



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Advances in the estimation of the reproduction number from compartmental models using contemporary Monte Carlo methods
Author(s)	Andrade, Jair
Publication Date	2024-03-08
Publisher	NUI Galway
Item record	http://hdl.handle.net/10379/18084

Downloaded 2024-04-28T14:14:49Z

Some rights reserved. For more information, please see the item record link above.





OLLSCOIL NA GAILLIMHÉ
UNIVERSITY OF GALWAY

SCHOOL OF COMPUTER SCIENCE
COLLEGE OF SCIENCE AND ENGINEERING
UNIVERSITY OF GALWAY

**Advances in the estimation of the
reproduction number from
compartmental models using
contemporary Monte Carlo methods**

JAIR ANDRADE, M.Sc., B.Eng.

Supervised by:
Prof. Jim Duggan

A thesis submitted for the degree of Doctor of Philosophy

November 2023

Abstract

The reproduction number represents the average number of secondary cases generated by a primary case. If the population is completely susceptible, it is referred to as the basic reproduction number (\mathcal{R}_0) and theoretically determines whether the pathogen can invade the population. Moreover, its magnitude is proportional to the effort needed to control the disease. Conversely, if the infection is spreading, it is referred to as the effective reproduction number (\mathcal{R}_t). It serves as an indicator of how extrinsic and intrinsic factors have affected transmission at any given time. Both \mathcal{R}_0 and \mathcal{R}_t can be estimated from compartmental models fitted to time series data. However, these estimates are sensitive to both model assumptions and calibration methods. Here, we show that by adhering to a rigorous inference workflow and utilising state-of-the-art algorithms and visualisation tools, one can obtain robust estimates. Using Hamiltonian Monte Carlo in a Bayesian approach, we found a linear relationship between the mean generation time and \mathcal{R}_0 . This discovery allowed us to formulate a parameterisation that produces accurate \mathcal{R}_0 estimates regardless of the distribution of the epidemiological delays. On the other hand, we demonstrated, through a complementary workflow that spanned three Data Generating Processes (semi-deterministic and deterministic) and both schools of thought for statistical inference, that incorporating mobility data into the workflow can reduce the uncertainty in \mathcal{R}_t estimates. Nevertheless, this incorporation requires caution, given that mobility data can only be a proxy measurement of the transmission rate. Our results emphasise the importance of envisioning model calibration as a learning process that confronts embedded assumptions. We anticipate these findings will serve as a reference point for modellers that fit SIR-like structures to time-series data. These guidelines include which information to prioritise, how to approach the inference procedure, and how to interpret calibration results.

Acknowledgements

Thank you:

Jim Duggan, Martha Ortiz, Jairo Andrade, Pilín Andrade, Yuki Andrade, Milú Andrade, Darío Ortiz, Alfonso Ortiz, Nayibe Navarro, Roberto Ortiz, Manuel Pacheco, Julián Lizarazo, Joaquín Vargas, Luis Carlos León, Javier León, Karen Morales, Gina Román, Anyon Cadena, Ailin Martínez, Hugo Andrade, Diana Basto, Nelson Cardozo, David Wheat, Ken Lyons, Karl Mason, Kevin McDonnell, and Andrew Simpkin.

I would not be in this position without your support.

Declaration

This thesis has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree other than Doctor of Philosophy of the University of Galway. This thesis is the result of my own investigations.

The material contained in this thesis has appeared in the following published papers:

1. Andrade, J., Duggan, J., 2021. A Bayesian approach to calibrate system dynamics models using Hamiltonian Monte Carlo. *System Dynamics Review*, 37(4).
2. Andrade, J., Duggan, J., 2022. Inferring the effective reproductive number from deterministic and semi-deterministic compartmental models using incidence and mobility data. *PLOS Computational Biology*, 18(6).
3. Andrade, J., Duggan, J., 2023. Anchoring the mean generation time in the SEIR to mitigate biases in \mathfrak{R}_0 estimates due to uncertainty in the distribution of the epidemiological delays. *Royal Society Open Science*, 10(8).



Contents

Abstract	i
Acknowledgements	ii
Declaration	iii
1 Introduction	1
1.1 Reproduction number	1
1.2 Compartmental models	4
1.3 Inference	7
1.3.1 Frequentist approach	8
1.3.2 Bayesian approach	11
1.4 Research questions	13
2 Hamiltonian Monte Carlo in a Bayesian workflow	17
2.1 Introduction	17
2.2 Context	20
2.3 Bayesian inference workflow	22
2.3.1 Prior information	23
2.3.2 Observational model and probability of the data	25
2.3.3 Expectation	28
2.3.4 Markov chain Monte Carlo	29
2.3.5 Posterior distribution	34
2.4 Performance comparison	40
2.5 Conclusion	42
3 The basic reproduction number and the mean generation time	43
3.1 Introduction	43
3.2 Data Generating Process	45
3.2.1 The system (latent) component	46

Contents

3.2.2	Measurement component	49
3.3	Inference	51
3.3.1	Three unknowns (traditional): β, ρ, I_0	53
3.3.2	Four unknowns: β, ρ, I_0, γ	54
3.3.3	Three unknowns (alternative): $\mathfrak{R}_0^{-1}, \rho, I_0$	58
3.3.4	Misspecifying the latent period distribution	61
3.3.5	Sensitivity analysis	63
3.4	Application to Influenza A	65
3.5	Conclusion	66
4	Time-varying transmission rate	69
4.1	Introduction	69
4.2	Results	72
4.2.1	Context	72
4.3	State-Space Models	75
4.3.1	DGP1 - Geometric Brownian Motion	76
4.3.2	DPG2 - Cox-Ingersoll-Ross	84
4.3.3	DGP3 - Adaptive expectations	85
4.4	Discussion	90
4.5	Materials and methods	94
4.5.1	SEI3R profile	95
4.5.2	Basic and effective reproductive number	95
4.5.3	Filtering distribution	96
5	Conclusion	97
5.1	Research findings	97
5.2	Contribution	101
5.3	Limitations and future work	103
5.4	Final remarks	104
A	Supplementary Information (Ch. 2)	107
A.1	Measurement model	107
A.2	Parameter space exploration by MCMC algorithm	109
A.3	Potential scale reduction factor (\hat{R}) & Effective Sample Size (\hat{n}_{eff})	110
A.4	Synthetic data	113
A.4.1	Case 1	113
A.4.2	Case 2	115
A.4.3	Case 3	123

