.18: Posterior distributions of the LNA draws for exposures, infections, and recoveries (blue boxplots) in the main country-specific SEIR LNA models fit data from the Ebola outbreak in Guinea, Liberia, and Sierra Leone. The lower and upper whisker tips correspond to the 2.5th and 97.5th posterior quantiles, the lower and upper hinges to the 25th and 75th quantiles, and the middle hash mark to the posterior median. The solid red lines are the theoretical quantiles of the posterior predictive distribution (or equivalently, the prior distribution) of the LNA draws, drawn at the quantiles of a standard normal distribution corresponding to the boxplot quantiles. The green ticks are the estimated quantiles of the posterior predictive distributions of the LNA draws, accounting for boundary conditions on the state space of the latent process and obtained by simulating LNA paths from the posterior predictive distribution. The posterior distributions of LNA draws are shaded according to the level of posterior shrinkage, computed as one minus the ratio of standard deviations of LNA draws in the posterior and prior.

Additional ODE results

Table B.31: Posterior estimates of initial numbers of exposed, infected, and recovered individuals for the main country specific SEIR ODE models fit to Ebola data from the West Africa outbreak under tight priors. The effective number of susceptibles is equal to the effective population size, reported in Table B.32, minus the numbers of exposed, infected, and recovered individuals, but is not reported since the effective population size is not marginally identifiable.

Parameter	Estimate
$E_{0,G}$	4 (0.6, 7.9)
$I_{0,G}$	3 (0.3, 6.3)
$R_{0,G}$	0.1 (0.1, 0.2)
$E_{0,L}$	2.6 (0.2, 6)
$I_{0,L}$	1.8 (0.1, 4.4)
$R_{0,L}$	0.1 (0, 0.2)
$E_{0,S}$	9.6 (3.5, 15.8)
$I_{0,S}$	5.1 (1.1, 9.3)
$R_{0,S}$	0.3 (0, 0.8)

Table B.32: Posterior medians (95% Bayesian credible intervals) for parameters of the country-specific SEIR ODE models (Figure 4.15) fit to Ebola outbreak data. Subscripts, G, L, S , indicate specific countries, or generic countries A, B . Adjusted reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$. Parameter interpretations and priors are given in Tables B.26 and B.27. The more informative of the two prior regimes was presented as the main model in Section 4.4.1.

Parameter	Posterior median (95% BCI)	
	Tight Priors	Loose Priors
$R_{eff,G}^{(1)}$	1.2 (1.1, 1.4)	1.2 (1.1, 1.5)
$R_{eff,G}^{(2)}$	1.5 (1.3, 1.9)	1.6 (1.2, 2.3)
$7/\omega_G$	10.4 (5.1, 19.9)	11.3 (4.8, 26.3)
$7/\mu_G$	10 (6.8, 14.4)	11 (5.7, 21)
$1000\alpha_G$	0.3 (0.1, 0.6)	0.3 (0.1, 1)
$\rho_G N_{eff,G}$	6500 (4800, 9400)	6200 (4300, 10200)
ϕ_G	3.7 (2.5, 5.5)	3.7 (2.5, 5.4)
$R_{eff,L}$	2.4 (1.7, 3.5)	2.9 (1.8, 4.8)
$7/\omega_L$	11.9 (5.6, 20.5)	15.8 (6.7, 24.3)
$7/\mu_L$	10.2 (7.1, 14.6)	12.9 (7.1, 21.7)
$1000\alpha_L$	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
$\rho_L N_{eff,L}$	5500 (4500, 7000)	5200 (4400, 6700)
ϕ_L	7.6 (3.6, 16.5)	8.6 (3.9, 19.2)
$R_{eff,S}$	1.8 (1.5, 2.2)	1.9 (1.4, 2.8)
$7/\omega_S$	10.8 (5.8, 18.2)	12.5 (5.7, 22.6)
$7/\mu_S$	10.1 (6.9, 14.3)	11.6 (6.1, 20.7)
$1000\alpha_S$	0.4 (0.1, 0.8)	0.5 (0.1, 1.1)
$\rho_S N_{eff,S}$	16700 (13300, 22300)	15600 (12000, 22900)
ϕ_S	5.2 (2.9, 9)	5.2 (2.9, 9)

Table B.33: Effective sample sizes and potential scale reduction factors for the main stratified SEIR ODE model with virtual migration of infecteds fit to data from the West Africa Ebola outbreak under tight priors.

Parameter	ESS	PSRF
$\log(1000\alpha_G)$	7815	1.00
$\log(R_{adj,G}^{(2)}/R_{adj,G}^{(1)})$	9581	1.00
$\log(R_{adj,G}^{(2)})$	10113	1.00
$\log(\omega_G/\mu_G)$	9720	1.00
$\log(1/mu_G)$	10004	1.00
$\text{logit}(\rho_G)$	9839	1.00
$\log(\phi_G)$	8836	1.01
$\log(N_{eff,G}\rho_G)$	9390	1.01
$\log(1000\alpha_L)$	4633	1.00
$\log(R_{adj,L}^{(2)})$	7385	1.00
$\log(\omega_L/\mu_L)$	5803	1.00
$\log(1/mu_L)$	9443	1.00
$\text{logit}(\rho_L)$	7308	1.00
$\log(\phi_L)$	4723	1.00
$\log(N_{eff,L}\rho_L)$	8263	1.00
$\log(1000\alpha_S)$	4096	1.00
$\log(R_{adj,S}^{(2)})$	10004	1.01
$\log(\omega_S/\mu_S)$	9824	1.00
$\log(1/mu_S)$	9330	1.00
$\text{logit}(\rho_S)$	9798	1.01
$\log(\phi_S)$	6459	1.00
$\log(N_{eff,S}\rho_S)$	10005	1.00

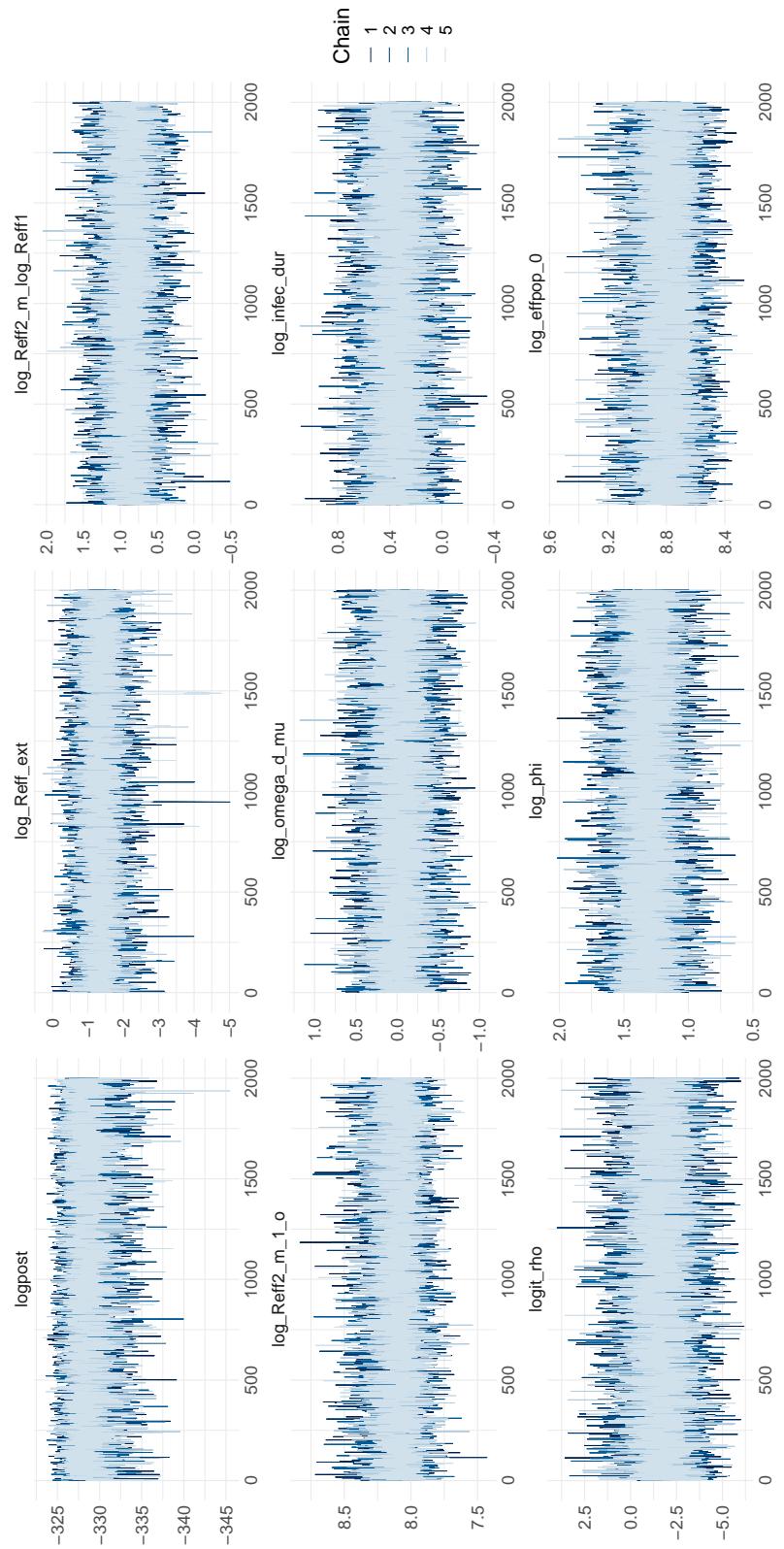


Figure B.19: Posterior traceplots for the SEIR ODE model fit to data from the Ebola outbreak in Guinea.

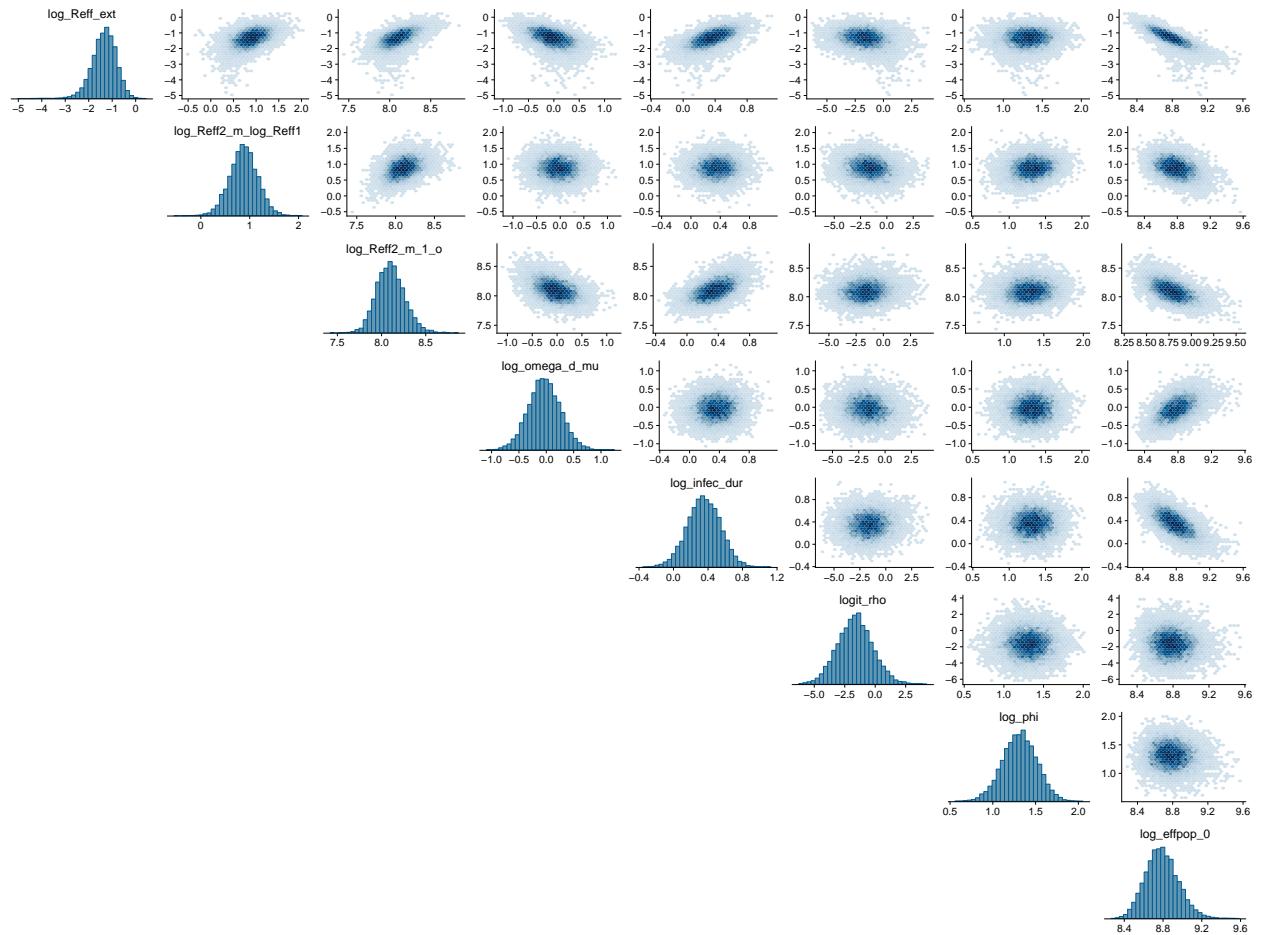


Figure B.20: Scatterplots of parameters for the SEIR ODE model fit to data from the Ebola outbreak in Guinea.

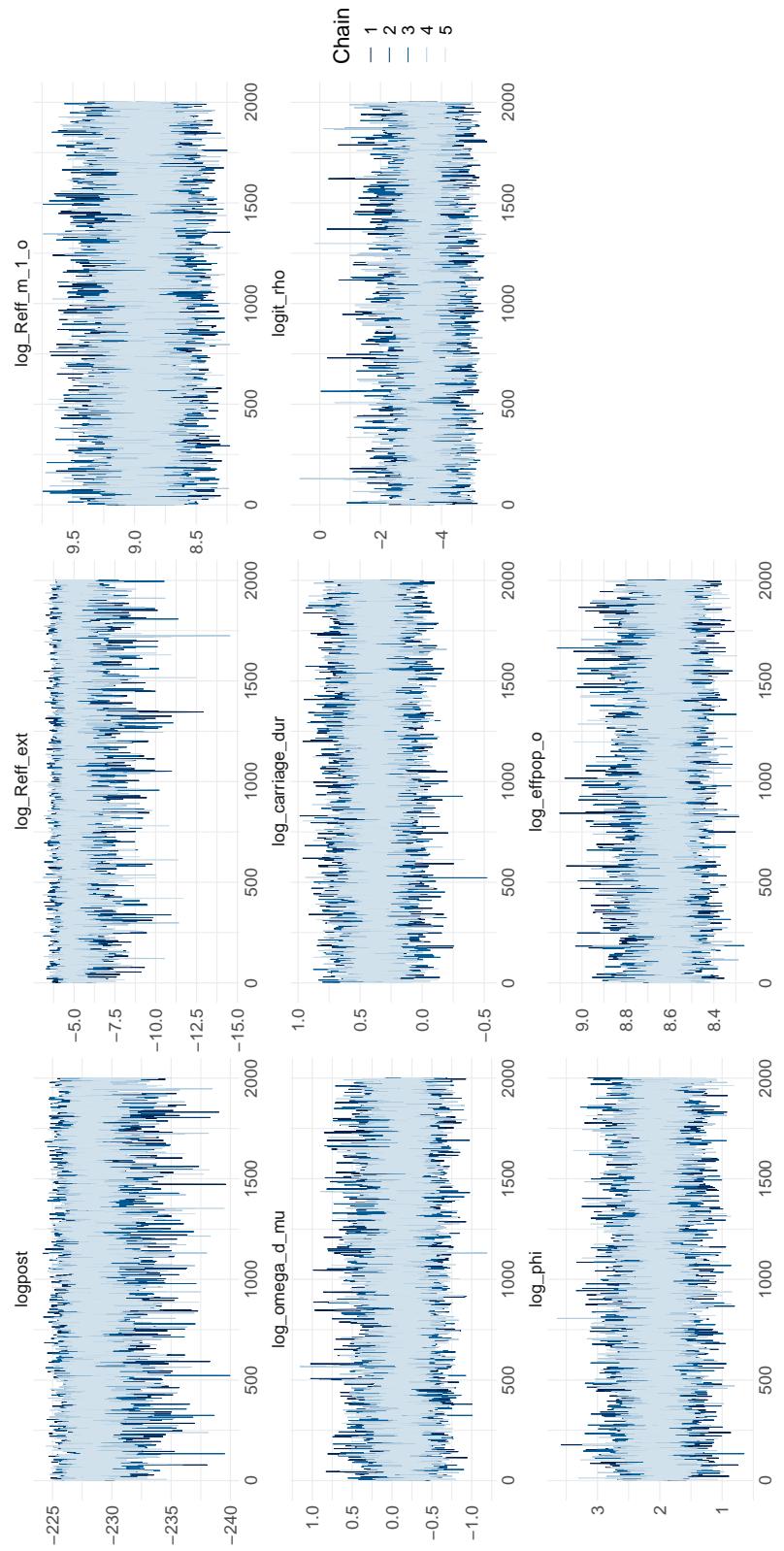


Figure B.21: Posterior traceplots for the SEIR ODE model fit to data from the Ebola outbreak in Liberia.

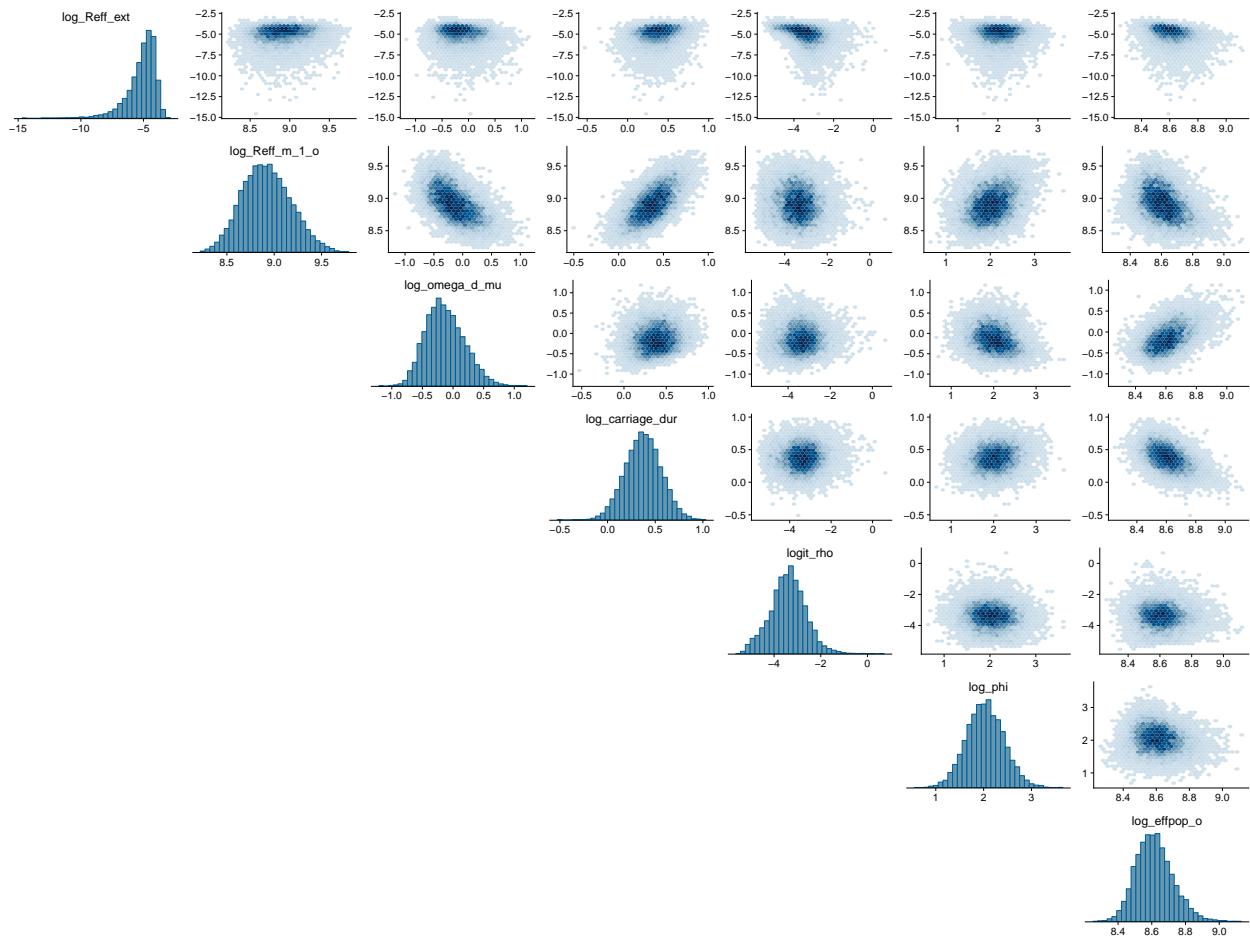


Figure B.22: Scatterplots of parameters for the SEIR ODE model fit to data from the Ebola outbreak in Liberia.

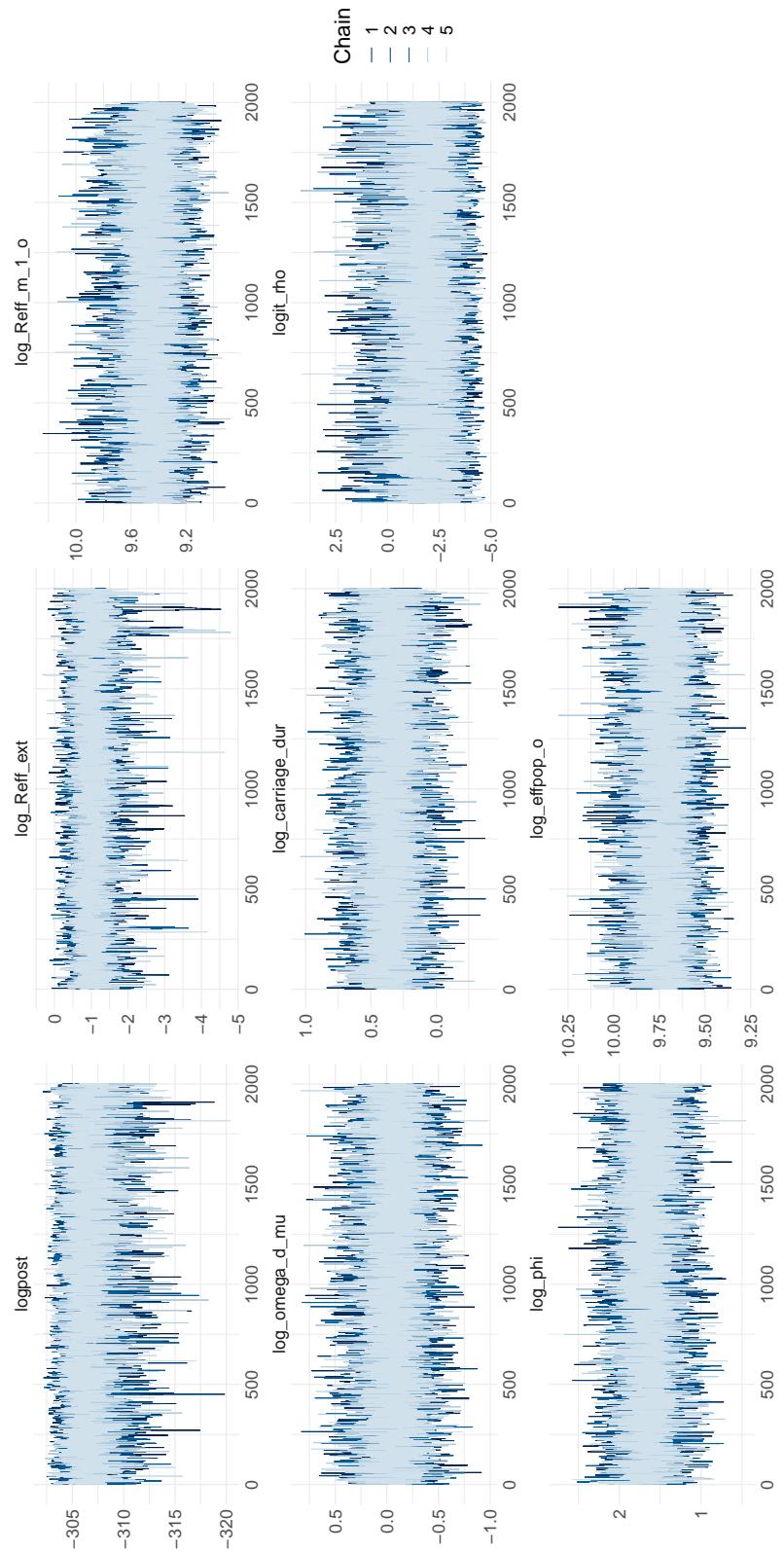


Figure B.23: Posterior traceplots for the SEIR model ODE fit to data from the Ebola outbreak in Sierra Leone.

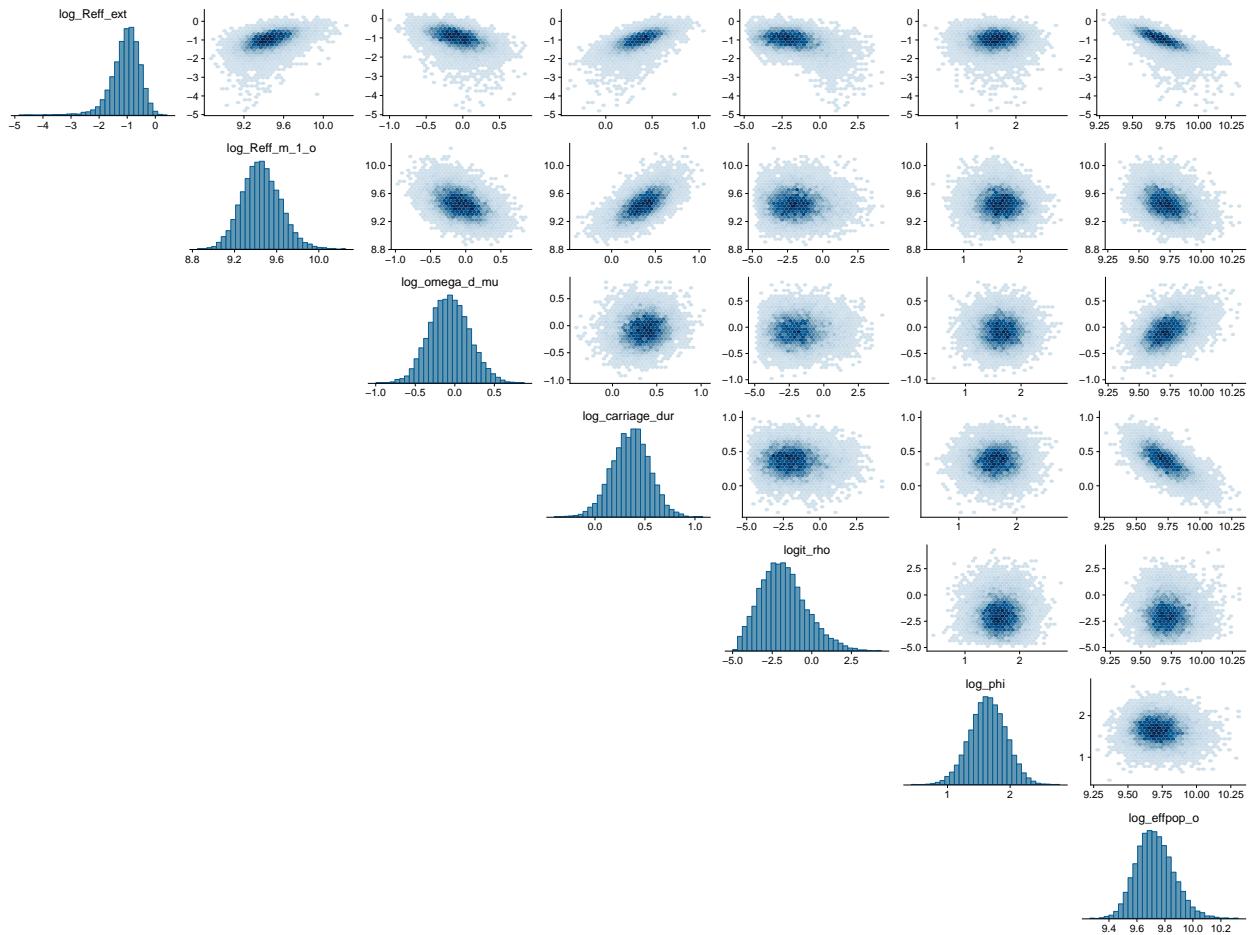


Figure B.24: Scatterplots of parameters for the SEIR ODE model fit to data from the Ebola outbreak in Sierra Leone.

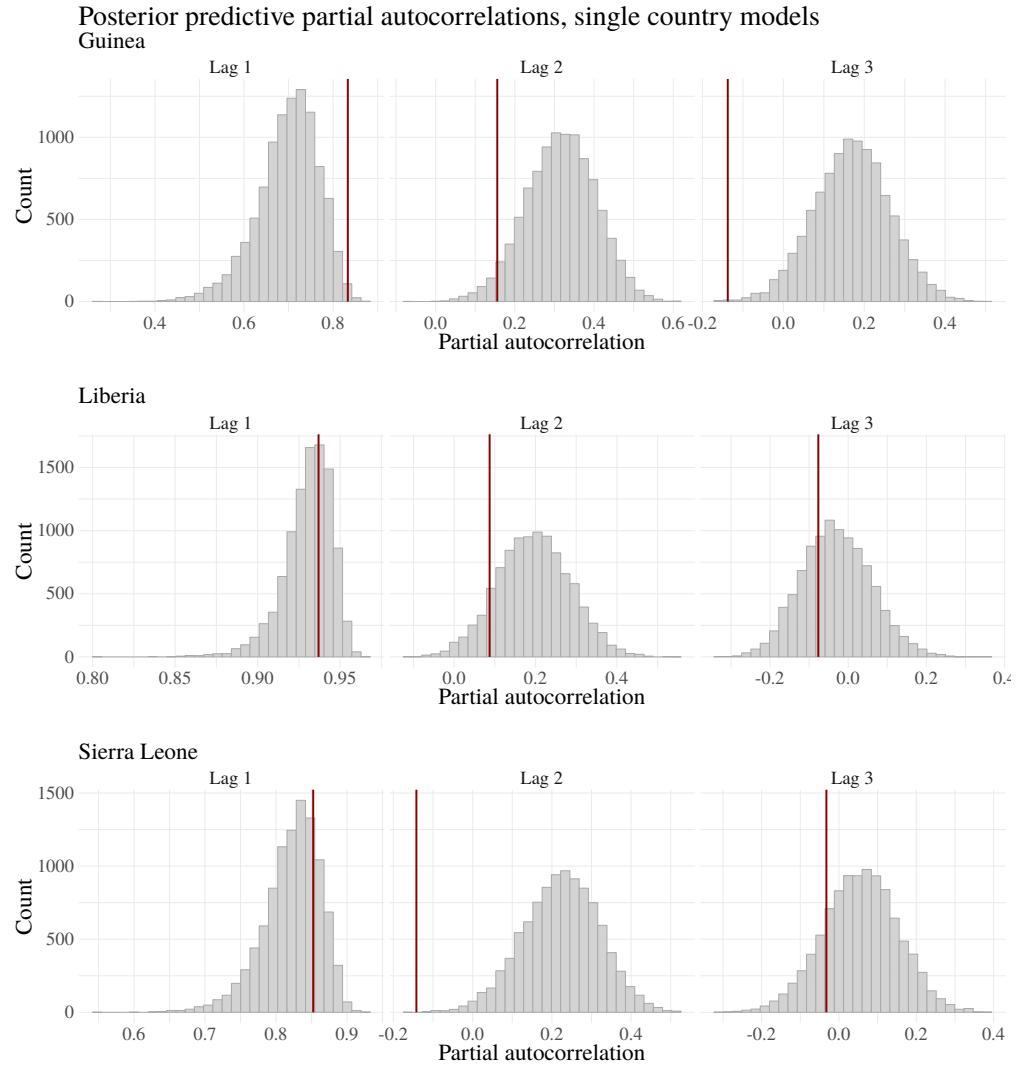


Figure B.25: Distributions of partial autocorrelations at lags 1, 2, and 3 for datasets generated under the full and partial posterior predictive distributions for the main country-specific SEIR ODE models fit to data from the West Africa Ebola outbreak. Vertical red lines are the partial autocorrelations for the observed incidence data at the respective lags.

B.8.2 Details and Supplementary Results for Multi-Country Models

Priors and MCMC details

We fit stratified SEIR models to the Ebola incidence data from Guinea, Liberia, and Sierra Leone under two prior regimes given in Tables B.35 and B.36. Cross-border transmission was incorporated into the rate of infection via virtual migration, as explained in Section 4.3.3. 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. We ran five chains for each model, initialized at random parameter values, for 200,000 iterations per chain. The model parameters, not including the initial compartment volumes, were jointly updated via MVNSS (see Section 2.2.1). 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 with covariances for parameters belonging to different countries set to zero. Thus, country specific parameters were proposed jointly but were independent in the proposal. Migration parameters were blocked with parameters corresponding to the destination country. The initial compartment volumes were updated jointly with the LNA paths in an EllipSS step. The MCMC alternated between three ElliptSS and two 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 B.3 with the one addition being the log ratio of adjusted reproductive numbers, subtract one, for Guinea, $\log \left((R_{eff,G}^{(2)} - 1)/(R_{eff,G}^{(1)} - 1) \right)$. Convergence was assessed visually by inspection of traceplots of posterior samples, and via potential scale reduction factors (PSRFs) [35] computed via the `coda` R package [173]. The LNA models each took roughly three days to run while run times for ODE models were on the order of 1–2 hours.

Table B.34: Priors for the initial compartment volumes at the times when transmission commenced in each country for single country models fit to data from the Ebola outbreak in West Africa. The initial compartment volumes for each country are assigned independent truncated multivariate normal priors that respect the boundary conditions on the state space of compartment volumes for that country (explained in Section B.2). 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	$(11.8 \times 10^6 - 7.1, 4, 3, 0.1)$
Liberia	$(4.4 \times 10^6 - 5.1, 3, 2, 0.1)$
Sierra Leone	$(7.1 \times 10^6 - 5.1, 3, 2, 0.1)$

Additional results for LNA models

Table B.35: Parameters for a country-stratified SEIR model fit to the West Africa Ebola outbreak, 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 shared. Adjusted reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$.

Parameter	Interpretation	Prior	Median (95% Interval)	References
$R_{eff,G}/R_{eff,G}^{(1)}$	Ratio of adjusted reproduction $\#_S$	LogNormal(0, 0.5 ²)	$R_{eff,G}/R_{eff,G}^{(1)} = 1.00 (0.38, 2.66)$	[49, 51, 52, 123]
$R_{eff,G}^{(2)}(t) - 1$	Adjusted reproduction $\# - 1, t \geq 33$	LogNormal(log(0.5), 0.75 ²)	$\Rightarrow R_{eff,G}^{(2)} = 1.50 (1.11, 3.17)$	[49, 51, 52, 123]
$R_{eff,L} - 1$	Adjusted reproduction $\# - 1$	LogNormal(log(0.5), 0.75 ²)	$\Rightarrow R_{eff,L} = 1.50 (1.11, 3.17)$	[49, 51, 52, 123]
$R_{eff,S} - 1$	Adjusted reproduction $\# - 1$	LogNormal(log(0.5), 0.75 ²)	$\Rightarrow R_{eff,S} = 1.50 (1.11, 3.17)$	[49, 51, 52, 123]
$\alpha_{GS}, \alpha_{GL}, \alpha_{LG}, \alpha_{LS}, \alpha_{SG}, \alpha_{SL}$	Infectious migration rate from country A to B	10000 $\alpha_{AB} \sim \text{Exponential}(0.25)$	# migrations per 1000 infected = 2.78 (0.10, 14.76)	[63]
ω_G/ω_G	Rate from $E_G \rightarrow I_G$	LogNormal(0.0, 0.3 ²)	$(1/\mu_G)/(1/\omega_G) = 1.00 (0.56, 1.80)$	[49, 212, 91]
ω_L/ω_L	Rate from $E_L \rightarrow I_L$	LogNormal(0.0, 0.3 ²)	$(1/\mu_L)/(1/\omega_L) = 1.00 (0.56, 1.80)$	[49, 212, 91]
ω_S/ω_S	Rate from $E_S \rightarrow I_S$	LogNormal(0.0, 0.3 ²)	$(1/\mu_S)/(1/\omega_S) = 1.00 (0.56, 1.80)$	[49, 212, 91]
$1/\mu_G$	Rate from $E_G \rightarrow I_G$	LogNormal(0.3, 0.2 ²)	$7/\mu_G = 9.45 (6.38, 13.98)$	[49, 212, 91]
$1/\mu_L$	Rate from $E_L \rightarrow I_L$	LogNormal(0.3, 0.2 ²)	$7/\mu_L = 9.45 (6.38, 13.98)$	[49, 212, 91]
$1/\mu_S$	Rate from $E_S \rightarrow I_S$	LogNormal(0.3, 0.2 ²)	$7/\mu_S = 9.45 (6.38, 13.98)$	[49, 212, 91]
$N_{eff,G}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^G, N_G/4$)	$N_{eff,G} \in (3627, 2.95 \times 10^6)$	Scale of counts
$N_{eff,L}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^L, N_L/4$)	$N_{eff,L} \in (4994, 1.1 \times 10^6)$	Scale of counts
$N_{eff,S}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^S, N_S/4$)	$N_{eff,S} \in (11317, 1.775 \times 10^6)$	Scale of counts
ρ_G	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_G = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ρ_L	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_L = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ρ_S	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_S = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ϕ_G, ϕ_L, ϕ_S	Neg.Binomial overdispersion	LogNormal(0, 5 ²)	$\Rightarrow = 1, (5.5 \times 10^{-5}, 1.8 \times 10^4)$	Diffuse.

Table B.36: Priors for a sensitivity analysis for the country-stratified SEIR model fit to the West Africa Ebola outbreak. Priors were less informative priors for the effective reproductive numbers, cross-border transmission, and infectious period duration vis-a-vis the main model. Subscripts, G, L, S , indicate specific countries, or generic countries A, B if a prior is shared. Effective reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$.

Parameter	Interpretation	Prior	Median (95% Interval)	References
$R_{eff,G}^{(2)}/R_{eff,G}^{(1)}$	Ratio of effective reproduction #'s	LogNormal(0, 0.5 ²)	$R_{eff,G}^{(2)}/R_{eff,G}^{(1)} = 1.00$ (0.38, 2.66)	[49, 51, 52, 123]
$R_{eff,G}^{(2)}(t) - 1$	Effective reproduction # - 1, $t \geq 33$	LogNormal(log(0.5), 1.5 ²)	$\Rightarrow R_{eff,G}^{(2)} = 1.50(1.02, 10.46)$	[49, 51, 52, 123]
$R_{eff,L} - 1$	Effective reproduction # - 1	LogNormal(log(0.5), 1.5 ²)	$\Rightarrow R_{eff,L} = 1.50(1.02, 10.46)$	[49, 51, 52, 123]
$R_{eff,S} - 1$	Effective reproduction # - 1	LogNormal(log(0.5), 1.5 ²)	$\Rightarrow R_{eff,S} = 1.50(1.02, 10.46)$	[49, 51, 52, 123]
$\alpha_{GS}, \alpha_{GL}, \alpha_{LG}, \alpha_{LS}, \alpha_{SG}, \alpha_{SL}$	Infectious migration rate from country A to B	1000 $\alpha_{AB} \sim \text{Exponential}(0.125)$	# migrations per 1000 infected = 5.55 (0.20, 29.51)	[63]
ω_G/ω_G	Rate from $E_G \rightarrow I_G$	LogNormal(0.0, 0.3 ²)	$(1/\mu_G)/(1/\omega_G) = 1.00$ (0.56, 1.80)	[49, 212, 91]
ω_L/ω_L	Rate from $E_L \rightarrow I_L$	LogNormal(0.0, 0.3 ²)	$(1/\mu_L)/(1/\omega_L) = 1.00$ (0.56, 1.80)	[49, 212, 91]
ω_S/ω_S	Rate from $E_S \rightarrow I_S$	LogNormal(0.0, 0.3 ²)	$(1/\mu_S)/(1/\omega_S) = 1.00$ (0.56, 1.80)	[49, 212, 91]
$1/\mu_G$	Rate from $E_G \rightarrow I_G$	LogNormal(0.3, 0.35 ²)	$7/\mu_G = 9.45$ (4.76, 18.76)	[49, 212, 91]
$1/\mu_L$	Rate from $E_L \rightarrow I_L$	LogNormal(0.3, 0.35 ²)	$7/\mu_L = 9.45$ (4.76, 18.76)	[49, 212, 91]
$1/\mu_S$	Rate from $E_S \rightarrow I_S$	LogNormal(0.3, 0.35 ²)	$7/\mu_S = 9.45$ (4.76, 18.76)	[49, 212, 91]
$N_{eff,G}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^G, N_G/4$)	$N_{eff,G} \in (3627, 2.95 \times 10^6)$	Scale of counts
$N_{eff,L}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^L, N_L/4$)	$N_{eff,L} \in (4994, 1.1 \times 10^6)$	Scale of counts
$N_{eff,S}$	Effective population size	Unif($\sum_{\ell=1}^L Y_\ell^S, N_S/4$)	$N_{eff,S} \in (11317, 1.775 \times 10^6)$	Scale of counts
ρ_G	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_G = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ρ_L	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_L = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ρ_S	Mean case detection prob.	LogitNormal(0, 1.4 ²)	$\Rightarrow \rho_S = 0.5, (0.06, 0.94)$	Very high and very low ρ unlikely.
ϕ_G, ϕ_L, ϕ_S	Neg.Binomial overdispersion	LogNormal(0, 5 ²)	$\Rightarrow = 1, (5.5 \times 10^{-5}, 1.8 \times 10^4)$	Diffuse.