A.5	Point estimates vs neighbourhoods	126
A.6	Mean Absolute Scaled Error	129
A.7	Performance comparison between HMC and RWM	130
A.7.1	Computational time	130
A.7.2	Trace plots	131
B	Supplementary Information (Ch. 3)	135
B.1	Synthetic data	135
B.2	Fitting high-fidelity D^{1j}	136
B.2.1	Three-unknown parameterisation (traditional)	136
B.2.2	Four-unknown parameterisation	139
B.2.3	Three-unknown parameterisation (alternative)	143
B.3	Fitting low-fidelity D^{1j}	149
B.3.1	Three-unknown parameterisation (traditional)	149
B.3.2	Four-unknown parameterisation	154
B.3.3	Three-unknown parameterisation (alternative)	156
B.4	Fitting low-fidelity D^{3j}	161
B.4.1	Three-unknown parameterisation (traditional)	161
B.4.2	Four-unknown parameterisation	163
B.4.3	Three-unknown parameterisation (alternative)	165
B.4.4	Three-unknown (Alternative) - Poisson	165
B.5	Fitting high-fidelity D^{3j}	169
B.5.1	Three-unknown parameterisation (traditional)	170
B.5.2	Four-unknown parameterisation	170
B.5.3	Three-unknown parameterisation (alternative)	172
B.6	Sensitivity analysis	172
B.6.1	Scenario 2	174
B.6.2	Scenario 3	177
B.6.3	Scenario 4	182
B.6.4	Scenario 5	184
B.7	Fitting Cumberland's data	187
B.7.1	Incidence report	188
B.7.2	Inference	188
C	Supplementary Information (Ch. 4)	191
C.1	Assumed parameters	191
C.2	Formulation of semi-deterministic candidates	191
C.2.1	Measurement components	191
C.3	Convergence tests	193

Contents

C.3.1	Candidate 1	194
C.3.2	Candidate 2	196
C.3.3	Candidate 3	197
C.3.4	Candidate 4	198
C.3.5	Candidate 5	199
C.3.6	Candidate 6	200
C.3.7	Candidates 7-12	201
C.4	Inference on DGP1's Candidate 2	201
C.4.1	Parameter inference	201
C.4.2	Hidden states	208
C.5	Inference on DGP1's Candidate 4	208
C.5.1	Parameter inference	208
C.5.2	Hidden states	210
C.6	Inference on DGP1's Candidate 5	210
C.6.1	Parameter inference	210
C.6.2	Hidden states	212
C.7	Inference on DGP1's Candidate 6	212
C.7.1	Parameter inference	212
C.7.2	Hidden states	214
C.8	Inference on DGP2	214
C.8.1	Parameter inference	214
C.8.2	Hidden states	217
C.9	Inference on DGP3	217
C.9.1	Prior distributions	220
C.9.2	Posterior distributions	220
C.9.3	Alternative measurement component	223
References		227

List of Figures

2.1	Daily number of influenza cases detected by the United States Public Health Service in Cumberland (Maryland) during the 1918 influenza pandemic, from 22 September 1918 to 30 November 1918.	21
2.2	Adaptation of a Bayesian workflow (Gelman et al., 2020) to calibrate System Dynamics models. Shaded boxes indicate modeller decisions. Dotted lines indicate alternative pathways.	23
2.3	Prior distribution. Shaded areas indicate probability mass. .	25
2.4	Prior predictive checks. A) Simulation of five hundred predicted incidence measurements (grey lines) from the proposed dynamic hypothesis and the prior distribution. Dots denote the actual data. B) Zoom to trajectories that may resemble the actual data. We show predicted measured incidences whose peak is lower than 200 new cases in a day.	27
2.5	A) Early warm-up phase for our three parameters (first 100 iterations). B) Complete sequence of iterations (2000) in the warm-up and sampling phases. The shaded area indicates the warm-up phase. In both figures, we present four Markov chains per calibrated parameter through sequences of points (samples obtained from Stan) joined by lines.	33
2.6	Comparison between marginal prior and marginal posterior distributions for our three parameters. Grey lines denote prior distributions. Blue lines denote posterior distributions	35
2.7	Posterior distribution for our three parameters. The diagonal shows the posterior marginal distributions. In the lower triangular part, the joint posterior distribution of each possible combination of two parameters is displayed. The upper triangular part shows the correlation among parameters. . .	36

List of Figures

2.8	A) 500 predicted incidence measurements (grey lines) from the posterior predictive distribution (model’s fit). To obtain a single predicted measurement, we draw a sample from the posterior distribution and use it to generate a trajectory from the observational model (Eq 2.7). Three different predicted incidence measurements are highlighted in Viridis colours. Dots denote the actual data. B) Posterior predictive distribution described in terms of the mean (dashed line) and 95% credible intervals (contour).	37
2.9	Forecast of scenarios (unmitigated and intervention) under two parameter interaction assumptions. These assumptions correspond to the correlations revealed by the calibration process and perfect independence. We simulated 500 trajectories per experiment in this two-by-two design. For each experiment, we plot a silhouette of the predicted measured incidence. The width of these silhouettes corresponds to 95 % credible intervals. Two silhouettes are superimposed per scenario. . .	39
2.10	A) Potential scale reduction factor by scenario and calibration algorithm for our three parameters. Red dashed line denote the acceptance threshold (1.01). B) Effective sample size (bulk and tail) by scenario and calibration algorithm for our three parameters. Red dashed line denote the acceptance threshold (400) C) Clock time (seconds) per effective sample by scenario and calibration algorithm.	41
3.1	Incidence reports generated by various instances of the SE^iI^jR framework. In this plot, we present two distributions of the latent period and four distributions of the infectious period. The colour of a line corresponds to a particular value of j (infectious period distribution). Solid lines indicate that the incidence report stems from an SEIR model with an exponentially-distributed latent period ($i = 1$). Dashed lines indicate that the incidence report stems from an SEIR model with a gamma-distributed latent period ($i = 3$).	50
3.2	Sample of synthetic data. This plot shows four representative incidence reports (dots) obtained from the simulation of two SE^1I^jR instances (lines). To obtain each report, we sample from the negative binomial distribution.	52

3.3	<p>Inference results obtained from the three-unknown parameterisation. This plot shows the results of fitting model candidates to incidence reports. A) Comparison of estimates for the basic reproduction number obtained from fitting four candidate models to four incidence reports. Error bars correspond to 95% credible intervals, and the vertical line denotes the true value. B) Comparison between inferred incidence (lines) obtained from two candidate models fitted to two incidence reports (dots). Twenty time series represent inferred incidence. Given the high-fidelity data, all inferred incidences are nearly identical, giving the impression of only one line in each panel.</p>	55
3.4	<p>Four-unknown parameterisation. A) Comparison of \mathfrak{R}_0 estimates obtained from four candidates fitted to four incidence reports. Error bars correspond to 95% CrI, and the vertical line denotes the true value. B) Linear relationship between \mathfrak{R}_0 and τ estimated from posterior distributions (represented via samples) obtained from four candidates fitted to four incidence reports. For each sample, we compute the predicted \mathfrak{R}_0 and τ (dots). C) This plot collapses the second column in B into a single panel.</p>	57
3.5	<p>Inference results obtained from the three-unknown alternative parameterisation. This plot compares estimates for the basic reproduction number obtained from fitting four candidate models to four incidence reports. Error bars correspond to 95% credible intervals, and the vertical line denotes the true value.</p>	59
3.6	<p>Comparing estimates from candidates (three-unknown parameterisation) with the wrong and correct latent period distributions. A) Incidence fit from two candidate models matching a D^{33} report. B) Comparison of I_0 and \mathfrak{R}_0 estimates by fitting model. Error bars correspond to 95% CrI, and vertical lines denote true values. C) Posterior distribution from an M^{13} candidate fitting a D^{33} report. The diagonal shows the marginal distributions, the lower triangular part represents each possible pairwise conditional distribution, and the upper triangular part represents the correlations among parameters. C) Posterior distribution from M^{33} candidate fitting a D^{33} report.</p>	62