Table B.37: Posterior estimates of initial numbers of exposed, infected, and recovered individuals for the main stratified approximate SEIR LNA model with virtual migration of infecteds fit to Ebola data from the West Africa outbreak under tight priors. The effective number of susceptibles is equal to the effective population size, reported in Table B.38, minus the numbers of exposed, infected, and recovered individuals, but is not reported since the effective population size is not marginally identifiable.

Parameter	Estimate
$E_{0,G}$	4.3 (1.1, 7.8)
$E_{0,L}$	3.4 (0.59, 6.6)
$E_{0,S}$	2.3 (0.23, 5.5)
$I_{0,G}$	3.4 (0.59, 6.6)
$I_{0,L}$	2.6 (0.47, 5.1)
$I_{0,S}$	2.4 (0.37, 4.8)
$R_{0,G}$	0.1 (0.07, 0.2)
$R_{0,L}$	0.1 (0.07, 0.2)
$R_{0,S}$	0.1 (0.06, 0.2)

Table B.38: Posterior medians (95% Bayesian credible intervals) for parameters of the approximate (Figure 4.9) stratified SEIR LNA models fit to Ebola outbreak data. Subscripts, G, L, S , indicate specific countries, or generic countries A, B . Effective reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$. Parameter interpretations and priors are given in Tables B.35 and B.36. The more informative of the two prior regimes was presented as the main model in Section 4.4.1.

Parameter	Posterior median (95% BCI)	
	<i>Tight Priors</i>	<i>Loose Priors</i>
$R_{eff,G}^{(1)}$	1.1 (1, 1.3)	1.1 (1, 1.2)
$R_{eff,G}^{(2)}$	1.2 (1.1, 1.4)	1.1 (1, 1.3)
$R_{eff,L}$	2.1 (1.6, 3.3)	1.8 (1.3, 3.4)
$R_{eff,S}$	1.3 (1.1, 1.4)	1.1 (1, 1.3)
$7/\omega_G$	7.5 (3.9, 13.6)	5.6 (2.6, 11.4)
$7/\omega_L$	9 (4.1, 19.9)	7.2 (2.5, 19.3)
$7/\omega_S$	4.1 (2.4, 6.6)	3 (1.5, 5.4)
$7/\mu_G$	8.4 (5.9, 11.8)	6.4 (3.6, 10.8)
$7/\mu_L$	9 (6, 13.2)	7.5 (3.4, 15.5)
$7/\mu_S$	6.1 (4.4, 8.2)	3.9 (2.4, 6.2)
$1000\alpha_{GL}$	3.0 (0.1, 15.5)	6.3 (0.3, 29.8)
$1000\alpha_{GS}$	2.6 (0.1, 13.5)	4.4 (0.2, 23.5)
$1000\alpha_{LG}$	3.9 (0.5, 14.2)	6.5 (1, 20.5)
$1000\alpha_{LS}$	1.4 (0.1, 7.2)	3 (0.2, 13.3)
$1000\alpha_{SG}$	2.6 (0.1, 13.4)	4.7 (0.3, 25)
$1000\alpha_{SL}$	3.0 (0.1, 16.7)	6.8 (0.2, 37.8)
$\rho_G N_{eff,G}$	9900 (6600, 15600)	14600 (7800, 29100)
$\rho_L N_{eff,L}$	5900 (4800, 7900)	6400 (4800, 11100)
$\rho_S N_{eff,S}$	26800 (20600, 38700)	39400 (24800, 85800)
ϕ_G	5.7 (3.4, 10.6)	6.3 (3.6, 12.9)
ϕ_L	8.8 (4.1, 19.3)	10.2 (4.7, 27.2)
ϕ_S	48 (22, 118)	56 (25, 140)

Table B.39: Effective sample sizes and potential scale reduction factors for the main stratified SEIR LNA model with virtual migration of infecteds fit to data from the West Africa Ebola outbreak under tight priors.

Parameter	ESS	PSRF
$\log(R_{eff,G}^{(2)}/R_{eff,G}^{(1)})$	647	1.00
$\log(R_{eff,G}^{(2)})$	626	1.01
$\log(R_{eff,L})$	347	1.02
$\log(R_{eff,S})$	459	1.08
$\log(\omega_G/\mu_G)$	1161	1.01
$\log(\omega_L/\mu_L)$	409	1.02
$\log(\omega_S/\mu_S)$	1217	1.01
$\log(1/\mu_G)$	1141	1.01
$\log(1/\mu_L)$	743	1.00
$\log(1/\mu_S)$	874	1.00
$\log(1000\alpha_{GL})$	1583	1.00
$\log(1000\alpha_{GS})$	972	1.02
$\log(1000\alpha_{LG})$	292	1.04
$\log(1000\alpha_{LS})$	478	1.02
$\log(1000\alpha_{SG})$	1238	1.00
$\log(1000\alpha_{SL})$	1198	1.06
$\log(\rho_G N_{eff,G})$	725	1.00
$\log(\rho_L N_{eff,L})$	521	1.01
$\log(\rho_S N_{eff,S})$	479	1.04
$\text{logit}(\rho_G)$	631	1.01
$\text{logit}(\rho_L)$	217	1.04
$\text{logit}(\rho_S)$	1414	1.01
$\log(\phi_G)$	596	1.03
$\log(\phi_L)$	1098	1.01
$\log(\phi_S)$	489	1.00

Posterior distributions of joint West Africa model parameters

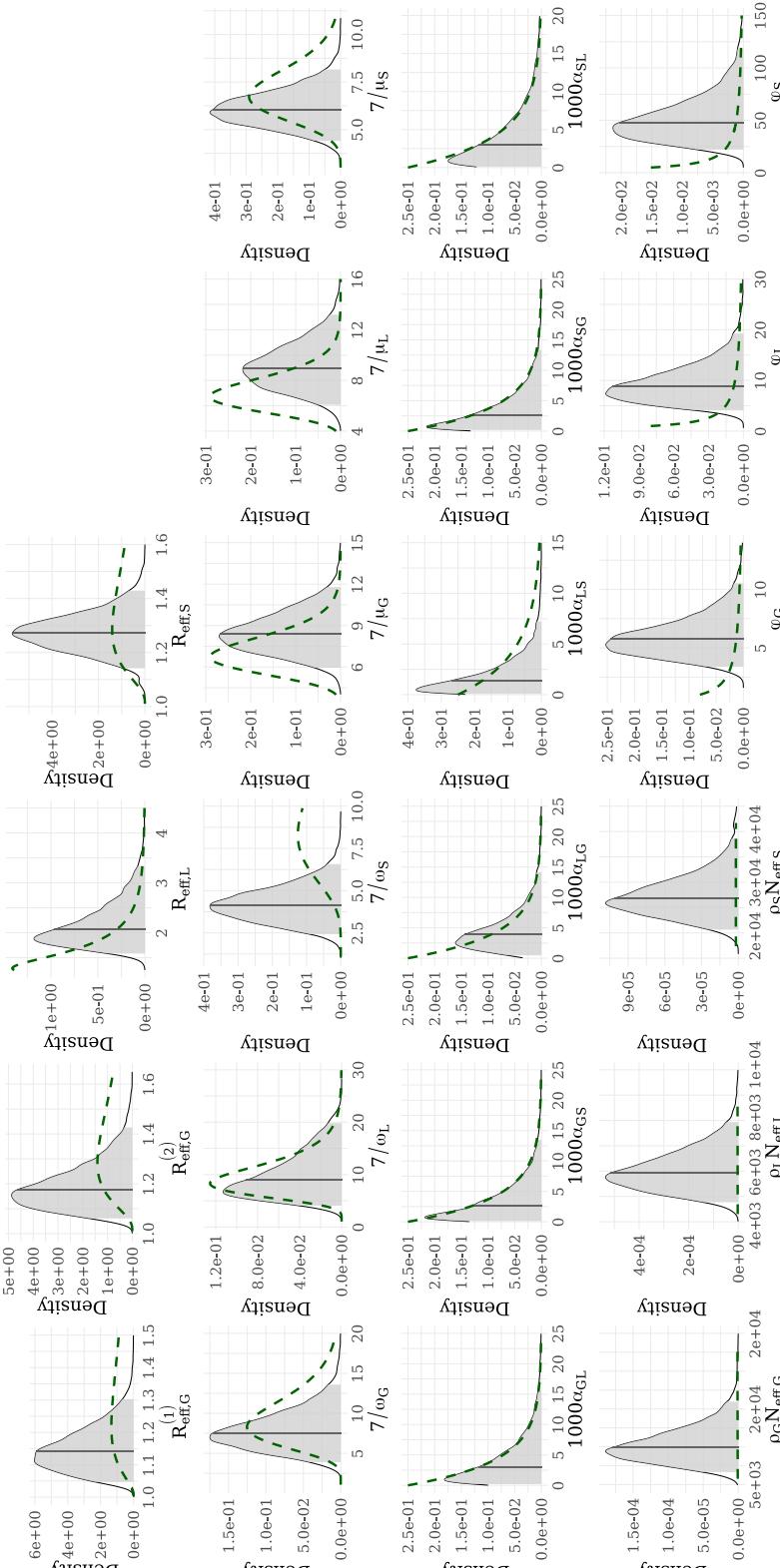


Figure B.26: Posterior distributions of parameters of the main joint SEIR LNA model fit via the LNA to data from the West Africa Ebola outbreak under tight priors. We show posterior medians (solid gray lines), 95% Bayesian credible intervals (light gray areas under the posterior densities), prior densities (induced priors for the reporting rate and latent period durations) over the posterior ranges (dashed green curves). $R_{adj} = \beta N_{eff}/\mu$ is the effective reproductive number, where β is the per-contact infection rate, N_{eff} is the effective population size, and μ is the recovery rate. The latent and infectious period durations, $7/\omega$ and $7/\mu$, respectively, are given in days. The effective number of cross-border infectious contacts in country B per 1000 infected individuals in country A is $1000\alpha_{AB}$. The effective reporting rate is ρN_{eff} , and ϕ is the negative binomial over-dispersion parameter. Subscripts indicate countries.

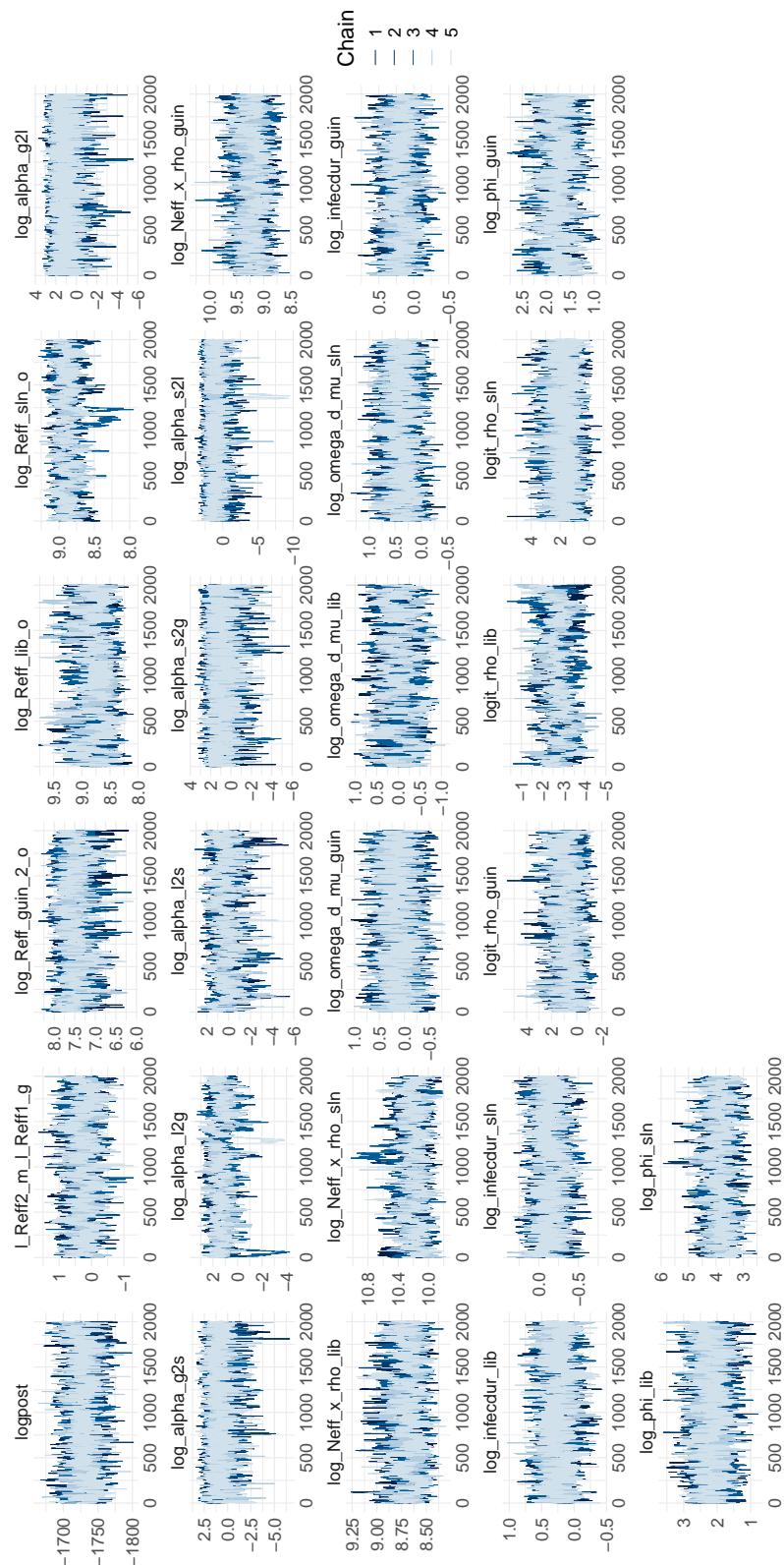


Figure B.27: Posterior traceplots for a stratified SEIR LNA model fit to a simulated Ebola outbreak.

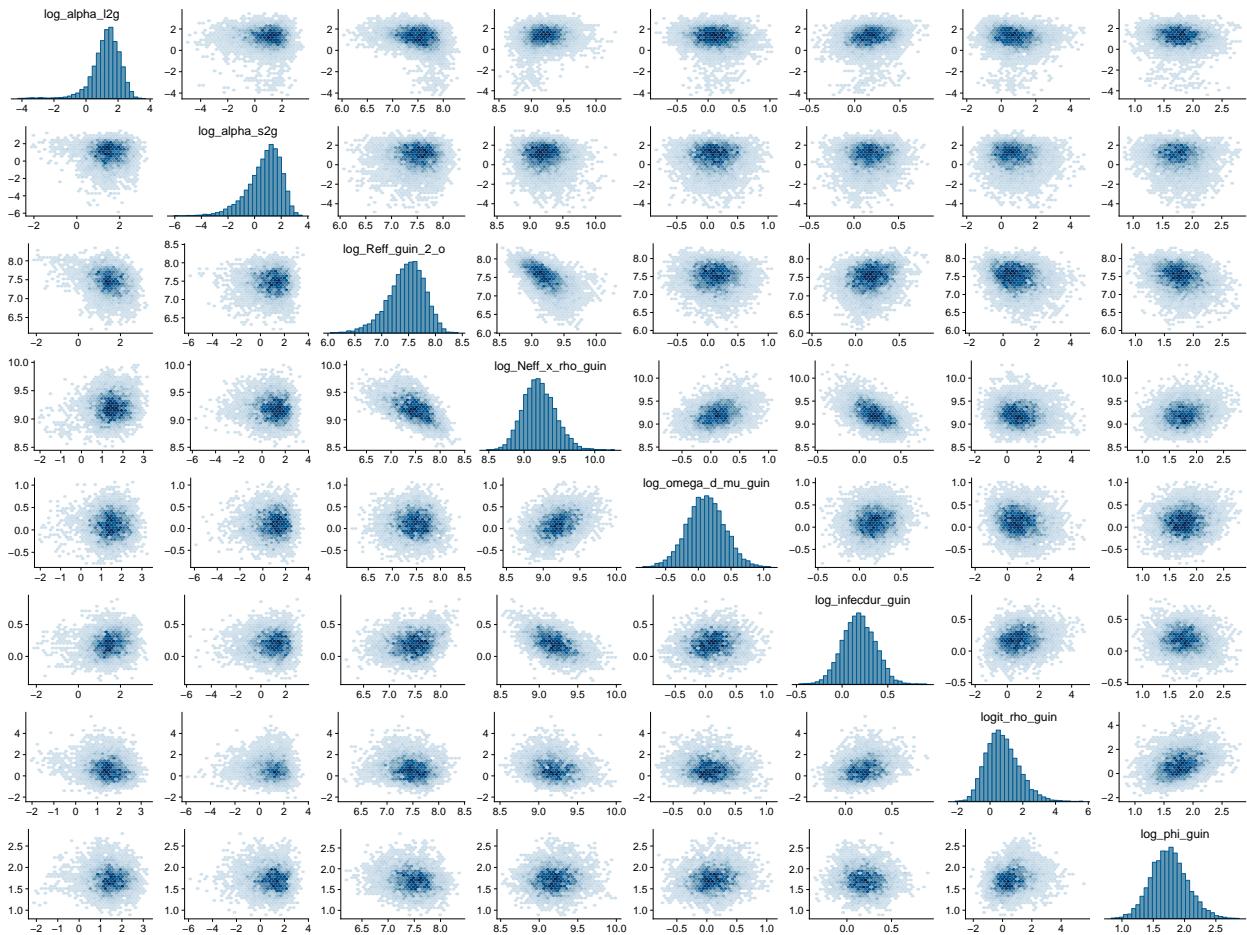


Figure B.28: Scatterplots of Guinea-specific parameters in a stratified SEIR LNA model fit to data from the West Africa Ebola outbreak.

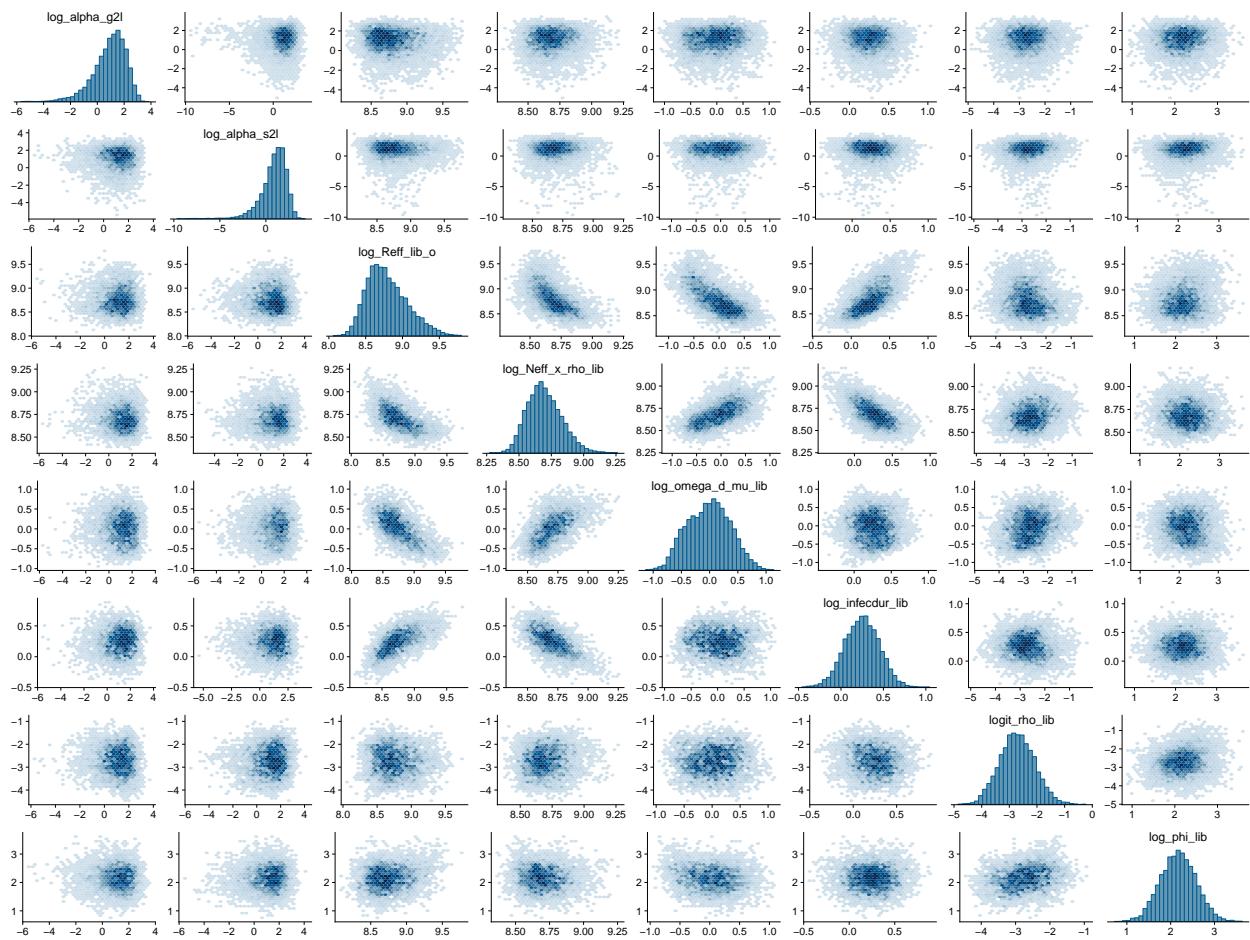


Figure B.29: Scatterplots of Liberia-specific parameters in a stratified SEIR LNA model fit to data from the West Africa Ebola outbreak.

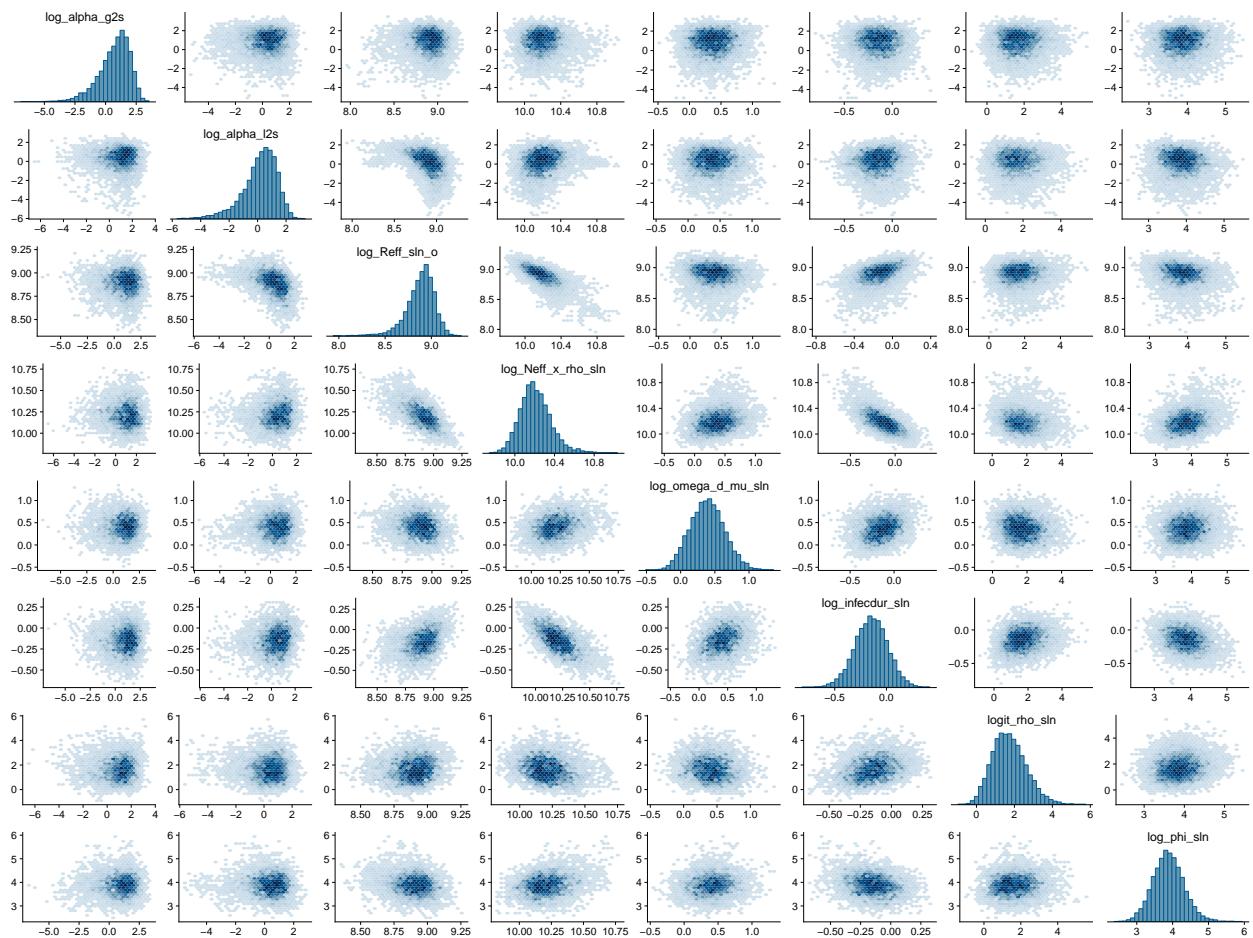


Figure B.30: Scatterplots of Sierra Leone-specific parameters in a stratified SEIR-LNA model fit to data from the West Africa Ebola outbreak.

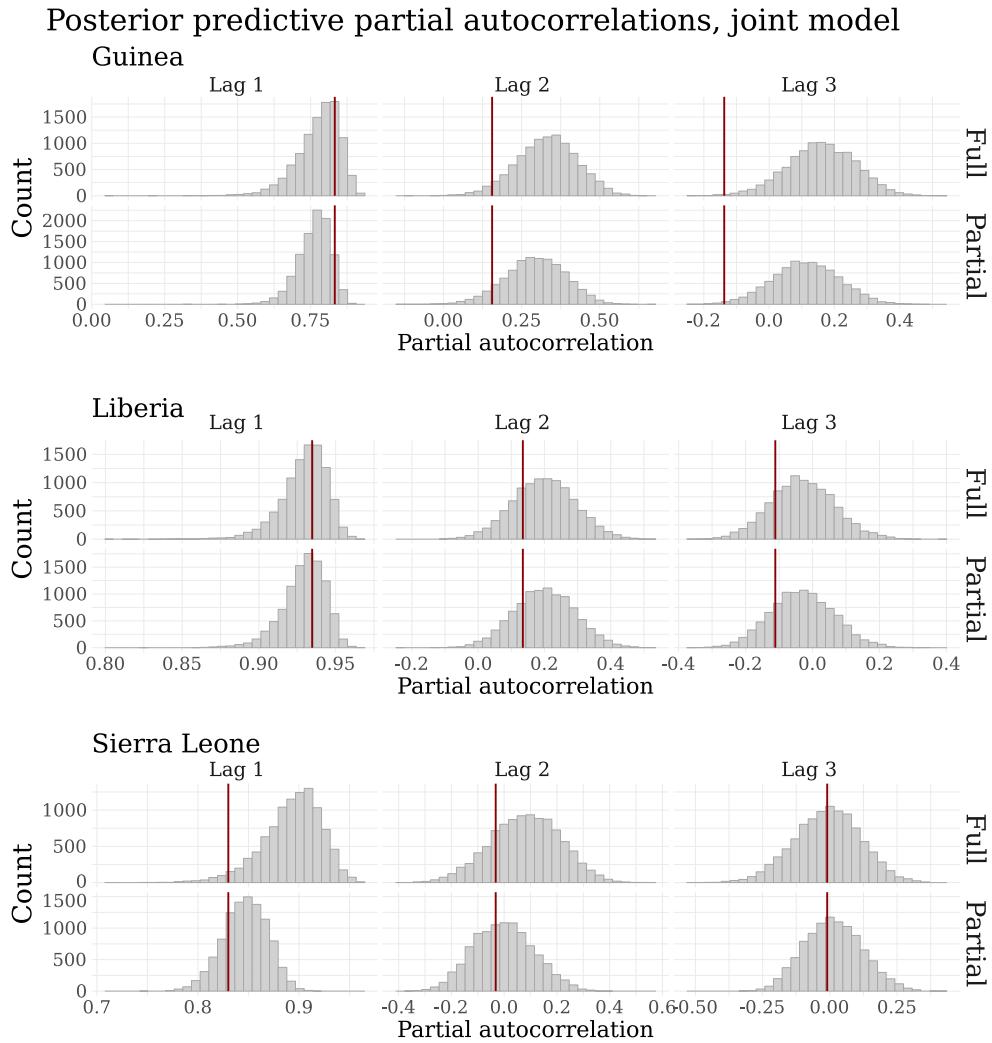


Figure B.31: Distributions of partial autocorrelations at lags 1, 2, and 3 for datasets generated under the full and partial posterior predictive distributions for the main stratified SEIR LNA model fit to data from the West Africa Ebola outbreak. Vertical red lines are the partial autocorrelations for the observed incidence data at the respective lags.

Posterior distributions of LNA draws for a joint Ebola model

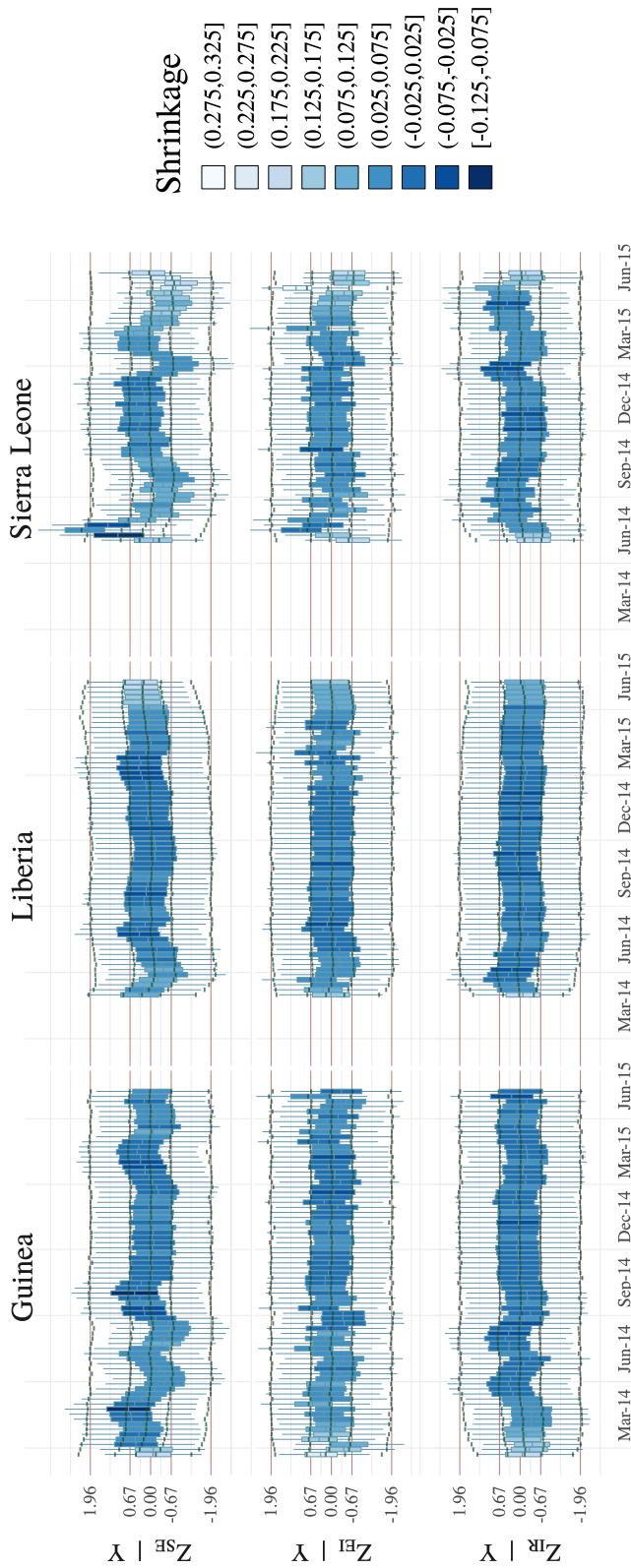


Figure B.32: Posterior distributions of the LNA draws for exposures, infections, and recoveries (blue boxplots) in the main stratified SEIR LNA model fit data from the Ebola outbreak in Guinea, Liberia, and Sierra Leone. The lower and upper whisker tips correspond to the 2.5th and 97.5th posterior quantiles, the lower and upper hinges to the 25th and 75th quantiles, and the middle hash mark to the posterior median. The solid red lines are the theoretical quantiles of the posterior predictive distribution (or equivalently, the prior distribution) of the LNA draws, drawn at the quantiles of a standard normal distribution corresponding to the boxplot quantiles. The green ticks are the estimated quantiles of the posterior predictive distributions of the LNA draws, accounting for boundary conditions on the state space of the latent process and obtained by simulating LNA paths from the posterior predictive distribution. The posterior distributions of LNA draws are shaded according to the level of posterior shrinkage, computed as one minus the ratio of standard deviations of LNA draws in the posterior and prior.

Additional results for ODE models

Table B.40: Posterior estimates of initial numbers of exposed, infected, and recovered individuals for the main stratified SEIR ODE model with virtual migration of infecteds fit to Ebola data from the West Africa outbreak under tight priors. The effective number of susceptibles is equal to the effective population size, reported in Table B.38, minus the numbers of exposed, infected, and recovered individuals, but is not reported since the effective population size is not marginally identifiable.

Parameter	Estimate
$E_{0,G}$	5.5 (2.21, 8.99)
$E_{0,L}$	2.35 (0.19, 5.56)
$E_{0,S}$	6.16 (3.07, 9.11)
$I_{0,G}$	4.62 (1.66, 7.67)
$I_{0,L}$	1.5 (0.09, 4.05)
$I_{0,S}$	4.82 (2.35, 7.31)
$R_{0,G}$	0.13 (0.1, 0.17)
$R_{0,L}$	0.09 (0.05, 0.14)
$R_{0,S}$	0.18 (0.14, 0.23)

Table B.41: Posterior medians (95% Bayesian credible intervals) for parameters of the approximate (Figure 4.9) stratified SEIR models fit to Ebola outbreak data. Subscripts, G, L, S , indicate specific countries, or generic countries A, B . Effective reproduction numbers are defined with respect to the effective population size as $R_{adj} = \beta N_{eff}/\mu$. Parameter interpretations and priors are given in Tables B.35 and B.36. The more informative of the two prior regimes was presented as the main model in Section 4.4.1.

Parameter	Posterior median (95% BCI)	
	<i>Tight Priors</i>	<i>Loose Priors</i>
$R_{eff,G}^{(1)}$	1.2 (1.1, 1.4)	1.2 (1.1, 1.4)
$R_{eff,G}^{(2)}$	1.3 (1.1, 1.5)	1.2 (1.1, 1.5)
$R_{eff,L}$	2.2 (1.7, 3.2)	2.5 (1.6, 4.2)
$R_{eff,S}$	1.3 (1.2, 1.5)	1.2 (1.1, 1.4)
$7/\omega_G$	10.5 (4.9, 19.1)	12.9 (5.2, 22.7)
$7/\omega_L$	3.9 (2.2, 6.6)	3 (1.7, 5.7)
$7/\omega_S$	8.7 (6.2, 12.3)	7.2 (4.2, 11.9)
$7/\mu_G$	8.7 (6.2, 12.3)	7.2 (4.2, 11.9)
$7/\mu_L$	9.6 (6.6, 13.8)	11.1 (5.8, 19.3)
$7/\mu_S$	5.9 (4.3, 8.2)	4 (2.5, 6.5)
$1000\alpha_{GL}$	2.3 (0.1, 12.5)	3.9 (0.2, 21.5)
$1000\alpha_{GS}$	4.7 (0.2, 24.3)	6.6 (0.2, 44)
$1000\alpha_{LG}$	1.7 (0.1, 7)	1.8 (0.1, 8.8)
$1000\alpha_{LS}$	0.7 (0, 5.2)	0.6 (0, 5.3)
$1000\alpha_{SG}$	2.6 (0.1, 13.5)	4.6 (0.2, 23.8)
$1000\alpha_{SL}$	2.9 (0.1, 15.4)	5.6 (0.2, 31)
$\rho_G N_{eff,G}$	8900 (6300, 13500)	10900 (6600, 19700)
$\rho_L N_{eff,L}$	5700 (4600, 7400)	5300 (4400, 7400)
$\rho_S N_{eff,S}$	27200 (20600, 36500)	34900 (23600, 53100)
ϕ_G	3.5 (2.3, 5.2)	3.5 (2.3, 5.2)
ϕ_L	7.4 (3.6, 16.2)	7.9 (3.7, 17.7)
ϕ_S	5.1 (2.7, 9.5)	5.7 (3, 10.5)

Table B.42: Effective sample sizes and potential scale reduction factors for the main stratified SEIR ODE model with virtual migration of infecteds fit to data from the West Africa Ebola outbreak under tight priors.

Parameter	ESS	PSRF
$\log(R_{eff,G}^{(2)}/R_{eff,G}^{(1)})$	4490	1.00
$\log(R_{eff,G}^{(2)})$	3917	1.00
$\log(R_{eff,L})$	4001	1.00
$\log(R_{eff,S})$	4749	1.00
$\log(\omega_G/\mu_G)$	4848	1.00
$\log(\omega_L/\mu_L)$	3032	1.00
$\log(\omega_S/\mu_S)$	4558	1.00
$\log(1/\mu_G)$	4431	1.00
$\log(1/\mu_L)$	5234	1.00
$\log(1/\mu_S)$	5326	1.00
$\log(1000\alpha_{GL})$	3964	1.00
$\log(1000\alpha_{GS})$	4704	1.00
$\log(1000\alpha_{LG})$	2368	1.02
$\log(1000\alpha_{LS})$	1995	1.01
$\log(1000\alpha_{SG})$	4289	1.00
$\log(1000\alpha_{SL})$	4005	1.00
$\log(\rho_G N_{eff,G})$	4647	1.00
$\log(\rho_L N_{eff,L})$	4693	1.00
$\log(\rho_S N_{eff,S})$	4420	1.00
$\text{logit}(\rho_G)$	4457	1.00
$\text{logit}(\rho_L)$	1346	1.01
$\text{logit}(\rho_S)$	5521	1.00
$\log(\phi_G)$	5474	1.00
$\log(\phi_L)$	5275	1.00
$\log(\phi_S)$	4475	1.00

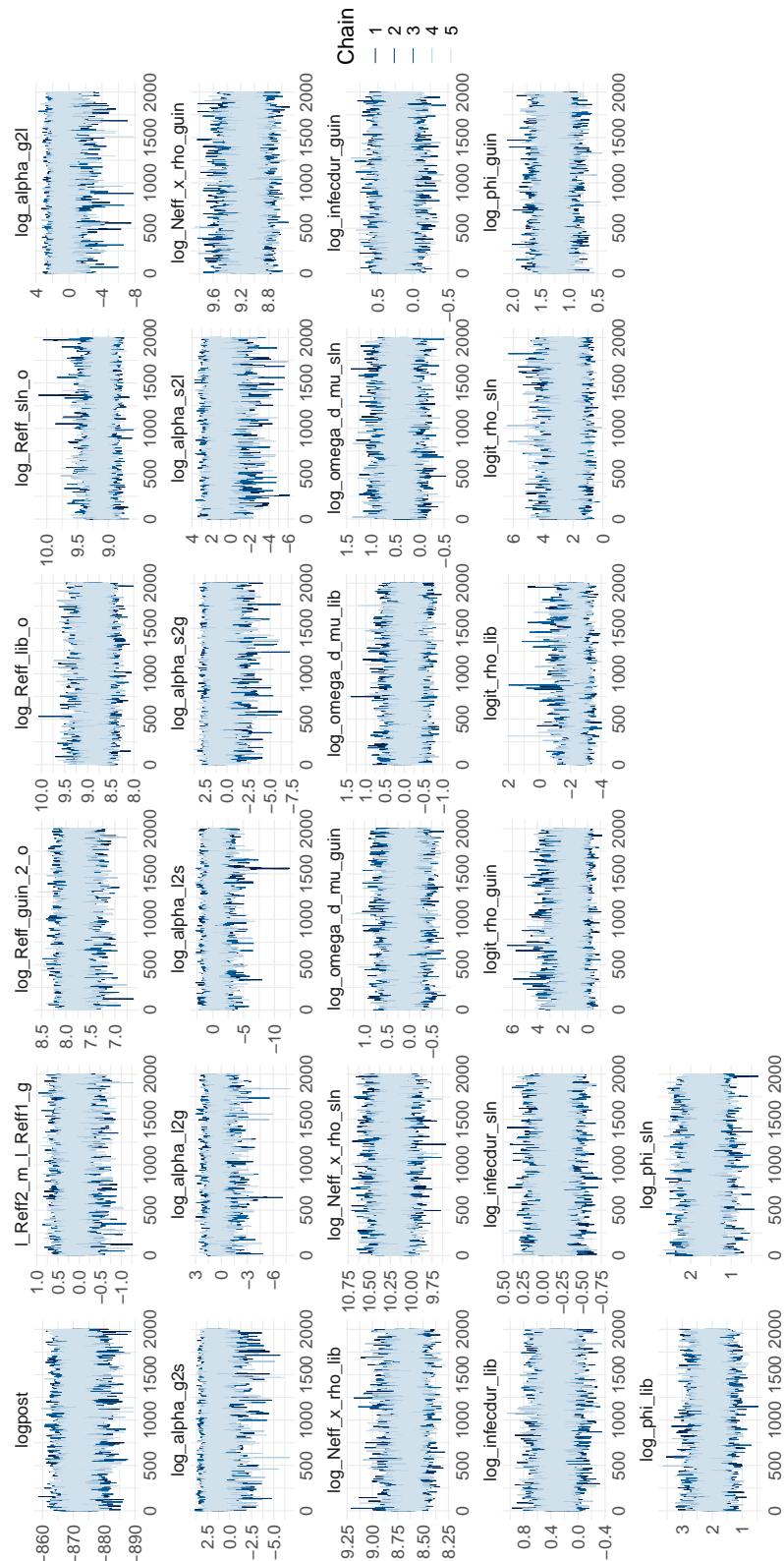


Figure B.33: Posterior traceplots for a stratified SEIR ODE model fit to a simulated Ebola outbreak.