3.7	Application of the three-unknown alternative parameterisation. This plot shows the estimates obtained from fitting four candidates (M^{1j}) to the daily number of influenza cases (rhombi) detected by the U.S. Public Health Service in Cumberland (Maryland) during the 1918 influenza pandemic, from 22 September 1918 to 30 November 1918. Ribbons correspond to 95% credible intervals of the predicted reported cases by two candidate models. The solid line denotes the median. The box inside the plot shows the estimates for the basic reproduction number by fitting model. Error bars correspond to 95% credible intervals.	67
4.1	Schematic diagram of the data generating processes (DGPs) explored in this chapter. This diagram aims to portray the DGPs as the ensemble of two components: a measurement or observational model (ellipse) and a process model (rounded rectangle). For instance, DGP1 is the amalgamation of the measurement model OM1 and the process model PM1. The process model is in turn the ensemble of two structures: a within-host profile (hexagon) and a time-dependent transmission rate (rhombus). Whereas all process models share a common within-host profile (SEI3R), they differ in the formulation of the transmission rate: Geometric Brownian Motion (GBM), Cox-Ingersoll-Ross (CIR), and n th-order exponential smoothing (NTH-SM). The inference method employed on each DGP depends upon the nature of the process model (Iterated Filtering + Particle Filter for stochastic structures and Hamiltonian Monte Carlo for deterministic ones).	73

List of Figures

<p>4.2 Incidence and mobility data. (A) Daily number (rhombus-shaped points) of COVID-19 cases detected during Ireland’s first wave, from the 29th of February 2020 to the 17th of May 2020. The x-axis indicates the date in which the infected individuals were swabbed. The line represents the smoothed trend (via LOESS method) from the data (B) Weekly number of COVID-19 cases detected in during Ireland’s first wave. The x-axis indicates the number of weeks since the first case was detected. (C) Apple data for Ireland from the 29th of February 2020 to the 17th of May 2020. Points represent the normalised amount of daily requests for driving directions. These indexes are normalised to the value on the 28th of February 2020. We highlight points every 7 days. These highlighted points are used to calibrate DGP1 and DGP2. The line represents the smoothed trend (via LOESS method) from the data. (D) Normalised amount of daily requests for driving directions at the end of each week starting from the 29th of February 2020. These bars correspond to the highlighted points in C.</p>	74
<p>4.3 Brownian motion trajectories. (A) 200 simulations from a transmission rate described by GBM. We generate these simulations from DGP1’s Maximum Likelihood Estimate (MLE) using the Euler-Maruyama (EM) algorithm. (B) 200 simulations from a transmission rate described by the CIR model. We generate these simulations from DGP2’s MLE using the EM algorithm. In both plots, highlighted trends corresponds to the mean trajectory.</p>	78
<p>4.4 Inference on DGP1. Predicted values stem from DGP1’s filtering distribution. In the LHS, the solid line indicates the median, and the darker and lighter ribbons represent the 50% and 95% CI, respectively. (A) Predicted incidence (solid line and ribbons in the LHS; violin plots in the RHS) against actual cases (rhombi in the LHS; horizontal dotted lines in the RHS). (B) Predicted relative transmission rate (solid line and ribbons in the LHS; violin plots in the RHS) compared to mobility index (points in the LHS; horizontal dotted lines in the RHS). (C) Predicted reproduction number (solid line and ribbons in the LHS; violin plots in the RHS). Horizontal dashed lines denote $\mathfrak{R}_t = 1$.</p>	83

List of Figures

4.5	Inference on DGP2. Predicted values stem from DGP2’s filtering distribution. In the LHS, the solid line indicates the median, and the darker and lighter ribbons represent the 50% and 95% CI, respectively. (A) Predicted incidence (solid line and ribbons in the LHS; violin plots in the RHS) against actual cases (rhombi in the LHS; horizontal dotted lines in the RHS). (B) Predicted relative transmission rate (solid line and ribbons in the LHS; violin plots in the RHS) compared to mobility index (points in the LHS; horizontal dotted lines in the RHS). (C) Predicted effective reproduction number (solid line and ribbons in the LHS; violin plots in the RHS). Horizontal dashed lines denote $\mathfrak{R}_t = 1$	86
4.6	Inference on DGP3. Comparison between expected values and data. On the LHS, for each model, we show 100 overlapped simulations of the predicted incidence against daily case counts. On the RHS, for each model, we show 100 overlapped simulations of the predicted relative transmission rate against Apple’s mobility data. In this plot, we estimate the predicted values from the posterior distribution of each of the DGP3’s nine candidate process models.	89
4.7	Comparison of predicted states. These values stem from the filtering distribution (DGP1 and DGP2) and the posterior distribution (DGP3). Here, DGP3’s process model is the 4th-order delay. Further, solid lines indicates the median, and the ribbons represent the 95% CI. (A) Predicted incidence by DGP and weekly cases (rhombi). (B) Predicted relative transmission rates by DGP and mobility data (points). (C) Predicted reproductive numbers by DGP. The horizontal line denotes the epidemics threshold.	93
A.1	Prior predictive checks by measurement component	108
A.2	Toy example data	109
A.3	Parameter space exploration	111
A.4	Synthetic incidence	114
A.5	Case 1. Prior predictive checks	114
A.6	Case 1 fit	116
A.7	Case 2. Prior predictive checks	116
A.8	Case 2.1. Pairs-plot	118
A.9	Case 2.2. Pairs-plot from second attempt	120

List of Figures

A.10 Case 2. Fit	121
A.11 Case 2. Marginal and joint distributions	122
A.12 Case 3. Prior predictive checks	123
A.13 Case 3 fit	124
A.14 Case 3. Marginal and joint distributions	125
A.15 Likelihood function by parameter	126
A.16 Point estimates	127
A.17 Fit comparison by point estimate	128
A.18 MASE distribution	130
A.19 Computational time by MCMC method	131
A.20 Trace plots of the 100-iterations scenario by method	132
A.21 Trace plots of the 200-iterations scenario by method	132
A.22 Trace plots of the 500-iterations scenario by method	133
A.23 Trace plots of the 1000-iterations scenario by method	133
A.24 Trace plots of the 1500-iterations scenario by method	134
A.25 Trace plots of the 1500-iterations scenario by method	134
B.1 Synthetic incidence data by measurement component and delay distribution. There are 80 time series per panel.	135
B.2 Prior distribution	136
B.3 Incidence fit	137
B.4 Fit scores by candidate model and type of data. x denotes latent incidence, whereas y indicates observed incidence. Ver- tical line denotes the mean.	138
B.5 Parameter estimates. Panel headings on the right-hand side indicate the data's origin.	138
B.6 Joint distribution	139
B.7 Score summary	140
B.8 Recovery rate's prior distribution	141
B.9 Incidence fit	141
B.10 Parameter estimates. Panel headings on the right-hand side indicate the data's origin.	142
B.11 Joint distribution	142
B.12 Comparison between the recovery rate's prior distribution (grey histograms) against four marginal posterior distributions obtained from each candidate (coloured histograms).	143
B.13 Scatter plot	145
B.14 Basic reproduction number's prior distribution	145
B.15 Incidence fit	146

List of Figures

B.16 Parameter estimates. Panel headings on the right-hand side indicate the data's origin.	147
B.17 Joint distribution	147
B.18 Overdispersion's prior distribution	149
B.19 Incidence fit	150
B.20 Fit scores by candidate model and type of data. x denotes latent incidence, whereas y indicates observed incidence. Vertical line denotes the mean.	151
B.21 Parameter estimates. Panel headings on the right-hand side indicate the data's origin.	152
B.22 Joint distribution	152
B.23 Score summary	154
B.24 Incidence fit	155
B.25 Parameter estimates. Panel headings on the right-hand side indicate the data's origin.	155
B.26 Joint distribution	156
B.27 Scatter plot	158
B.28 Incidence fit	158
B.29 Parameter estimates. Panel headings on the right-hand side indicate the data's origin.	159
B.30 Joint distribution	159
B.31 Incidence fit	162
B.32 MASE	162
B.33 Parameter estimates.	163
B.34 Joint distribution	165
B.35 Scatter plot	167
B.36 Scatter plot	172
B.37 Scenario 2's simulated incidence reports. Measurement noise from the Poisson (no overdispersion) and Negative Binomial distributions was added to the smooth trajectories obtained from SEIR instances with an exponential-distributed latent period. There are 20 incidence reports overlapped in each panel.	175
B.38 Estimates of the basic reproduction number by model candidate.	177
B.39 Scenario 3's simulated incidence reports. Measurement noise from the Poisson (no overdispersion) and Negative Binomial distributions was added to the smooth trajectories obtained from SEIR instances with an exponential-distributed latent period.	179

List of Figures

B.40 Scenario 4’s simulated incidence reports. Measurement noise from the Poisson (no overdispersion) and Negative Binomial distributions was added to the smooth trajectories obtained from SEIR instances with an exponential-distributed latent period.	182
B.41 Scenario 5’s simulated incidence reports. Measurement noise from the Poisson (no overdispersion) and Negative Binomial distributions was added to the smooth trajectories obtained from SEIR instances with an exponential-distributed latent period.	185
B.42 Cumberland’s incidence data	188
B.43 Posterior predictive checks	189
B.44 \mathcal{R}_0 estimates	190
C.1 Candidate 1’s convergence test	195
C.2 Candidate 2’s convergence test	196
C.3 Candidate 3’s convergence test	197
C.4 Candidate 4’s convergence test	198
C.5 Candidate 5’s convergence test	199
C.6 Candidate 6’s convergence test	200
C.7 Candidates 7-12 convergence test	201
C.8 Local search	202
C.9 Global search	203
C.10 Likelihood estimates	204
C.11 Raw likelihood	205
C.12 Exploration space	206
C.13 MLE estimation while holding ζ constant	206
C.14 ζ confidence intervals	207
C.15 Likelihood surface	207
C.16 95% confidence intervals by method	208
C.17 Candidate 2’s hidden states	209
C.18 95% confidence intervals by method	210
C.19 Candidate 4’s hidden states	211
C.20 95% confidence intervals by method	212
C.21 Candidate 5’s hidden states	213
C.22 95% confidence intervals by method	214
C.23 Candidate 6’s hidden states	215
C.24 95% confidence intervals by method	216
C.25 Likelihood estimates	217

List of Figures

C.26 Analysis of outliers	218
C.27 DGP's hidden states	219
C.28 Prior distributions	220
C.29 Log-likelihood estimates by model candidate	221
C.30 MASE by indicator. Error bars indicate 95% credible intervals.	221
C.31 Parameter estimates. Error bars indicate 95% credible intervals.	222
C.32 Log-lik density by candidate	224
C.33 Candidate 4's trace plot. This candidate has been amal- gamated with the Nbin distribution.	224
C.34 Incidence fit region	225
C.35 Bimodality with one unknown parameter.	226

List of Tables

3.1	Scenarios	64
4.1	Measurement model candidates	80
A.1	Diagnostics	113
A.2	Case 2 estimates	122
A.3	Case 3 estimates	125
A.4	Point estimates	127
A.5	Performance metrics	129
B.1	Traditional parameterisation's coverage table	140
B.2	Four-unknown parameterisation's coverage table	144
B.3	Alternative parameterisation's coverage table	148
B.4	Traditional parameterisation's coverage table	153
B.5	Four-unknown parameterisation's coverage table	157
B.6	Alternative parameterisation's coverage table	160
B.7	Traditional parameterisation's coverage table	164
B.8	Four-unknown parameterisation's coverage table	166
B.9	Alternative parameterisation's coverage table	168
B.10	Alternative parameterisation's coverage table	169
B.11	Traditional parameterisation's coverage table	170
B.12	Four-unknown parameterisation's coverage table	171
B.13	Alternative parameterisation's coverage table	173
B.14	Sensitivity analysis scenarios	174
B.15	Scenario 2's coverage table. Poisson noise	176
B.16	Scenario 2's coverage table. Overdispersion	178
B.17	Scenario 3's coverage table. Poisson noise	180
B.18	Scenario 3's coverage table. Overdispersion	181
B.19	Scenario 4's coverage table. Poisson noise	183

List of Tables

B.20 Scenario 4's coverage table. Overdispersion	184
B.21 Scenario 5's coverage table. Poisson noise	186
B.22 Scenario 5's coverage table. Overdispersion	187
C.1 Parameter values	192
C.2 Semi-deterministic model candidates	192
C.3 Probes	194
C.4 Candidate 2's estimated parameters	202
C.5 Candidate 4's estimated parameters	209
C.6 Candidate 5's estimated parameters	211
C.7 Candidate 6's estimated parameters	213
C.8 DGP2's estimated parameters	215
C.9 Parameter estimates from the well-behaved chains	225

Chapter 1

Introduction

Pandemics are large-scale outbreaks of an infectious disease that lead to increases in morbidity and mortality over a wide geographic area (Morens, Folkers, and Fauci 2009). Within the last 100 years, respiratory viruses have caused five pandemics that killed millions of people (Monto and Webster 2013; Msemburi et al. 2023). In particular, strains of type A Influenza are responsible for four of these seismic events (Monto and Fukuda 2019), whereas a novel coronavirus, SARS-CoV-2 (Huang et al. 2020), is the culprit of the recent COVID-19 pandemic that impacted virtually every country in the world. In light of these threats, decision-makers employ countermeasures to curtail transmission within a population so as to minimise the occurrence of disease, severe illness, and, ultimately, death (Keeling and Rohani 2011). These countermeasures can be pharmaceutical (e.g., vaccination (Daems, Del Giudice, and Rappuoli 2005)) or non-pharmaceutical (e.g., social distancing (World Health Organization 2019)). However, the success of these interventions hinges on an appropriate understanding of the basic epidemiological characteristics of the invading pathogen that allows decision-makers to design effective policies that achieve the intended target while balancing the trade-off between benefits and harms (Hollingsworth et al. 2011; World Health Organization 2019).