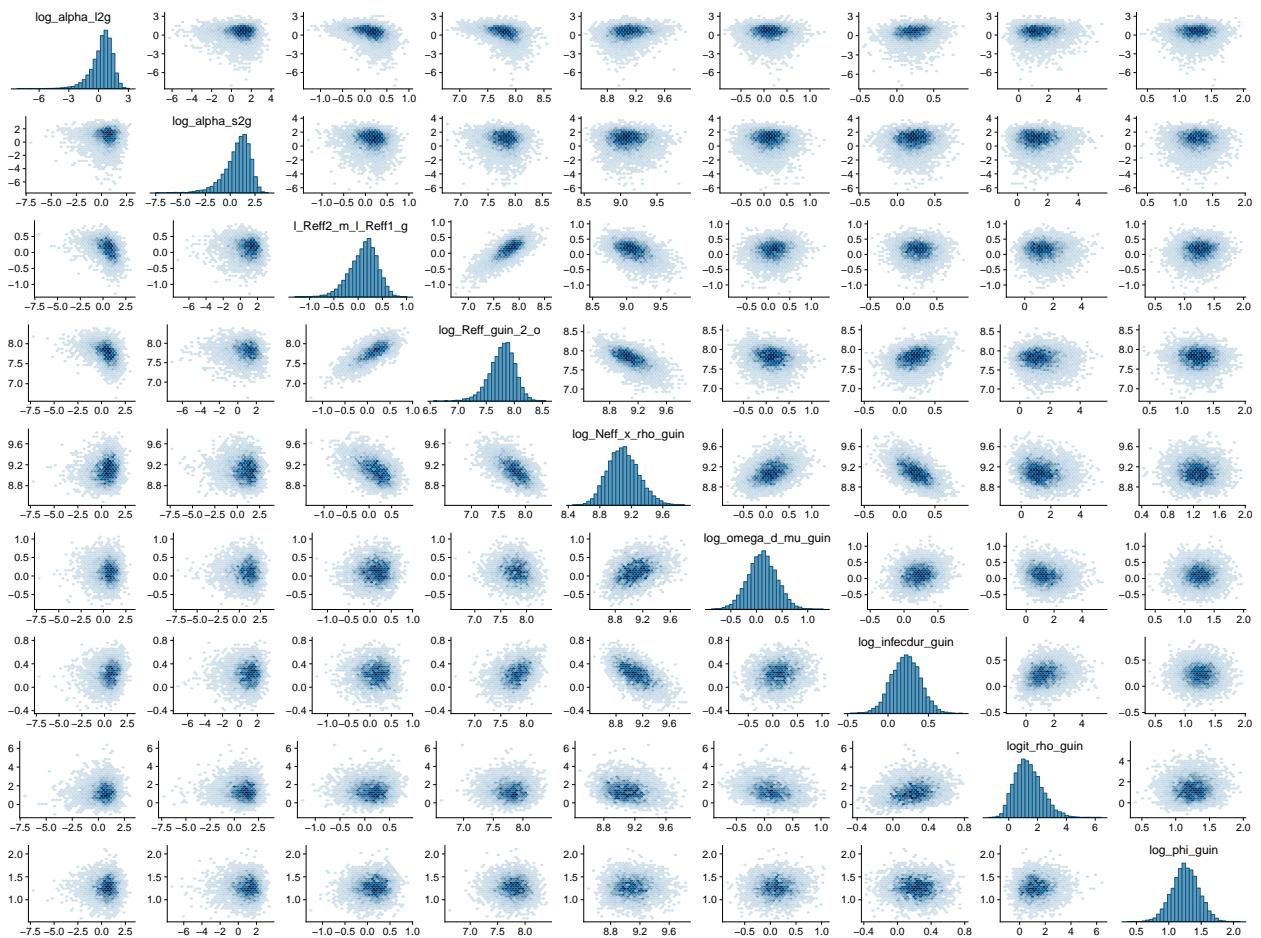


Figure B.34: Scatterplots of Guinea-specific parameters in a stratified SEIR ODE model fit to data from the West Africa Ebola outbreak.

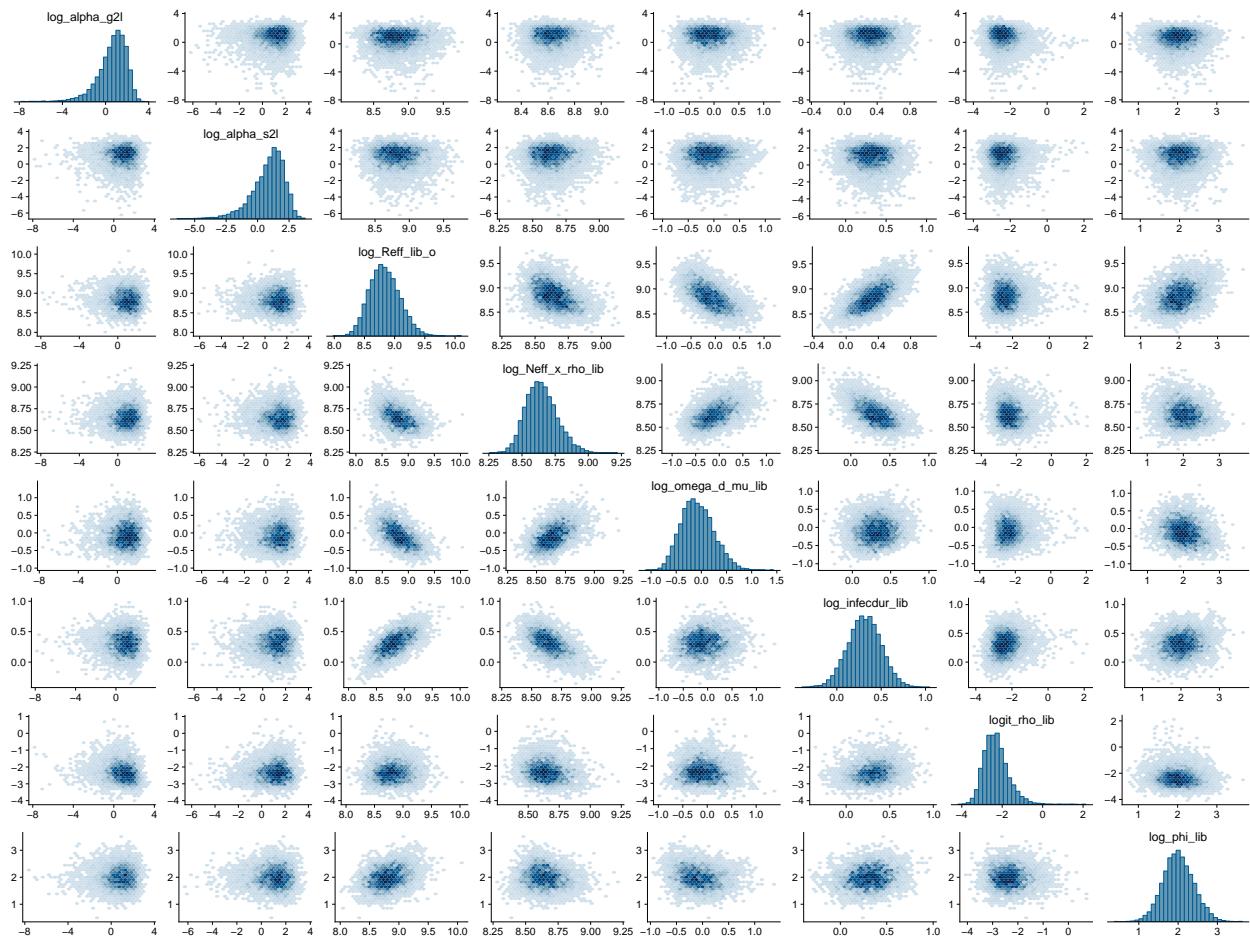


Figure B.35: Scatterplots of Liberia-specific parameters in a stratified SEIR ODE model fit to data from the West Africa Ebola outbreak.

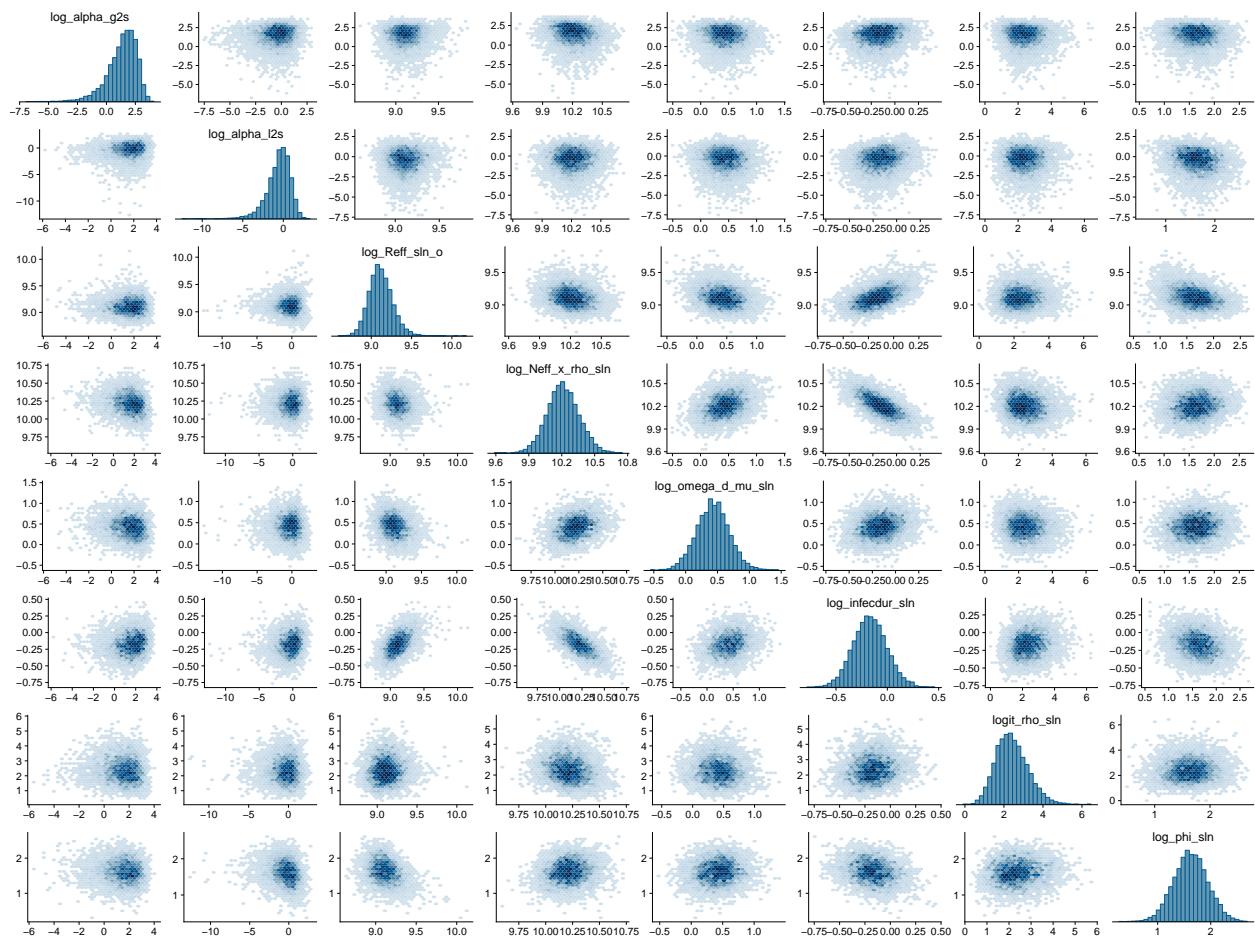


Figure B.36: Scatterplots of Sierra Leone-specific parameters in a stratified SEIR ODE model fit to data from the West Africa Ebola outbreak.

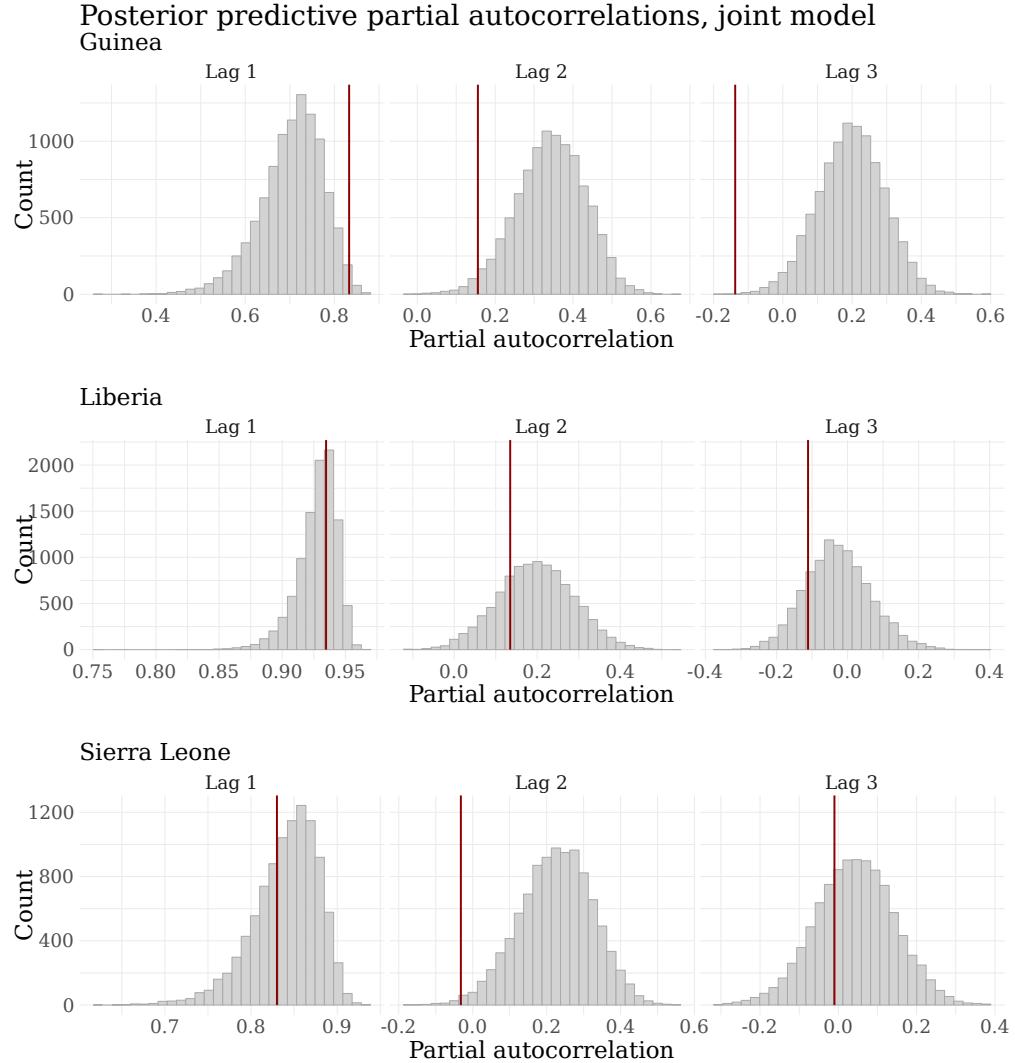


Figure B.37: Distributions of partial autocorrelations at lags 1, 2, and 3 for datasets generated under the full and partial posterior predictive distributions for the main stratified SEIR model fit via the ODE to data from the West Africa Ebola outbreak. Vertical red lines are the partial autocorrelations for the observed incidence data at the respective lags.

Appendix C

APPENDIX TO CHAPTER 5

C.1 *Simulation Details for Results in Section 5.2*

C.1.1 *Simulation setup*

We simulated an outbreak with SIRS dynamics in a population of $N = 100,000$ individuals, ten of whom were initially infected, spanning two wave of the outbreak over thirty weeks. The mean infectious period duration was one week, and the mean duration of immunity was four weeks. The basic reproduction numbers of the outbreak, and hence the per-contact infectivity rates, varied sinusoidally with additive noise, over the course of the outbreak:

$$R_0(t) = \beta(t)N/\mu = 2 + 0.35(\sin(t/3.5) + 0.1Z_t), \quad (\text{C.1})$$

where $Z_t \sim N(0, 1)$. The per-contact infection rates are given by $\beta(t_\ell) = R_0(t_\ell)\mu/N$. The observed incidence was a negative binomial sample of the true incidence with mean case detection rate ρ and overdispersion parameter ϕ :

$$Y_\ell \sim \text{Neg.Binom.}(\mu = \rho(N_{SI}(t_\ell) - N_{SI}(t_{\ell-1})), \sigma^2 = \mu(1 + \mu/\phi)). \quad (\text{C.2})$$

Table C.1: Parameters and priors used in fitting two SIRS models, one with constant FOI and another with time-varying FOI, to data from an outbreak with SIRS dynamics where the per-contact infection rate varied sinusoidally over time.

Parameter	Interpretation	Truth	Prior	Median (95% Interval)
$R_0 - 1$	Basic reproduction #1	Eqn. C.1	LogNormal(0, 0.5)	$\implies R_0(t_0) = 2.00$ (1.14, 8.10)
$1/\mu$	Mean infectious period	1	LogNormal(0.05, 0.5)	1.05 (0.39, 2.80)
$\rho/(1 - \rho)$	Odds of case detection	0.25	LogNormal(-0.847, 0.5)	$\implies \rho = 0.3$ (0.14, 0.53)
$1/\sqrt{\phi}$	Neg.Binom. overdispersion	50	Exponential(5)	$\implies \phi = 52$ (1.84, 39000).

We fit two models, one where the force of infection was constant over the course of the outbreak and another where the rates of infectious contact were time varying, with log-differences in the basic reproduction numbers penalized using a first order Gaussian Markov random field (GMRF) shrinkage prior where the standard deviation of the log-differences is given by σ_{GMRF} (analogous to the models in Section 5.3). The first order GMRF shrinkage prior penalizes the magnitude of one-step differences in log reproduction numbers, $\log(R_0(t_\ell)) - \log(R_0(t_{\ell-1})) \sim N(0, \sigma_{GMRF}^2)$. Sample paths of the GMRF are obtained by sampling the initial state and differences,

$$\log(R_0(t_0)) \sim N(\mu_0 = 0, \sigma_0^2 = 0.5^2) \quad (C.3)$$

$$\log(\sigma_{GMRF}) \sim N(-2.75, 0.25^2) \quad (C.4)$$

$$\log(R_0(t_\ell)) - \log(R_0(t_{\ell-1})) \sim N(0, \sigma_{GMRF}^2) \quad (C.5)$$

$$\log(R_0(t_\ell)) = \log(R_0(t_0)) + \sum_{k=1}^{\ell} \log(R_0(t_\ell)) - \log(R_0(t_{\ell-1})), \quad \ell = 1, \dots, 30 \quad (C.6)$$

We used non-centered parameterizations for both the LNA and the GMRF. The models were fit under informative priors, summarized in Table C.1. We ran five MCMC chains per model for 100,000 iterations per chain, alternating between five elliptical slice sampling (EllipSS) updates and one multivariate normal slice sampling (MVNSS) updates per MCMC iteration. The estimation scale on which the MCMC explored the parameter space was given by $\{\log(R_0(t_0)), \log(1/\mu), \text{logit}(\rho), 1/\sqrt{(\phi)}\}$. The MVNSS proposal covariance matrix was adapted throughout the first 50,000 iterations, samples from which were subsequently discarded, using the gain factor sequence: the gain factor sequence, $\gamma_n = 0.5(1 + 0.01n)^{-0.99}$. The contribution of isotropic Gaussian noise to the MVNSS proposal was initialized at 0.001 and reduced throughout the adaptation phase according to the sequence $\iota_n = 0.001(1 + 0.01n)^{-0.99}$. Convergence was assessed visually by inspection of traceplots of posterior samples, and via potential scale reduction factors (PSRFs) [35] computed via the `coda` R package [173].