1.1 Reproduction number

Undoubtedly, one of the most crucial concepts that mathematical thinking has contributed to understanding infectious disease dynamics is the epidemi-

1.1. Reproduction number

ological metric known as the reproduction number or ratio (Heesterbeek and Dietz 1996). This concept was originally developed in the field of demographics (Dublin and Lotka 1925; Kuczynski 1928) to characterise the growth and decline of populations, and it is defined as the expected future offspring of a new-born individual (Heesterbeek and Dietz 1996). If this value is above one, the population will grow; otherwise, it will decline. In parallel, Ronald Ross is credited (Heesterbeek and Dietz 1996; Heesterbeek 2002) as the first scholar to employ rudimentarily the reproduction number in epidemiology. Specifically, he identified a critical mosquito density below which malaria transmission cannot be maintained in the population (Ross 1910). A decade later, Kermack and McKendrick further advanced this work (Kermack and McKendrick 1927) by postulating a theorem which states that, in order for an infectious agent to be sustained in a population, the population density has to exceed a certain critical density (Heesterbeek 2002). Then, in 1952, George Macdonald (Macdonald 1952) linked the dynamics of infectious diseases to the critical threshold of unity and set the foundations for developing its current definition (Delamater et al. 2019). However, we had to wait for another few decades to develop the concept and fully comprehend its potency (Heesterbeek 2002), thanks to the efforts of Klaus Dietz, Roy Anderson and Robert May.

Assuming that the population is completely susceptible, one refers to the reproduction number as the *basic* reproduction number or ratio (\mathfrak{R}_0), and it corresponds to the expected number of secondary cases produced by a typical infected individual during its entire period of infectiousness (Diekmann, Heesterbeek, and Metz 1990; R. Anderson and May 1992; Fine 1993; Heffernan, Smith, and Wahl 2005). Epidemiologists employ this quantity to determine whether the pathogen can invade the population (Heffernan, Smith, and Wahl 2005) by way of its threshold property (Heesterbeek 2002). If an infection is to persist, each infected individual must, on average, transmit that infection to more than one other individual ($\mathfrak{R}_0 > 1$); otherwise, the infection will disappear progressively from the population (Fine 1993). Furthermore, \mathfrak{R}_0 serves as an indicator of a pathogen's transmissibility, given that its magnitude is positively correlated to the predicted number of infected individuals at the end of an outbreak (Ma and Earn 2006; Delamater et al. 2019). This feature implies that \mathfrak{R}_0 also gauges the effort needed to control the spread of a disease (Heesterbeek and Dietz 1996; Dietz 1993). On the other hand, decision-makers can use \mathfrak{R}_0 to calculate the level of vaccination required to eradicate an infection from a population (Keeling and Rohani

2011).

Although \mathfrak{R}_0 provides crucial insights about the epidemiological characteristics of a disease, once an infection spreads through a population, it is convenient to complement the analysis with a metric that forgoes the premise of a totally susceptible population (Wallinga and Lipsitch 2007; Delamater et al. 2019) and other static assumptions. This quantity corresponds to the *actual* or *effective* reproduction number, denoted by \mathfrak{R}_t , and it is defined as the *time-varying* average number of secondary cases caused by a primary case at a calendar time t (R. Anderson and May 1992). Given its definition, one can interpret \mathfrak{R}_t as a theoretical indicator of how policy changes, population immunity, and other factors have affected transmission at specific points in time (Gostic et al. 2020). Interestingly, \mathfrak{R}_t generalises \mathfrak{R}_0 (Eq (1.1)) by taking into consideration the susceptible fraction of the population (s_t), thereby extending the threshold property at any given time t . If \mathfrak{R}_t is and remains below unity (epidemiological threshold), the outbreak will die out. On the contrary, if $\mathfrak{R}_t > 1$, a sustained outbreak is likely (Thompson et al. 2019). This property entails that a decision-maker’s ultimate purpose concerns reducing \mathfrak{R}_t to a level below the epidemiological threshold to control the disease. Moreover, its estimation helps assess whether current control efforts are effective or additional interventions are required (Thompson et al. 2019).

$$\mathfrak{R}_t = \mathfrak{R}_0 s_t \tag{1.1}$$

Consequently, accurate estimation of the reproduction number is crucial to planning for the control of an infection (Dietz 1993; Wallinga and Lipsitch 2007). However, it should be noted that \mathfrak{R}_t summarises in a single scalar numerous biological, sociobehavioural, and environmental factors that govern pathogen transmission (Delamater et al. 2019) for which direct observation is not possible. Although in specific contexts, such as certain vector-borne infections and sexually transmitted diseases (Dietz 1993), it is possible to measure those factors directly, it is more likely that the estimation will rely on surrogate measurements. On the one hand, if the inference process concerns the basic reproduction number (\mathfrak{R}_0) exclusively, there exist methods that leverage various sources of information, such as the number of cases at the start of the outbreak and seroprevalence data, to obtain plausible approximations if certain conditions are met. Readers are referred to Dietz (1993) and Heffernan, Smith, and Wahl (2005) for details of these methods.

On the other hand, if the quantity of interest is the actual reproduction number (\mathcal{R}_t), where $t > 0$, another toolset is available, specifically tailored to perform this task. This toolset stems from an inductive approach (Coulson, Rohani, and Pascual 2004) that does not aim to capture the underlying process governing the system behaviour but instead explicitly uses the patterns in the time-series data to construct and parameterise the predictive models. In this case, using statistical inference methods, one obtains near real-time \mathcal{R}_t estimates from models fitted to periodic (daily or weekly) counts of confirmed cases and reported deaths (Thompson et al. 2019; Abbott et al. 2020; Gostic et al. 2020; Ogi-Gittins et al. 2023). This type of model accounts for the delays between infection and the eventual reporting of a case or death.