C.1.2 Additional results

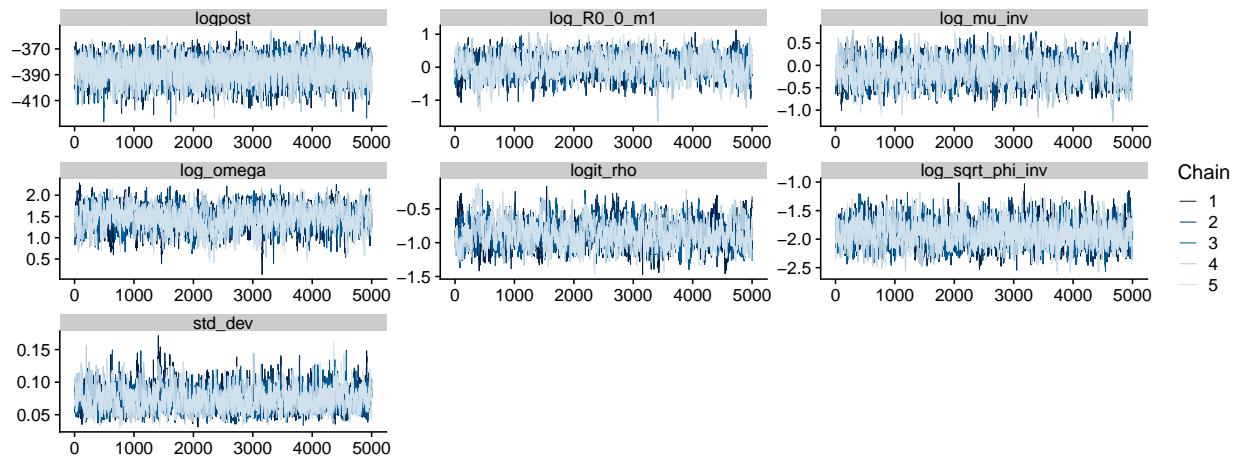


Figure C.1: Posterior traceplots for SIRS model parameters with time-varying force of infection.

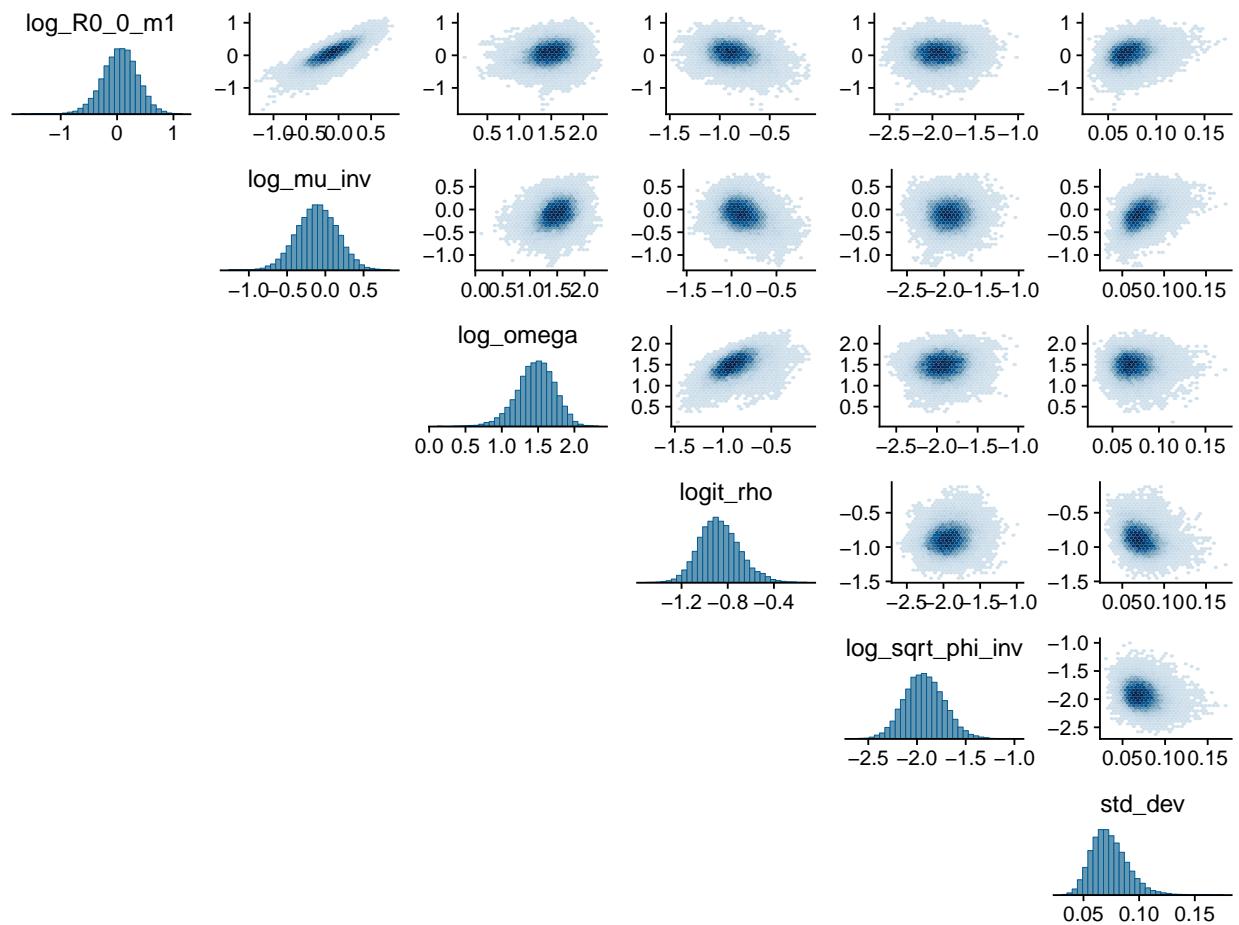


Figure C.2: Posterior histograms and pairwise hexplots for SIRS model parameters with time-varying force of infection.

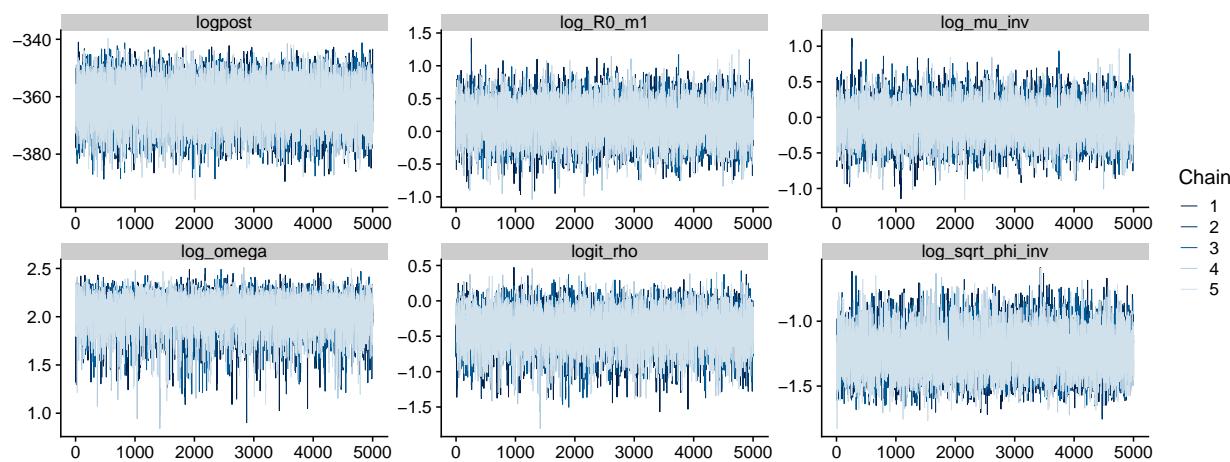


Figure C.3: Posterior traceplots for SIRS model parameters with time-homogeneous force of infection.

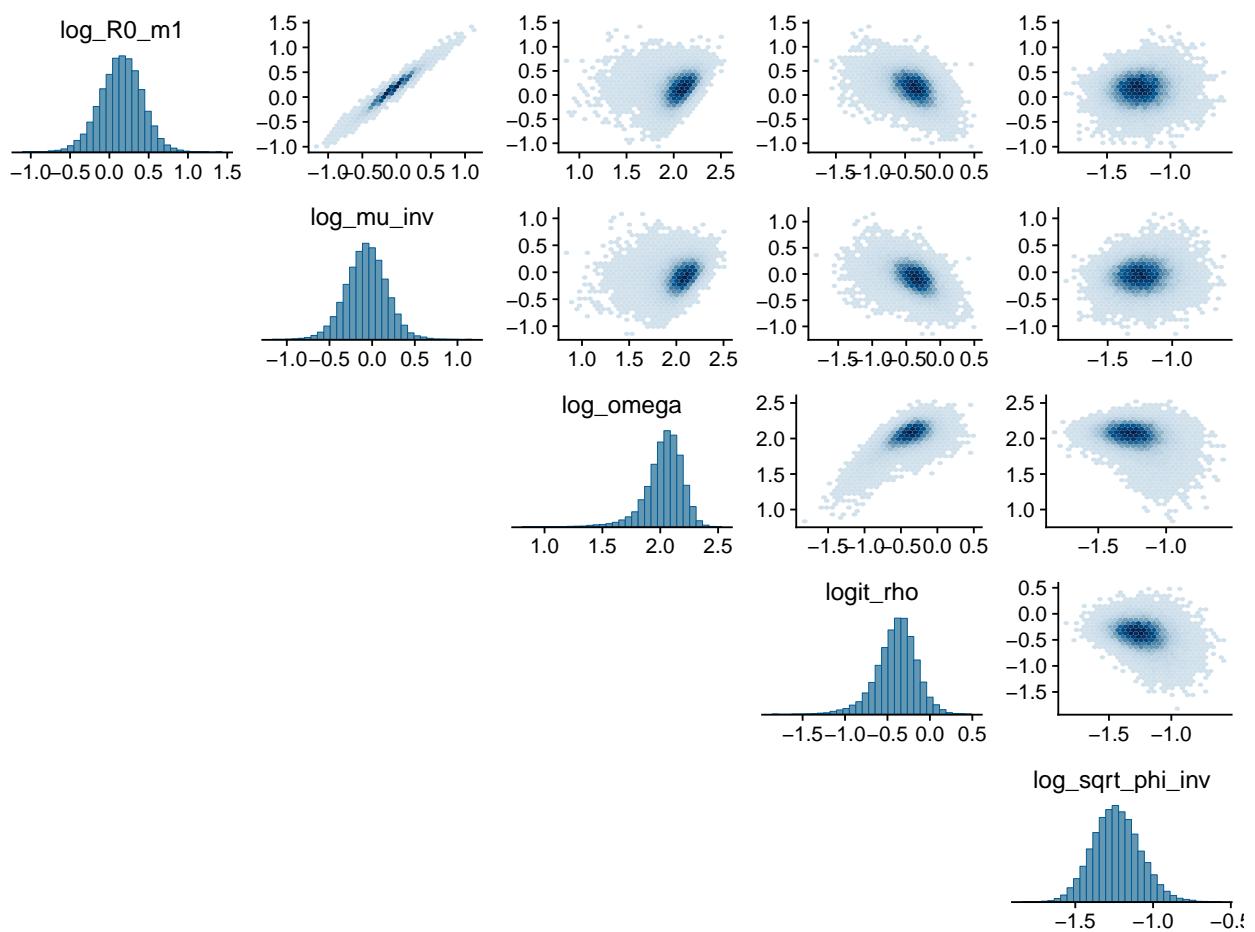


Figure C.4: Posterior histograms and pairwise hexplots for SIRS model parameters with time-homogeneous force of infection.

Posterior distributions of LNA draws

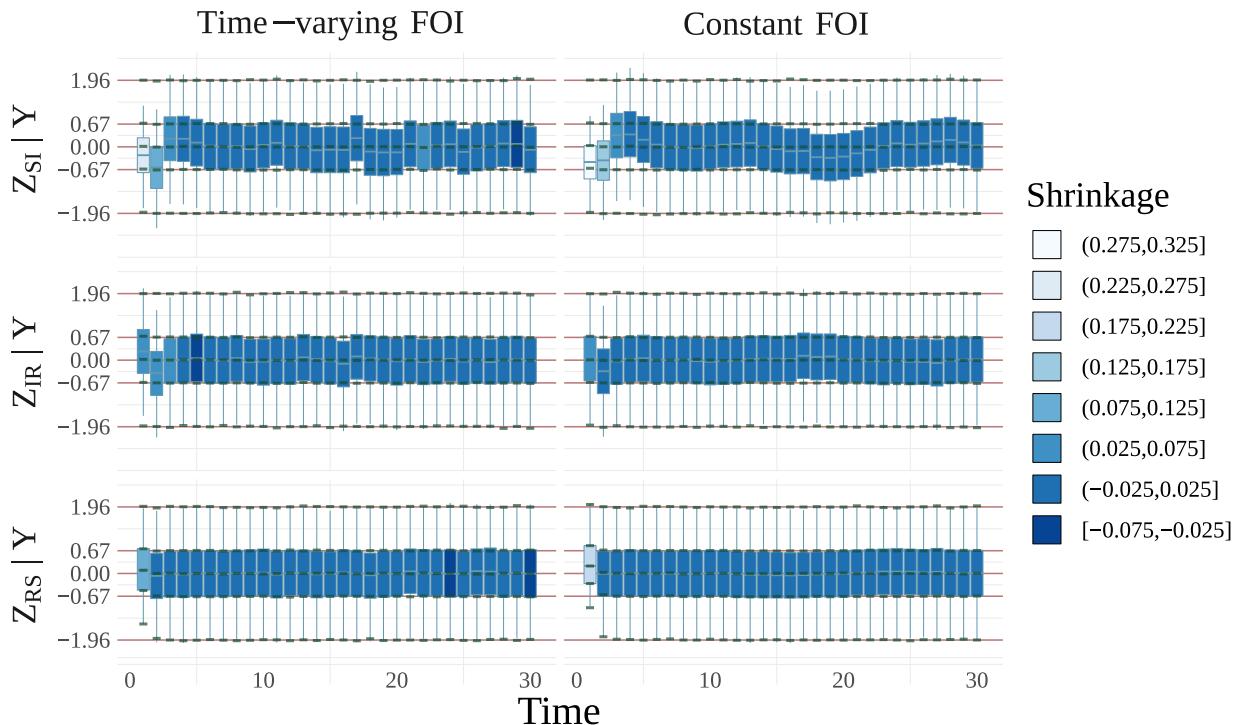


Figure C.5: Posterior distributions of the LNA draws for infection, recovery, and loss of immunity in SIRS models with time-varying and constant force of infection (blue boxplots). The lower and upper whisker tips correspond to the 2.5th and 97.5th posterior quantiles, the lower and upper hinges to the 25th and 75th quantiles, and the middle hash mark to the posterior median. The solid red lines are the theoretical quantiles of the posterior predictive distribution (or equivalently, the prior distribution) of the LNA draws, drawn at the quantiles of a standard normal distribution corresponding to the boxplot quantiles. The green ticks are the estimated quantiles of the posterior predictive distributions of the LNA draws, accounting for boundary conditions on the state space of the latent process and obtained by simulating LNA paths from the posterior predictive distribution. The posterior distributions of LNA draws are shaded according to the level of posterior shrinkage, computed as one minus the ratio of standard deviations of LNA draws in the posterior and prior.

C.2 Specification of Contact Rates

Contact rates within and between age-strata were obtained from the Finnish arm of the POLYMOD survey [156, 157], and were accessed via the `socialmixr` R package [79]. The survey was a population-based, prospective survey that recorded contact rates by sex, age, location, duration, and frequency. We queried the symmetric form of the estimated mean contact matrix aggregated age-strata in our analysis (0–19 years and 20+ years), which based on a sample of 1,006 individuals. A contact matrix is said to be symmetric if the total number of contacts between two groups is equal. In particular, if c_{ij} is the mean number of contacts an individual in group i has with members of group j , and group i is of size N_i , then symmetry requires that $c_{ij}N_i = c_{ji}N_j$. Due to the vagaries of sampling variability, this relation does not always hold. The symmetric contact matrix is obtained by taking $c'_{ij} = (c_{ij}N_i + c_{ji}N_j)/2N_i$.

The estimated mean contact rates are

$$\mathbf{C}_{tot} = \begin{matrix} & \begin{matrix} 0-19 & 20+ \end{matrix} \\ \begin{matrix} 0-19 \\ 20+ \end{matrix} & \begin{pmatrix} 7.44 & 5.05 \\ 1.55 & 8.56 \end{pmatrix} \end{matrix}.$$

Following [151], we row-normalize the matrix, thus removing differences in absolute contact rates. This is helpful in parameterizing and assigning sensible priors to other aspects of the model, such as basic reproduction numbers. The normalized contact rates are

$$\mathbf{C} = \begin{matrix} & \begin{matrix} 0-19 & 20+ \end{matrix} \\ \begin{matrix} 0-19 \\ 20+ \end{matrix} & \begin{pmatrix} 0.60 & 0.40 \\ 0.15 & 0.85 \end{pmatrix} \end{matrix}.$$

C.3 Mapping Standard Normal Draws to LNA Sample Paths with Forcings

Algorithm 11 Mapping standard normal draws onto LNA sample paths with forcings.

```

1: procedure DOLNA2( $\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I}, \boldsymbol{\xi}$ )
2:   initialize:  $\mathbf{X}(t_0) \leftarrow \mathbf{X}_0$ ,  $\mathbf{N}(t_0) \leftarrow \mathbf{0}$ ,  $\tilde{\mathbf{N}}(t_0) \leftarrow \mathbf{0}$ ,  $\boldsymbol{\mu}(t_0) \leftarrow \mathbf{0}$ ,  $\Sigma(t_0) \leftarrow \mathbf{0}$ 
3:   for  $\ell = 1, \dots, L$  do
4:      $\boldsymbol{\mu}(t_\ell)$ ,  $\Sigma(t_\ell) \leftarrow$  solutions to (4.15) and (4.17) over  $(t_{\ell-1}, t_\ell]$ 
5:      $\tilde{\mathbf{N}}(t_\ell) \leftarrow \boldsymbol{\mu}(t_\ell) + \Sigma(t_\ell)^{1/2} \mathbf{Z}(t_\ell)$   $\triangleright$  non-centered parameterization
6:      $\mathbf{N}(t_\ell) \leftarrow \mathbf{N}(t_{\ell-1}) + \exp(\tilde{\mathbf{N}}(t_\ell)) - \mathbf{1}$ 
7:   restart initial conditions:
       $\mathbf{X}(t_\ell) \leftarrow \mathbf{X}(t_{\ell-1}) + \mathbf{A}^T(\mathbf{N}(t_\ell) - \mathbf{N}(t_{\ell-1}))$ 
       $\mathbf{X}(t_\ell^+) \leftarrow \mathbf{X}(t_\ell) + \boldsymbol{\xi}(t_\ell)$ 
       $\tilde{\mathbf{N}}(t_\ell) \leftarrow \mathbf{0}$ ,  $\boldsymbol{\mu}(t_\ell) \leftarrow \mathbf{0}$ ,  $\Sigma(t_\ell) \leftarrow \mathbf{0}$ 
8: return  $\triangleright$  return incidence and/or prevalence sample paths
9:    $\mathbf{N} = \{\mathbf{N}(t_0), \mathbf{N}(t_1), \dots, \mathbf{N}(t_L)\}$ ,  $\mathbf{X} = \{\mathbf{X}(t_0), \mathbf{X}(t_1), \dots, \mathbf{X}(t_\ell)\}$ 

```

C.4 Joint Sampling of LNA Paths and Gaussian Markov Random Fields

Jointly sampling the LNA draws and the values of a GMRF, \mathbf{F} , in addition to possibly any other latent Gaussian parameters including the initial compartment volumes as in Section B.2, requires trivially modifying Algorithm 10 to include the additional parameters. We implement the same computational strategy used throughout Chapters 4 and 5 of using non-centered parameterizations for the parameters of interest and the LNA path. As we have noted, this has the advantage of breaking the autocorrelation that would otherwise lead to poor MCMC mixing when using a centered parameterization and alternating between updates to different parameters in the model. We will denote by $\mathbf{Z} = (\mathbf{Z}^X, \mathbf{Z}^{X_0}, \mathbf{Z}^F, \mathbf{Z}^{\theta_F})$ the multivariate standard normal draws for the latent LNA process, the initial compartment volumes, the draws to be mapped to the GMRF, and \mathbf{Z}^{θ_F} the draws that are mapped to the GMRF hyper-parameters that are to be updated along with the GMRF. Unless otherwise stated (and we never do so state), we will at the very least update the precision or standard deviation of the GMRF increments jointly with the field to avoid poor MCMC mixing. We will assume that non-centered parameterizations for the GMRF and the hyper-parameters are linear transformations and thus do not require Jacobian terms to be included in the ElliptSS acceptance probability. We will denote by \mathbf{m}_0 , \mathbf{m}_θ , \mathbf{V}_0 , and \mathbf{V}_θ the means and covariance matrices of the initial volumes and hyper-parameters, and by `doGMRF` the linear operator for mapping \mathbf{Z}^F to a GMRF sample, i.e., the GMRF values are obtained by $\text{doGMRF}(\mathbf{Z}^F; \boldsymbol{\theta}_F)$.

Algorithm 12 Sampling LNA draws, initial volumes, GMRF values, and GMRF hyper-parameters via elliptical slice sampling.