1.2 Compartmental models

Alternatively, one can employ compartmental models to estimate the reproduction number (basic and actual). These models describe systems in terms of a finite number of macroscopic subsystems, called compartments or pools, each of which is homogeneous and well mixed, and the compartments interact by exchanging materials. There may be inputs from the environment into one or more of the compartments, and there may be outputs from one or more compartments into the environment (Jacquez 1972). Furthermore, compartmental models are deductive in nature (Coulson, Rohani, and Pascual 2004), given that a scientific understanding of the system guides their formulation process (Bretó et al. 2009). In epidemiology, modellers subdivide the population into a number of discrete compartments according to their infection status and describe via equations the transition of individuals between compartments (Keeling and Rohani 2011). Moreover, transition equations can be stochastic or deterministic. The former is preferred for addressing questions involving small populations and analysing the persistence of infections (Vynnycky and White 2010; Keeling and Grenfell 2002). This type of model employs Continuous Time Markov Chains or stochastic differential equations (Allen 2017) to introduce randomness into the system. Stochastic models, nonetheless, can also be used for studying large populations (Camacho et al. 2014).

On the other hand, deterministic structures (R. Anderson and May 1992) formulated via ordinary differential equations (ODEs) may enable sufficient biological understanding of the system, provided that population numbers

never become too small (Renshaw 1993). Namely, ODE models are appropriate for relatively large populations (Heesterbeek and Dietz 1996), where the average behaviour of the system can be described with high accuracy (Keeling and Rohani 2011). Undoubtedly, the main advantage of deterministic models is their mathematical convenience to perform various kinds of studies, such as stability (Krylova and Earn 2013; Strogatz 2018), sensitivity (Epstein, Hatna, and Crodelle 2021), and simulation analyses (Gostic et al. 2020; Park and Bolker 2020), that facilitate the identification and interpretation of cause-and-effect relationships. Likewise, these structures are amenable to well-known inference methods such as non-linear optimisation techniques (Brauer and Castillo-Chavez 2012) and Markov chain Monte Carlo (Chatzilena et al. 2019). Since this work focuses on pandemics caused by respiratory viruses, we exclusively consider deterministic transition equations. Of note, the insights derived from deterministic models can serve as the cornerstone for formulating more elaborated extensions that incorporate stochastic equations.

When modelling the transmission of directly transmitted infections with deterministic models, it is commonplace to employ the Susceptible-Infected-Recovered (SIR) framework (Keeling and Rohani 2011), which dates back to the early 20th century (Ross and Hudson 1917; Kermack and McKendrick 1927). This framework considers the introduction of one or more infectious persons into a closed population (no birth nor migration flows) of susceptible individuals who acquire infection through contact with infectious individuals. Each infected individual runs through the course of his or her infection until he or she recovers or dies (R. Anderson 1991). Kermack and McKendrick, in their seminal work (Kermack and McKendrick 1927), considered the general case in which the chance of removal (recovering or dying) increases over time (Lloyd 2009). Furthermore, they assumed that population size would only decline due to disease-induced mortality and that recovery conferred lasting immunity (R. Anderson 1991; Ross and Hudson 1917). Simply put, this framework stratifies individuals into three classes: susceptible (S_t), infected (I_t), and recovered (R_t).

If we further assume that the recovery rate is independent of the time since infection and that there is no disease-induced mortality, one obtains the classical SIR model (Eq (1.2)). This structure represents the interaction between susceptible and infectious individuals following the mass action principle ($\frac{\beta S_t I_t}{N}$), an idea borrowed from chemical reaction kinetics. In short, this principle states that the number of contacts between susceptible and

infectious individuals per unit of time is proportional to the product of the respective densities (Heesterbeek and Dietz 1996). In other words, the number of contacts is independent of the population size. This property is also known as the frequency-dependent transmission (Keeling and Rohani 2011). In this formulation, β corresponds to the transmission coefficient, the average number of contagions per infectious individual per unit time, which is given by the product of the contact rate and the transmission probability (Lloyd 2009). Finally, the rate at which individuals transition from the infectious to the recovered compartment is denoted by γ . Its inverse (γ^{-1}) characterises the average number of days an individual can transmit the disease, namely the mean infectious period.

$$\begin{aligned}\dot{S}_t &= -\frac{\beta S_t I_t}{N} \\ \dot{I}_t &= \frac{\beta S_t I_t}{N} - \gamma I_t \\ \dot{R}_t &= \gamma I_t\end{aligned}\tag{1.2}$$

Defining the SIR model leads us to the basic reproduction number (Eq (1.3)). Recall that \mathfrak{R}_0 was previously defined as the expected number of secondary cases (β) produced by a typical infected individual during its entire period of infectiousness (γ^{-1}). Analytically, modellers deduce this equation using the next-generation matrix method (Diekmann, Heesterbeek, and Metz 1990). Nevertheless, \mathfrak{R}_0 is not merely a by-product of the model. It is the quantity that governs its dynamics. Consider the case where $S_0 = N$, $I_0 = 0$, and $R_0 = 0$. Stability analysis (van den Driessche 2017) predicts the existence of a disease-free equilibrium (DFE) at this coordinate. If $\mathfrak{R}_0 < 1$, the DFE is locally asymptotically stable. That is, the system tends to return to the equilibrium state or *fixed point* after experiencing sufficiently small disturbances (Strogatz 2018). In practical terms, this scenario implies that introducing infectious individuals into the system will not lead to an outbreak, and the number of infectious individuals will decrease monotonically to 0. Conversely, when $\mathfrak{R}_0 > 1$, the DFE is unstable, which entails that any small perturbation will grow in time, moving the system away from the fixed point. Namely, the introduction of infectious individuals will propagate the disease throughout the population. Therefore, \mathfrak{R}_0 acts as a sharp threshold between the disease dying out or causing an outbreak (van den Driessche 2017). At least in the idealised mathematical world. In reality, what seems far more likely is that the probability of an outbreak occurring rises substantially near

the epidemiological threshold (Renshaw 1993).

$$\mathfrak{R}_0 = \beta\gamma^{-1} \tag{1.3}$$

Although the classical SIR model encompasses sufficient realism to be useful in some contexts (Renshaw 1993), its appeal stems from the flexibility that the framework offers to incorporate more realistic features. For instance, the SIR allows the introduction of additional compartments that reflect more accurately the natural history of the disease, such as latent (Wearing, Rohani, and Keeling 2005), pre-clinical (Davies et al. 2020), and sub-clinical states (Chowell et al. 2006). One can also incorporate demographic effects (Krylova and Earn 2013), age-specific variation in contact rates (Andrade and Duggan 2020), loss of immunity (Duggan et al. 2024), and spatial coupling (Bolker and Grenfell 1995; Keeling and Rohani 2002). Additionally, these enhanced models can be further enriched by including structures that account for control strategies such as vaccination (Bubar et al. 2021) or testing, tracing, and isolation (Sturniolo et al. 2021). Irrespective of the particular formulation, \mathfrak{R}_0 can be formulated as a function of the model parameters using the next-generation matrix method (Diekmann, Heesterbeek, and Metz 1990; van den Driessche 2017). Then, one estimates \mathfrak{R}_t at any given t via simulation. Of course, more complex models entail more involved calculations, but in principle, the reproduction number can be computed from parameter values and simulation.