- 1: **procedure** `doELLIPTSS3`($\mathbf{Z} = (\mathbf{Z}_{cur}^X, \mathbf{Z}_{cur}^{X_0}, \mathbf{Z}^F, \mathbf{Z}^{\theta_F}), \boldsymbol{\theta}, \mathbf{Y}, \mathcal{I}, \omega = 2\pi$)
- 2: Sample ellipse: $\mathbf{Z}_{prop} \sim N(\mathbf{0}, \mathbf{I})$
- 3: Sample threshold: $u | \mathbf{x} \sim \text{Unif}(0, L(\mathbf{Y} | \text{doLNA}(\mathbf{Z}_{cur}, \boldsymbol{\theta}, \mathcal{I})))$
- 4: Position the bracket:

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

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

$$\phi \sim \text{Unif}(L_\psi, R_\psi)$$
- 5: Compute the proposal:

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

$$\implies \mathbf{X}'_0 = \mathbf{m}_0 + \mathbf{V}_0^{1/2} \mathbf{Z}^{X'_0}$$

$$\implies \boldsymbol{\theta}'_F = \mathbf{m}_\theta + \mathbf{V}_\theta^{1/2} \mathbf{Z}^{\theta'_F}$$

$$\implies \mathbf{F}' = \text{doGMRF}(\mathbf{Z}^F; \boldsymbol{\theta}_F)$$
- 6: **if** $L(\mathbf{Y} | \text{doLNA}(\mathbf{Z}^{X'}, \mathbf{X}'_0, \mathbf{F}', \boldsymbol{\theta}'_F, \boldsymbol{\theta}, \mathcal{I})) > u$ **then** accept \mathbf{Z}'
- 7: **return** \mathbf{Z}'
- 8: **else**
- 9: Shrink bracket and try a new angle:
If: $\phi < 0$ **then:** $L_\phi \leftarrow \phi$ **else:** $R_\phi \leftarrow \phi$
- 10: $\phi \sim \text{Unif}(L_\phi, R_\phi)$
- 12: **GoTo:** 5

C.5 Modeling Pandemic A(H1N1) Influenza in Finland — Additional Details and Supplementary Results

C.5.1 MCMC Details

We fit age–vaccination stratified SIRS ODE models with time varying dynamics to pandemic A(H1N1) influenza incidence data. The vaccination adjusted reproduction numbers were modeled using either a GMRF or as piecewise homogeneous within three epochs. The models were fit using the same prior regimes, given in Table C.3. We ran five chains for each model, initialized at random parameter values, for 100,000 iterations per chain. Model parameters (excluding noncentered GMRF draws and hyperparameters, as described in Section 5.3.4) were jointly updated via MVNSS (see Section 2.2.1). The empirical covariance for the MVNSS algorithm was adapted over the first 50,000 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 $\iota_n = 0.001(1 + 0.01n)^{-0.99}$. In the GMRF model, the non–centered GMRF draws were updated via ElliptSS along with their hyperparameters, hence MCMC alternated between one ElliptSS and one MVNSS update 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 the non–GMRF parameters were parameterized as in Table ??, and the GMRF was parameterized using the NCP given in Section 5.3.3. Convergence was assessed visually by inspection of traceplots of posterior samples. The models took between roughly one hour to run.

C.5.2 Prior Specification

Priors for effective population sizes

At the start of the modeling period, individuals in the population were assigned to either be susceptible or detached from the transmission process. The fraction of the population that was susceptible defines the effective population size. We identify the scale of the prior for the susceptible fraction of the population by inflating the total number of cases that were detected in each season by the case detection rate and the fraction of the effective population size that we would expect to be infected in a deterministic outbreak with SIR dynamics over a reasonable range of basic reproduction numbers. We ignored age structure for the sake of simplicity and assumed that the same fraction of youths and adults were initially susceptible. Results from a sensitivity analysis in which we assumed that a larger fraction of youths were initially susceptible are presented in Section C.5.4.

The final size relation for the SIR model [152] relates the fraction of the population that is infected under SIR dynamics, π , to the basic reproduction number via:

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

Roughly 10,000 cases were detected in all age groups over the course of both seasons. If the reproduction numbers and detection rates were stable across seasons, which might not be unreasonable if the lower attack rate in the second season is the result of depletion of susceptibles in the first season, a very crude estimate of the number of susceptibles is

$$\hat{s} = \frac{10000}{\rho(1 - \pi)},$$

where ρ is the detection rate. Hence, the fraction of the population that is in the susceptible compartment is $\hat{s}N$. Effective population sizes for different reproduction numbers that are typical of A(H1N1)pdm09 [21] and detection rates [194] are given in Table C.2.

In the main analysis presented in Section 5.4, we used a multivariate normal approximation to a dirichlet–multinomial distribution for the initial distribution of individuals of each

age stratum. The hyperparameter,

$$\boldsymbol{\alpha}_j = (S_{j,0}^{(u)} = 25, I_{j,0}^{(u)} = 0, R_{j,0}^{(u)} = 0, D_{j,0}^{(v)} = 75),$$

was centered so that 25% of each stratum was initially susceptible, with the central 80% of the prior mass between 20% and 30%.

Table C.2: Effective fraction of the population that is susceptible for different reproduction numbers and detection rates.

	Detection rate (ρ)			
R_0	0.0075	0.01	0.015	0.02
1.25	0.40	0.30	0.20	0.15
1.5	0.60	0.45	0.30	0.22
1.75	0.87	0.65	0.43	0.33

Priors for time-homogeneous parameters

Table C.3: Parameters and priors for age-vaccination stratified SIRS ODE models fit to the A(H1N1)pdm09 influenza outbreak in Finland, prior distributions used in the main analysis. Vaccination adjusted intrinsic reproduction numbers are defined as in equation (5.9). The form of the parameter listed in the leftmost column corresponds to the estimation scale used in the MCMC. All parameters in this table were jointly updated using multivariate normal slice sampling.

Param.	Interpretation	Prior	Median (90% Interval)	References/Justification
$\log(\psi_{Y,0})$	Intrinsic reproduction #	$N(\log(1.1), 0.06^2)$	$\psi_{Y,0}^{adj} = 1.10 (1.00, 1.21)$	Nearly endemic dynamics [21]
$\log(\psi_{Y,19})$	Intrinsic reproduction #	$N(\log(1.5), 0.175^2)$	$\psi_{Y,19} = 1.50 (1.12, 2.00)$	
$\log(\psi_{Y,71}^{adj})$	Adjusted intrinsic reproduction #	$N(\log(1.3), 0.11^2)$	$\psi_{Y,71}^{adj} = 1.3 (1.08, 1.56)$	[21] and $\psi_{Y,71}^{adj} < \psi_{A,71}^{adj}$
$\log(\psi_{A,0}^{adj})$	Intrinsic reproduction #	$N(\log(1.05), 0.025^2)$	$\psi_{A,0}^{adj} = 1.10 (1.00, 1.21)$	Nearly endemic dynamics [21]
$\log(\psi_{A,19}^{adj})$	Intrinsic reproduction #	$N(\log(1.3), 0.11^2)$	$\psi_{A,19}^{adj} = 1.3 (1.10, 1.78)$	[21] and $\psi_{A,19}^{adj} < \psi_{Y,19}^{adj}$
$\log(\psi_{A,71}^{adj})$	Adjusted intrinsic reproduction #	$N(\log(1.4), 0.145^2)$	$\psi_{A,71}^{adj} = 1.4 (1.10, 1.78)$	[21] and $\psi_{A,71}^{adj} > \psi_{Y,71}^{adj}$
$\log(\alpha_{Y,0} N_Y s_y)$	Scaled rate of exogenous infection	$N(1.8, 0.54^2)$	6.0 (2.5, 14.7)	Small w.r.t. $s_Y N_Y$
$\log(\alpha_{Y,19} \alpha_{Y,0} N_Y s_Y)$	Scaled rate of exogenous infection	$N(2.3, 0.54^2)$	10.0 (4.1, 24.2)	Small w.r.t. $s_Y N_Y$
$\log(\alpha_{Y,71} \alpha_{Y,0} N_Y s_Y)$	Scaled rate of exogenous infection	$N(1.6, 0.54^2)$	5.0 (2.0, 12.0)	Small w.r.t. $s_Y N_Y$
$\log(\alpha_{A,0} \alpha_{A,0} N_A s_A)$	Scaled rate of exogenous infection	$N(3, 0.54^2)$	20 (8.2, 48.8)	Small w.r.t. $s_A N_A$
$\log(\alpha_{A,19} N_A s_A)$	Scaled rate of exogenous infection	$N(3.2, 0.54^2)$	24.5 (10.1, 59.6)	Small w.r.t. $s_A N_A$
$\log(\alpha_{A,71} N_A s_A)$	Scaled rate of exogenous infection	$N(2, 0.54^2)$	7.4 (3.0, 18.0)	Small w.r.t. $s_A N_A$
$7/\mu_Y - 1$	Avg. infec. period, youths (days-1)	$N(0.41, 0.32^2)$	$7/\mu_Y = 2.5 (1.9, 3.55)$	[21, 42, 54, 214]
$7/\mu_A - 1$	Avg. infec. period, adults (days-1)	$N(0.22, 0.165^2)$	$7/\mu_A = 2.25 (1.95, 2.63)$	$1/\mu_Y > 1/\mu_A$, [21, 42, 54, 214]
$\log(52/\omega)$	Mean duration of immunity (years)	$N(1.1, 0.3^2)$	$52/\omega = 3 (1.8, 4.9)$	Pr(Lose immunity in 1 year) = 0.3
$\text{logit}(\nu)$	1 - VE for susceptibility	$N(-1.39, 1)$	$\nu = 0.2 (0.05, 0.56)$	[135, 199]
$\text{logit}(\rho_Y)$	Mean case detection rate, youths	$N(\text{logit}(0.0125), 0.31^2)$	$\rho = 0.0125 (0.0075, 0.02)$	[194]
$\text{logit}(\rho_A)$	Mean case detection rate, adults	$N(\text{logit}(0.01), 0.31^2)$	$\rho = 0.01 (0.006, 0.017)$	$\rho_A < \rho_Y$, [194]
$\log(1/\sqrt{\phi_Y})$	Neg. binom. overdispersion, youths	$\exp(1)$	0.69 (0.05, 3.0)	It works.
$\log(1/\sqrt{\phi_A})$	Neg. binom. overdispersion, adults	$\exp(1)$	0.69 (0.05, 3.0)	It works.

Priors for GMRFs

Table C.4: Priors for initial values of GMRFs for the numbers of exogenous infection in each age stratum. Parameters are subscripted by age stratum and week number, indexed from epiweek 15, 2009.

Parameter	Prior	Median (90% Interval)
$\log(\alpha_{Y,0})$	$N(1.8, 0.54^2)$	6.0 (2.5, 14.7)
$\log(\alpha_{Y,19})$	$N(2.3, 0.54^2)$	10.0 (4.1, 24.2)
$\log(\alpha_{Y,71})$	$N(1.6, 0.54^2)$	5.0 (2.0, 12.0)
$\log(\alpha_{A,0})$	$N(3, 0.54^2)$	20 (8.2, 48.8)
$\log(\alpha_{A,19})$	$N(3.2, 0.54^2)$	24.5 (10.1, 59.6)
$\log(\alpha_{A,71})$	$N(2, 0.54^2)$	7.4 (3.0, 18.0)

Table C.5: Priors for standard deviations of GMRF increments of reproduction numbers and rates of exogenous infection. Parameters are superscripted by whether they apply to a GRMF for intrinsic reproduction numbers, ψ , or for numbers of exogenous infections, α , and are subscripted by age stratum and week number, indexed from epiweek 15, 2009.

Parameter	Prior	Median (90% Interval)
$\log(\sigma_{Y,0}^\psi)$	$N(-5.25, 0.125^2)$	0.0052 (0.0043, 0.0064)
$\log(\sigma_{Y,19}^\psi)$	$N(-4, 0.125^2)$	0.018 (0.015, 0.022)
$\log(\sigma_{Y,71}^\psi)$	$N(-5.25, 0.125^2)$	0.014 (0.012, 0.018)
$\log(\sigma_{A,0}^\psi)$	$N(-5.25, 0.125^2)$	0.0052 (0.0043, 0.0064)
$\log(\sigma_{A,19}^\psi)$	$N(-3.75, 0.125^2)$	0.024 (0.019, 0.029)
$\log(\sigma_{A,71}^\psi)$	$N(-5.25, 0.125^2)$	0.014 (0.012, 0.018)
$\log(\sigma_{Y,0}^\alpha)$	$N(-4.25, 0.25^2)$	0.014 (0.0095, 0.022)
$\log(\sigma_{Y,19}^\alpha)$	$N(-3.75, 0.25^2)$	0.024 (0.016, 0.035)
$\log(\sigma_{A,71}^\alpha)$	$N(-4, 0.25^2)$	0.018 (0.012, 0.028)
$\log(\sigma_{Y,0}^\alpha)$	$N(-4.25, 0.25^2)$	0.014 (0.0095, 0.022)
$\log(\sigma_{A,19}^\alpha)$	$N(-3.75, 0.25^2)$	0.024 (0.016, 0.035)
$\log(\sigma_{A,71}^\alpha)$	$N(-4, 0.25^2)$	0.018 (0.012, 0.028)

Prior distribution of vaccination adjusted basic reproduction numbers

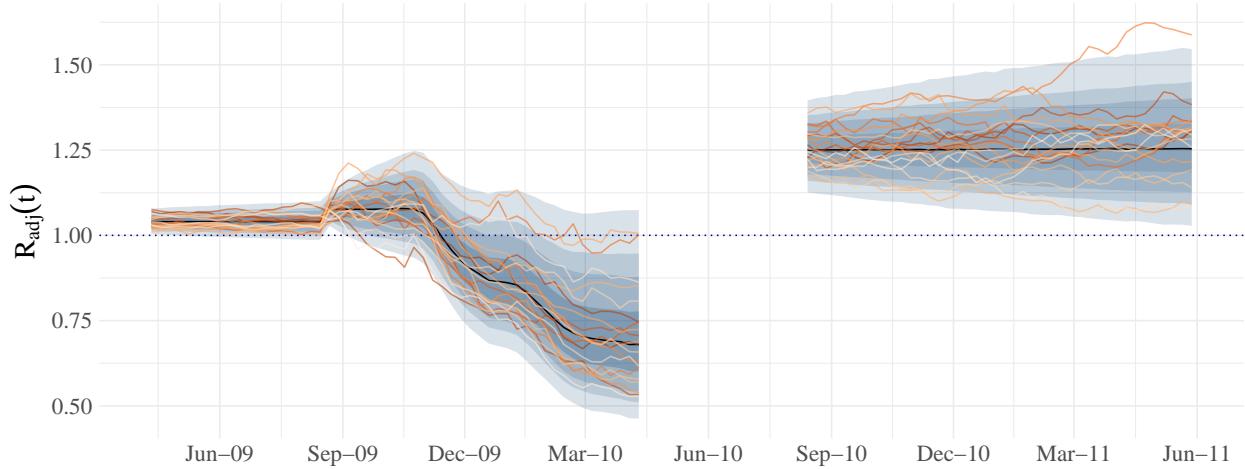


Figure C.6: Induced prior distribution of for vaccination adjusted basic reproduction numbers A(H1N1)pdm09 influenza in Finland. Shaded bands correspond to the central 50%, 80%, and 95% prior probability intervals.

Prior distributions of numbers of exogenous infections

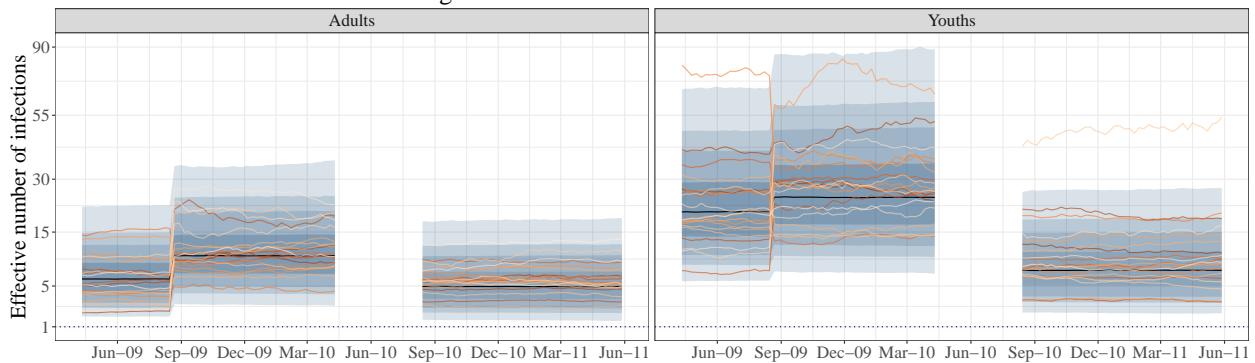


Figure C.7: Induced prior distribution of for the effective number of exogenous infections among youths and adults. Shaded bands correspond to the central 50%, 80%, and 95% prior probability intervals.

C.5.3 Additional Results

Stratified SIRS model with time-varying dynamics

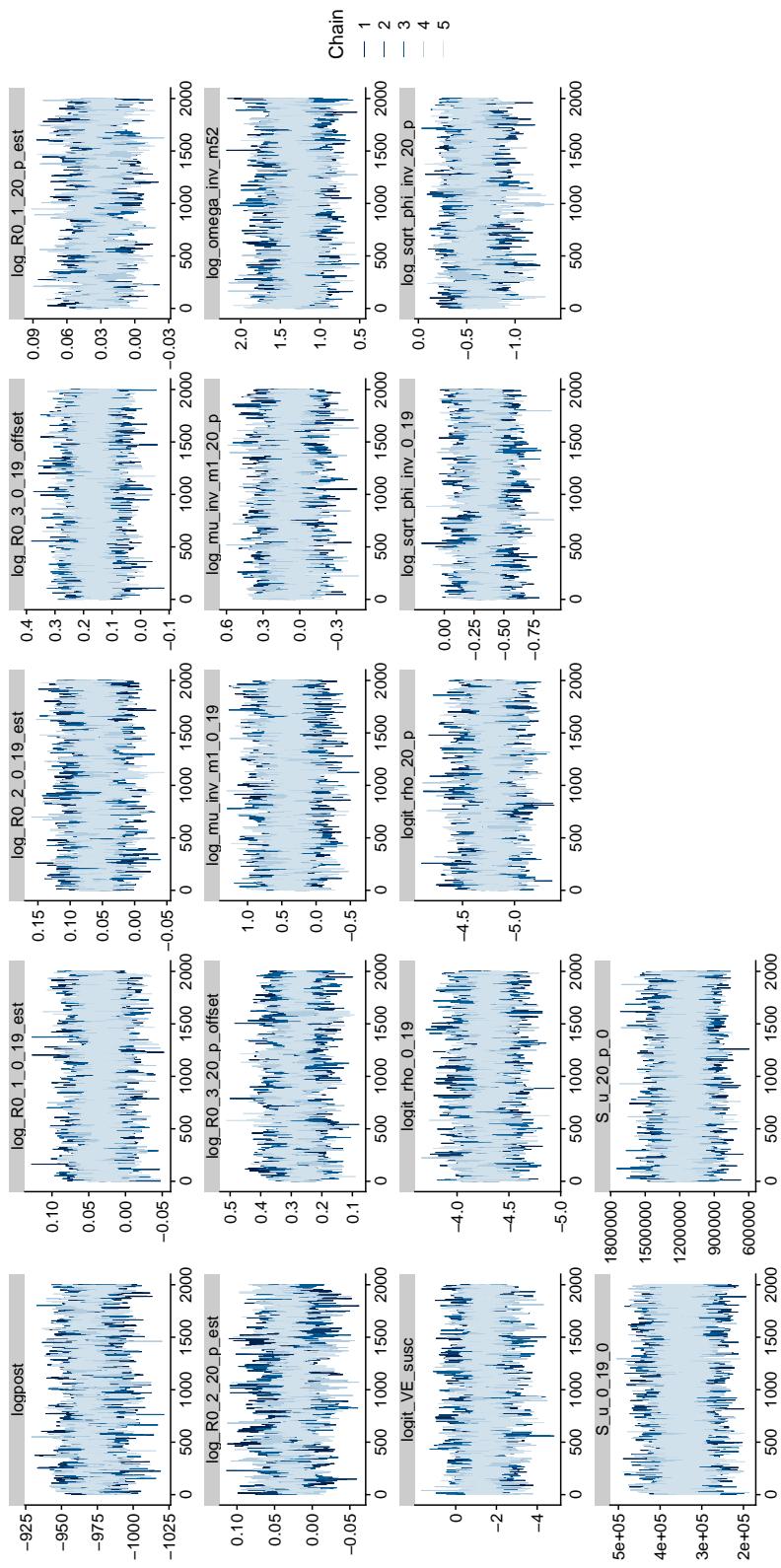


Figure C.8: Posterior traceplots for a stratified SIRS ODE model with time-varying dynamics fit to data from the A(H1N1) influenza pandemic in Finland.