1.3 Inference

In practice, though, modellers may only have access to a subset of those values, while the remaining (unknown) parameters must be estimated through a particular application of statistical inference. This procedure, known as *model calibration* (Oliva 2003) or *trajectory matching* (Wearing, Rohani, and Keeling 2005), consists of fitting the ODE model (Eq (1.4)) to available data (\vec{y}), usually a time series of case numbers (Vynnycky and Edmunds 2008), hospitalisations (Chowell et al. 2006), or deaths (Diaz et al. 2018). Here, Eq (1.4) corresponds to any SIR-like system of differential equations, where \vec{x} indicates the vector of compartments and $\vec{\theta}$ the unknown parameters.

$$\frac{d\vec{x}}{dt} = \vec{f}(\vec{x}, t, \vec{\theta}) \tag{1.4}$$

1.3.1 Frequentist approach

One approach to performing model calibration involves the formulation of a probability density function, $\pi(\vec{y}|\vec{f}(\vec{\theta}))$, that expresses how well the observed data is compatible with a given model (Dangerfield and Duggan 2020). Since the model and the observed data are fixed in this formulation, the probability density function only depends on the unknown parameters. We refer to this function as the likelihood: $\mathcal{L}(\theta)$ (Eq (1.5)). From this point onward, we omit the vector representation for notational convenience. Once the likelihood function has been defined, modellers run optimisation routines to explore its parameter space in order to identify the set of values that best describes the available information (Dangerfield and Duggan 2020). Namely, the Maximum Likelihood Estimate (MLE): $\hat{\theta}_{ML}$. Algorithms such as the Nelder-Mead simplex method (Nelder and Mead 1965) offered by R (R Core Team 2021) or Powell’s conjugate direction method (Powell 1964) included in System Dynamics software (e.g., Vensim or Stella) achieve this task.

$$\mathcal{L}(\theta) = \pi(y|f(\theta)) \tag{1.5}$$

If $\hat{\theta}_{ML}$ produces simulated behaviour that resembles the observed one, it is fair to conclude that *the model fits the data*. However, this statement is about the model, not about the MLE itself. Naturally, one wonders about the *quality* of this estimate. That is, which values near $\hat{\theta}_{ML}$ produce simulations consistent with the data? In other words, we are interested in the uncertainty surrounding $\hat{\theta}_{ML}$. Generally, one approaches these questions through the estimation of uncertainty regions. From the *frequentist* tradition, these regions are known as *confidence intervals* (CI). Box and colleagues (Box, Hunter, and Hunter 2005) define a CI in terms of hypothesis testing, where the sample estimate corresponds to the null hypothesis ($H_0 : \hat{\theta}_{ML}$) and α the significance level. Then, a $1 - \alpha$ CI for H_0 , using a two-sided significance test, implies that values within this neighbourhood do not produce significant discrepancy with the data at the chosen significance level. Conversely, all the values outside the interval do show significant discrepancies. Additionally, a CI has the theoretical property that in *repeated sampling from the same population*, a proportion $1 - \alpha$ of intervals constructed in this fashion contains the true value (Box, Hunter, and Hunter 2005). We refer to this property as *coverage*.

Nevertheless, estimating CIs from ODE models is far from straightforward.

One approach concerns the computation of the standard deviation of a parameter estimate. Namely, the standard error (SE) (Gelman 2023). Leveraging a classical asymptotic result in the spirit of a central limit theorem permits the approximation of the SE. This result establishes that given a sufficient sample size, the MLE stems from a multivariate normal sampling distribution (Brauer and Castillo-Chavez 2012): $\hat{\theta}_{ML} \sim N(\theta, \Sigma)$. Here, θ denotes the true value and Σ the covariance matrix, a square matrix whose size is equal to the number of estimated parameters. Importantly, the square root of its diagonal corresponds to each parameter's standard error. Then, a CI is defined as follows: $\hat{\theta}_{ML} \pm z \times SE$, where z represents the number of standard deviations above or below the mean of a standard normal distribution.

This covariance matrix can be approximated using the second derivative (Hessian) of the log-likelihood evaluated at the MLE. This Hessian, also known as the observed Fisher information matrix, measures the curvature around the MLE (Pawitan 2013) and can be interpreted as the information content of the data corresponding to the model parameters (Banks et al. 2009). More importantly, the inverse of the Fisher information matrix is an estimator of the asymptotic covariance matrix (Σ). Optimisation routines, such as *optim* in R, return a numerical approximation of the Hessian evaluated at the MLE. However, the validity of these estimates rests upon certain regularity conditions (Banks et al. 2009) and a sufficient sample size (Cintr3n-Arias et al. 2020). As regards these requirements, software packages do not verify the former, whereas the latter is rarely met in practice. For instance, an incidence time series of an outbreak corresponds to only one measurement.

Instead of relying on the standard error for constructing confidence regions, an alternative approach that is likely to be more robust (better coverage) in small samples is related to the normalised likelihood or likelihood ratio (Royston 2007). Following this approach, a confidence interval for a single parameter (θ) is defined as a set of parameter values with a high enough likelihood such that $\frac{\mathcal{L}(\theta)}{\mathcal{L}(\hat{\theta}_{ML})} > c$ (Pawitan 2013). Notice that the threshold or cut-off value c is a value between 0 and 1. However, determining the specific value at which one rejects the null hypothesis requires asymptotic theory. If the sample size is sufficiently large, $2 \log \frac{\mathcal{L}(\hat{\theta}_{ML})}{\mathcal{L}(\theta)} \sim \chi_{1,1-\alpha}^2$, which implies that $c = e^{0.5\chi_{1,1-\alpha}^2}$. See Pawitan (2013) for the complete details. After some algebraic manipulation, a $1 - \alpha$ confidence interval is defined as the values of θ that satisfy the inequality Eq (1.6).

$$\log \mathcal{L}(\theta) > \log \mathcal{L}(\hat{\theta}_{ML}) - 0.5\chi_{1,1-\alpha}^2 \quad (1.6)$$

While Eq (1.6) is defined for a single parameter, in actual applications, θ consists of various parameters. In consequence, each parameter must be addressed individually. To explain this procedure, we decompose θ into two components: ψ and ϕ . Here, ψ is a scalar that denotes the parameter of interest, whereas ϕ is a vector of one or more elements representing the remaining parameters, treated as a *nuisance* (Pawitan 2013). Then, one fixes ϕ to certain values and varies ψ to construct the confidence interval at the desired significance level. For instance, System Dynamics software employs the *likelihood slice* approach (Dogan 2007), which is a method that fixes ϕ at the MLE and then performs a grid search over ψ to identify values that satisfy Eq (1.6). Although this shortcut produces quick results, Bolker (Bolker 2008) notes that: “A simple, but usually wrong, way of [constructing confidence intervals] is to calculate a likelihood slice, fixing the values of all but one parameter (usually at their maximum likelihood estimates) and then calculating the likelihood for a range of values of the focal parameter”. The reason for this assessment is that confidence regions from likelihood slices are accurate only if the elements of θ are uncorrelated; otherwise, they are statistically misleading (Bolker 2008).