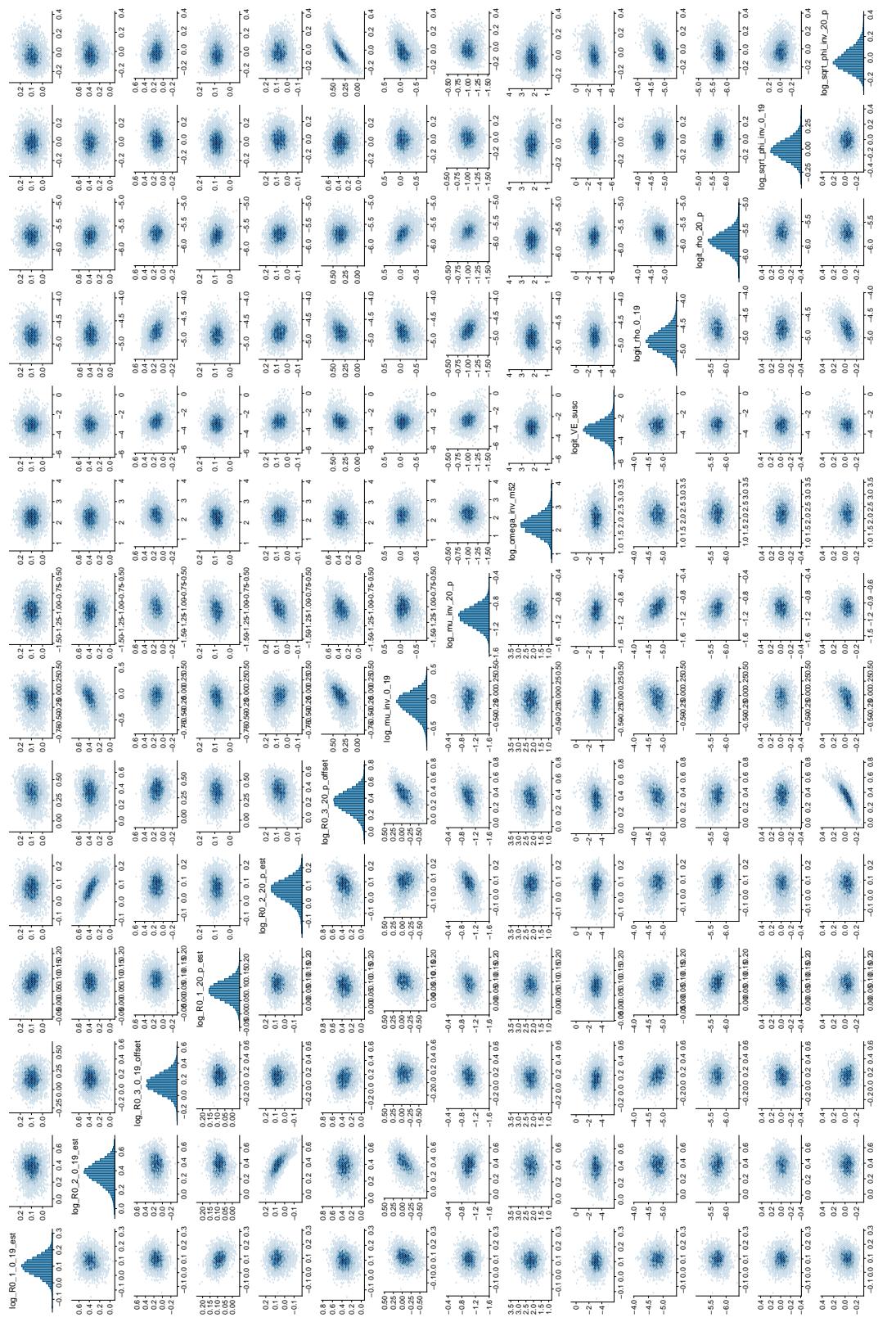


Figure C.9: Histograms and pairwise scatterplots of posterior samples for the parameters of a stratified SIRS ODE model with time varying dynamics that were sampled via MVNNS. The model was fit to data from the A(H1N1) influenza pandemic in Finland.

Stratified SIRS model with piecewise-homogeneous dynamics

Table C.6: Parameters and priors for age-vaccination stratified piecewise homogeneous SIRS ODE models fit to the A(H1N1)pdm09 influenza outbreak in Finland, prior distributions used in the main analysis. Vaccination adjusted intrinsic reproduction numbers are defined as in equation (5.9)

Parameter	Interpretation	Prior	Median (95% Interval)	References
$\log(\psi_{Y,0}^{adj})$	Intrinsic reproduction #	$N(\log(1.05), 0.025^2)$	$\psi_{Y,0}^{adj} = 1.05 (1.00, 1.09)$	\approx endemic dynamics
$\log(\psi_{Y,19}^{adj})$	Intrinsic reproduction #	$N(\log(1.075), 0.03^2)$	$\psi_{Y,19}^{adj} = 1.075 (1.00, 1.09)$	Nearly endemic and $\psi_{Y,19}^{adj} > \psi_{Y,0}^{adj}$
$\log(\psi_{Y,71}^{adj})$	Adjusted intrinsic reproduction #	$N(\log(1.2), 0.06^2)$	$\psi_{Y,71}^{adj} = 1.2 (1.09, 1.32)$	Nearly endemic with 10% immune
$\log(\psi_{A,0}^{adj})$	Intrinsic reproduction #	$N(\log(1.05), 0.025^2)$	$\psi_{A,0}^{adj} = 1.05 (1.00, 1.09)$	Nearly endemic dynamics
$\log(\psi_{A,19}^{adj})$	Intrinsic reproduction #	$N(\log(1.075), 0.03^2)$	$\psi_{A,19}^{adj} = 1.075 (1.00, 1.09)$	Nearly endemic and $\psi_{A,19}^{adj} > \psi_{A,0}^{adj}$
$\log(\psi_{A,71}^{adj})$	Adjusted intrinsic reproduction #	$N(\log(1.3), 0.07^2)$	$\psi_{A,71}^{adj} = 1.3 (1.16, 1.46)$	Nearly endemic w/ 10% immune, $\psi_{A,71}^{adj} > \psi_{A,0}^{adj}$
$7/\mu_Y - 1$	Avg. infectious period, youths (days-1)	$N(0.41, 0.32^2)$	$7/\mu_Y = 2.5 (1.9, 3.55)$	[21, 42, 54, 214]
$7/\mu_A - 1$	Avg. infectious period, adults (days-1)	$N(0.22, 0.165^2)$	$7/\mu_A = 2.25 (1.95, 2.63)$	$1/\mu_Y > 1/\mu_A$, [21, 42, 54, 214]
$\log(52/\omega)$	Mean duration of immunity (years)	$N(1.1, 0.3^2)$	$52/\omega = 3 (1.8, 4.9)$	Pr(Lose immunity in 1 year) = 0.3
$\text{logit}(\nu)$	1 - VE for susceptibility	$N(-1.39, 1)$	$\nu = 0.2 (0.05, 0.56)$	[135, 199]
$\text{logit}(\rho_Y)$	Mean case detection rate, youths	$N(\text{logit}(0.0125), 0.31^2)$	$\rho = 0.0125 (0.0075, 0.02)$	[194]
$\text{logit}(\rho_A)$	Mean case detection rate, adults	$N(\text{logit}(0.01), 0.31^2)$	$\rho = 0.01 (0.006, 0.017)$	$\rho_A < \rho_Y$, [194]
$\log(1/\sqrt{\phi_Y})$	Neg. binom. overdispersion, youths	$\exp(1)$	$0.69 (0.05, 3.0)$	It works.
$\log(1/\sqrt{\phi_A})$	Neg. binom. overdispersion, adults	$\exp(1)$	$0.69 (0.05, 3.0)$	It works.

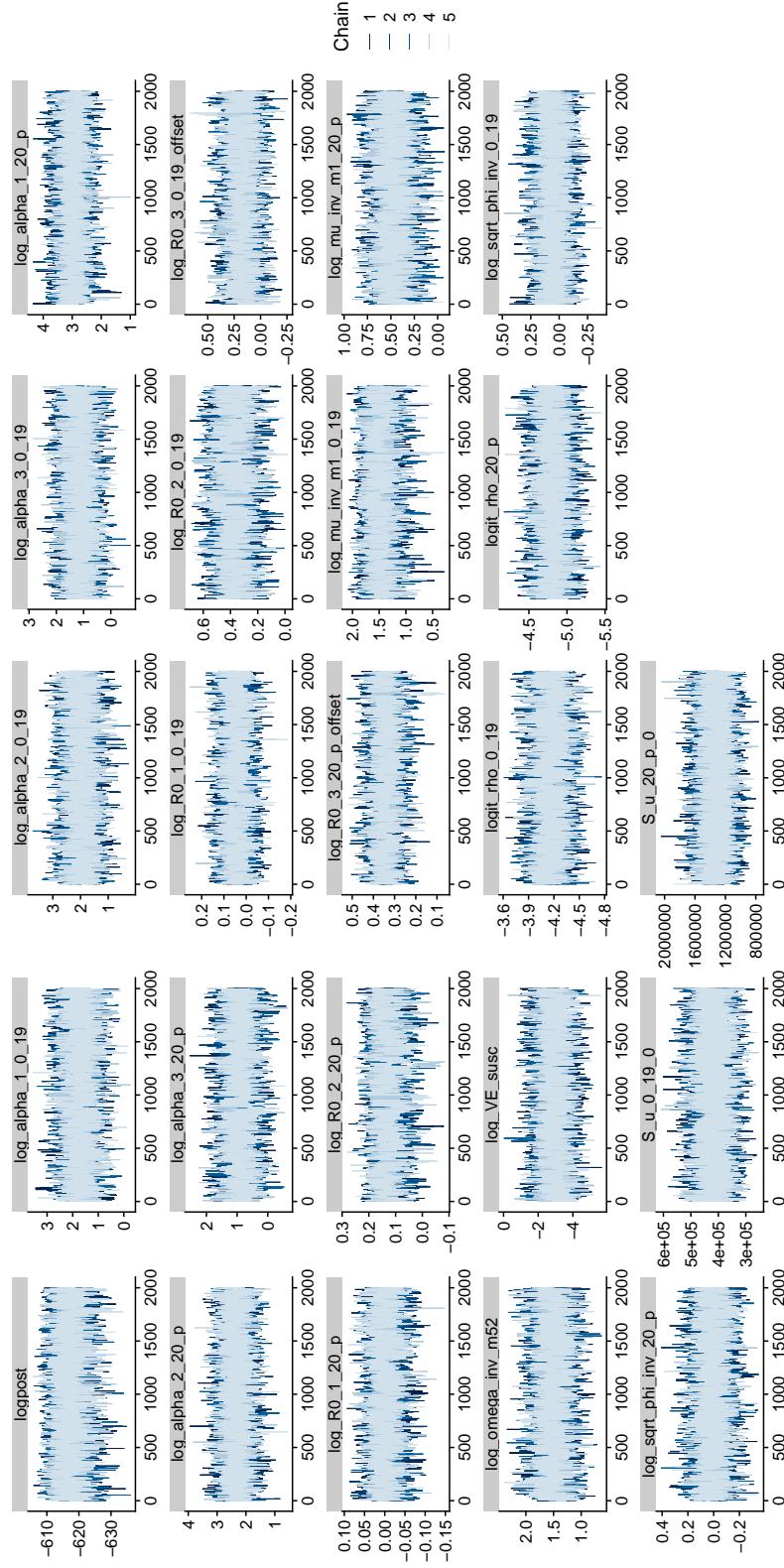


Figure C.10: Posterior traceplots for a stratified SIRS ODE model with piecewise-homogeneous dynamics fit to data from the A(H1N1) influenza pandemic in Finland.

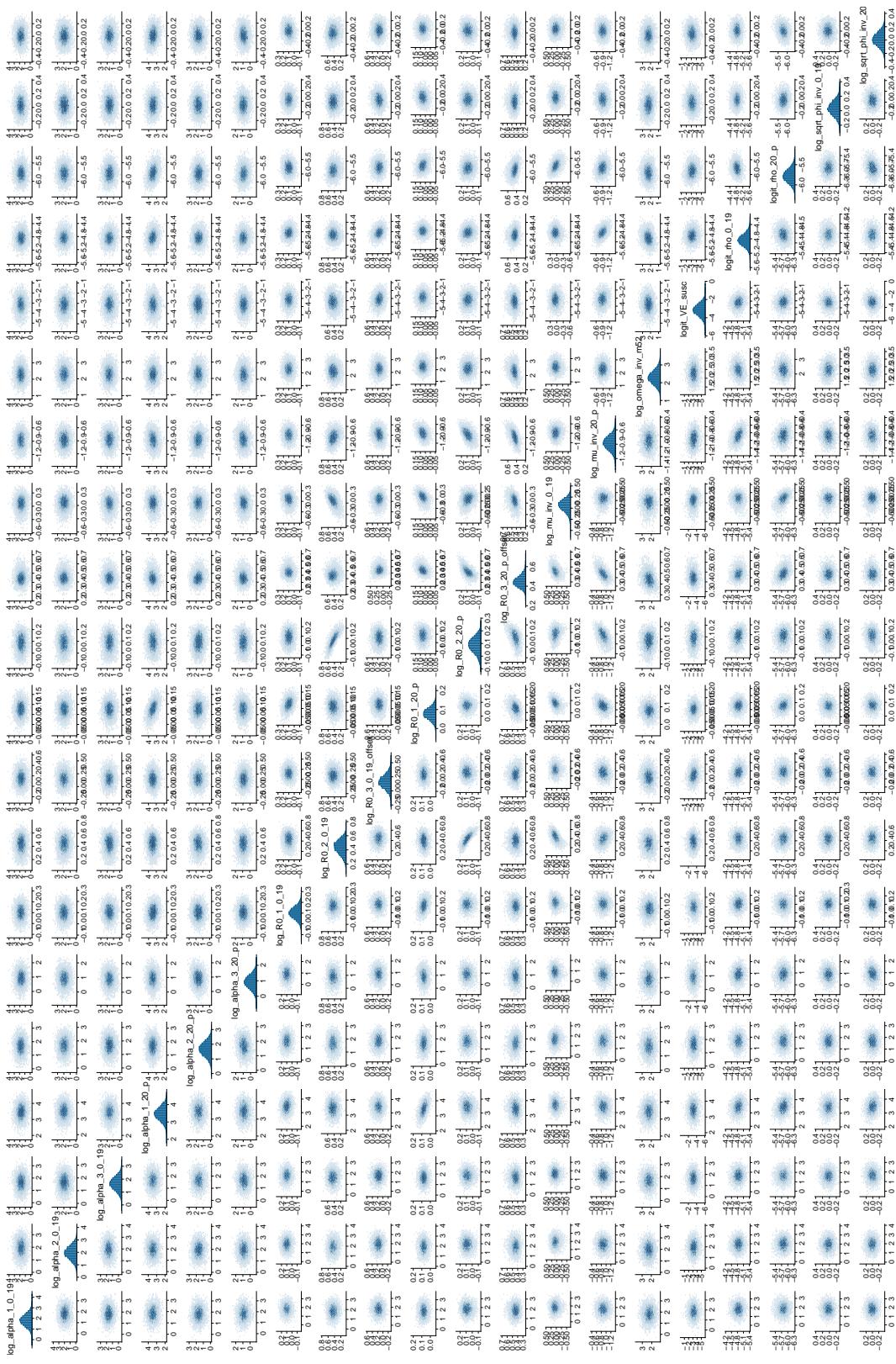


Figure C.11: Histograms and pairwise scatterplots of posterior samples for the parameters of a stratified SIRS ODE model with piecewise homogeneous dynamics that were sampled via MVNSS. The model was fit to data from the A(H1N1) influenza pandemic in Finland.

C.5.4 Sensitivity to Distribution of Initial Numbers of Susceptibles

We fit a supplementary model to explore the sensitivity of estimates of model parameters and incidence to the prior for the susceptible fraction of each age stratum. We again used a multivariate normal approximation to a dirichlet–multinomial distribution for the initial distribution of individuals. The hyperparameter for youths was

$$\boldsymbol{\alpha}_Y = (S_{Y,0}^{(u)} = 25, I_{Y,0}^{(u)} = 0, R_{Y,0}^{(u)} = 0, D_{Y,0}^{(v)} = 25),$$

so that 50% of youths were initially susceptible, with the central 80% of the prior mass between 40% and 60%. The hyperparameter for adults was

$$\boldsymbol{\alpha}_A = (S_{A,0}^{(u)} = 20, I_{A,0}^{(u)} = 0, R_{A,0}^{(u)} = 0, D_{A,0}^{(v)} = 60),$$

so that 25% of each stratum was initially susceptible, with the central 80% of the prior mass between 19% and 31%. The other priors and various aspects of the model and MCMC were otherwise the same as in the main model with time varying dynamics.

The effect of centering the prior for the fraction of youths who were susceptible, but roughly holding constant the fraction of susceptible adults (the prior in the sensitivity analysis was slightly more diffuse), is that attack rates among youths are inflated, while attack rates among adults are largely similar (Table C.7). Estimates of the rate parameters governing the model dynamics are largely unchanged from estimates from the main analysis (Table C.8). However, we find that the detection rate among youths is meaningfully lower than the estimated mean case detection rate in the main analysis, which is unsurprising since the estimated incidence is higher. The corresponding estimates for adults are essentially unchanged. In both analyses, the estimated susceptible fractions of each age stratum essentially recover the prior. This suggests that there is not enough signal in the partially observed incidence data to disentangle the effective population size and detection rate.

Table C.7: Estimated infections (thousands) and attack rates by season and age stratum. Attack rates are calculated as the number of infections divided by the size of each stratum, assuming that cases are unique.

Main analysis						
	Season 1		Season 2		Both Seasons	
	Incidence ($\times 10^3$)	Attack rate (%)	Incidence ($\times 10^3$)	Attack rate (%)	Incidence ($\times 10^3$)	Attack rate (%)
Ages 0-19	174 (127, 231)	14.2 (10.4, 18.9)	68.6 (46.6, 100)	5.6 (3.8, 8.2)	244 (181, 321)	19.9 (14.8, 26.2)
Ages 20+	356 (243, 501)	8.6 (5.9, 12.1)	171 (117, 245)	4.1 (2.8, 5.9)	530 (378, 720)	12.8 (9.2, 17.4)
All ages	532 (393, 703)	9.9 (7.4, 13.1)	240 (172, 333)	4.5 (3.2, 6.2)	774 (586, 1,010)	14.5 (11, 18.8)

Higher susceptible %						
	Season 1		Season 2		Both Seasons	
	Incidence ($\times 10^3$)	Attack rate (%)	Incidence ($\times 10^3$)	Attack rate (%)	Incidence ($\times 10^3$)	Attack rate (%)
Ages 0-19	281 (198, 384)	23 (16.2, 31.4)	95.2 (66, 141)	7.8 (5.4, 11.5)	377 (276, 502)	30.9 (22.6, 41)
Ages 20+	349 (232, 496)	8.4 (5.6, 12)	175 (119, 251)	4.2 (2.9, 6.1)	526 (367, 717)	12.7 (8.9, 17.4)
All ages	632 (470, 823)	11.8 (8.8, 15.4)	272 (193, 374)	5.1 (3.6, 7)	907 (689, 1,170)	16.9 (12.9, 21.8)

Table C.8: Posterior estimates of SIRS model parameters from the main analysis compared to estimates obtained with a higher number of susceptible youths. Estimated posterior medians (95% Bayesian credible intervals).

Parameter	Interpretation	Dynamics	
		Main analysis	Higher susceptible %
$\psi_{Y,0}$	Intrinsic R_0 for youths at epiweek 15, 2009	1 (0.99, 1.1)	1 (0.99, 1.1)
$\psi_{Y,19}$	Intrinsic R_0 for youths at epiweek 35, 2009	1.1 (1, 1.1)	1.1 (1, 1.1)
$\psi_{Y,71}^{adj}$	Vaccination adjusted intrinsic R_0 for youths at epiweek 33, 2010	1.2 (1, 1.3)	1.2 (1, 1.3)
$\psi_{A,0}$	Intrinsic R_0 for adults at epiweek 15, 2009	1 (1, 1.1)	1 (1, 1.1)
$\psi_{A,19}$	Intrinsic R_0 for adults at epiweek 35, 2009	1 (0.98, 1.1)	1 (0.97, 1.1)
$\psi_{A,71}^{adj}$	Vaccination adjusted intrinsic R_0 for adults at epiweek 33, 2010	1.3 (1.2, 1.5)	1.3 (1.2, 1.5)
$1/\mu_Y$	Mean infectious period for youths (days)	2.5 (1.9, 3.5)	2.3 (1.8, 3.1)
$1/\mu_A$	Mean infectious period for adults (days)	2.1 (1.8, 2.4)	2.1 (1.8, 2.5)
$1/\omega$	Mean duration of immunity (years)	3.6 (2.4, 5.8)	3.6 (2.4, 5.8)
ν	1 - VE for susceptibility	0.19 (0.039, 0.54)	0.17 (0.036, 0.5)
s_Y	% susceptible youths, $S_Y^{(u)}(t_0)/N_Y$	0.28 (0.21, 0.36)	0.5 (0.38, 0.63)
s_A	% susceptible adults, $S_A^{(u)}(t_0)/N_A$	0.28 (0.21, 0.35)	0.28 (0.21, 0.36)
ρ_Y	Mean case detection rate for youths	0.013 (0.0097, 0.019)	0.0099 (0.0072, 0.013)
ρ_A	Mean case detection rate for adults	0.0084 (0.0061, 0.012)	0.0081 (0.006, 0.011)
$1/\sqrt{\phi_Y}$	Negative binomial overdispersion for youths	0.71 (0.55, 0.92)	0.65 (0.5, 0.86)
$1/\sqrt{\phi_A}$	Negative binomial overdispersion for adults	0.56 (0.37, 0.76)	0.54 (0.38, 0.74)