A more robust approach to eliminate the nuisance parameter (ϕ), albeit more computationally intensive, is to replace it with its MLE at each fixed value of the parameter of interest. This method is called the *profile likelihood* (Pawitan 2013). In practice, one specifies a number of points within a plausible range for the parameter of interest. For each point in this range, we optimise the likelihood with respect to the nuisance parameter, yielding the vector θ_{prof} . Subsequently, we use Eq (1.6) to compute the confidence region at the desired confidence level, where θ is replaced by θ_{prof} .

Another technique for computing confidence intervals is the *Bootstrap* method (Efron and Tibshirani 1986), which uses resampling to approximate the MLE’s sampling distribution. This resampling can be parametric or non-parametric (Dogan 2007). In the context of compartmental models, the parametric variant is favoured (Chowell, Nishiura, and Bettencourt 2007), given that modellers have access to only one time series. This constraint imposes an additional assumption to generate measurement error at each time t if the modeller wants to avoid the assumption of equal variance across measurements. Following this approach, first, one estimates $\hat{\theta}_{ML}$ using an

optimisation routine. Then, $\hat{\theta}_{ML}$ is plugged into $\pi(y|\theta)$ to fabricate new data sets. Here, we turn $\pi(y|\theta)$ into a sampling distribution by fixing θ to $\hat{\theta}_{ML}$. Finally, the simulation model is fitted, using the optimisation routine, to each of these new sets, generating a distribution of parameter estimates. This empirical distribution of parameters allows the estimation of confidence intervals (Dogan 2007).

1.3.2 Bayesian approach

A competing paradigm ushered in by Bayes and Laplace stems from the revolutionary shift in the conception of parameters as unknowns to the notion of parameters as random variables (Robert 2007). The key contribution of this idea is that one can formulate a probability distribution ($\pi(\theta)$) on the model parameters to quantify their uncertainty. We refer to this distribution as the prior distribution. Moreover, a prior distribution can be updated based on observations y that contain information about θ . In particular, the likelihood function ($\pi(y|\theta)$) is responsible for summarising this information. Here, we omit the ODE model (f) in the notation of the likelihood function to maintain consistency with mainstream texts. As a result, we obtain an updated device known as the posterior distribution ($\pi(\theta|y)$) (Gelman et al. 2013). In statistical terms (Eq (1.7)), the posterior distribution arises from the Bayes theorem (Robert 2007), where $\pi(y) = \int \pi(\theta)\pi(y|\theta)d\theta$ denotes the average likelihood (Gelman et al. 2013), a constant that ensures that $\pi(\theta|y)$ integrates to one (McElreath 2020).

$$\pi(\theta|y) = \frac{\pi(\theta)\pi(y|\theta)}{\pi(y)} \tag{1.7}$$

In a nutshell, the technical core of Bayesian inference involves the formulation of a model that comprises a likelihood function and a prior distribution and then performing the appropriate computation to obtain the posterior distribution (Gelman et al. 2013). Since $\pi(\theta|y)$ encapsulates information from all parts of the Bayesian model, the answer to any statistical question should involve the manipulation of this mathematical device. As regards the uncertainty of parameter estimates, marginalising $\pi(\theta|y)$ over ϕ results in a univariate distribution of the parameter of interest (ψ). From this marginal distribution, we can estimate uncertainty regions known as *credible intervals* (CrI). In this context, a $1 - \alpha$ CrI implies that a proportion $1 - \alpha$ of the values compatible with the model and the data lies in the parameter range specified

by the interval bounds (Lambert 2018; McElreath 2020). Furthermore, one can marginalise $\pi(\theta|y)$ over a subset of ϕ to obtain multivariate distributions that allow the identification of parameter interactions with ψ . Undoubtedly, the main advantage of a Bayesian approach is that the range of inference questions extends beyond θ . Drawing on the concept of expectations (Eq (1.8)), $\pi(\theta|y)$ can answer any question properly posed in terms of a function (g) that depends on θ . For instance, $g(\theta)$ can be the prediction of an unobserved state in a SIR-like model, a performance metric (e.g., mean absolute scaled error), or even the SIR-like model simulated beyond the period for which it was fitted (forecast). In all of these cases, answers are always given in terms of probability statements (Gelman et al. 2013), which are valid for any sample size (McElreath 2020).

$$\mathbb{E}_\pi(g) = \int g(\theta)\pi(\theta|y)d\theta \quad (1.8)$$

However, the computation of $\pi(\theta|y)$ for non-trivial models requires the use of simulation-based approaches such as Markov chain Monte Carlo or MCMC (Geyer 2011). In short, this procedure consists of generating a sequence of random variables $\theta_1, \theta_2, \dots$, where the probability distribution of θ_n only depends on the previous one: θ_{n-1} . In other words, the Markov property (Blitzstein and Hwang 2019). Also, the conditional probability distribution for generating random variables, $T(\theta_n|\theta_{n-1})$, is called a transition kernel or Markov kernel (Robert and Casella 2010) that, if appropriately crafted, will induce a *stationary* or *invariant* distribution. This property entails that every new random variable (θ_n) will be identically distributed to its previous state (θ_{n-1}). Furthermore, if one makes $\pi(\theta|y)$ the target of the invariant T , then generating new θ_n from T would be equivalent to producing samples from the posterior distribution, and from there, the computation of expectations or uncertainty regions is straightforward (Betancourt 2018).

Early work on this problem dates back to the development of the Metropolis-Hastings algorithm (Metropolis et al. 1953), which experienced further refinements over time (Betancourt 2017). Essentially, these algorithms generate new samples that are improved at each step in the simulation in the sense of converging to the target distribution (Gelman et al. 2013). Convergence, in this context, means that the Markov chain becomes stationary. However, only asymptotics (infinite samples) guarantee convergence, which implies that an infeasible number of samples may be required to reach this stationary state when dealing with complex parameter spaces (Betancourt 2018). This

fragility stems from the random walk behaviour embedded in their transition kernel (Neal 1995), where the Markov chain can take a long time zigging and zagging while exploring the target distribution (Gelman et al. 2013).

At this point, asymptotics (whether it is based on a large sample size or a large Markov chain) is the only basis for ensuring the validity of inference results. In contrast, the interplay between statistics and physics (Betancourt 2017) produced an enhanced transition kernel based on Hamiltonian dynamics that all but suppresses the random behaviour in the generation of random variables. This method, known as *Hamiltonian Monte Carlo* or HMC (Duane et al. 1987; Neal 2011), not only allows the Markov chain to move much more efficiently through the target distribution (Andrade and Duggan 2020) but also provides information during the generation of random variables, indicating whether they should be rejected (Betancourt 2018). Although HMC is proficient in traversing parameter spaces, it cannot provide reliable estimates for every model taxonomy. However, the next best job that an algorithm can do is inform the modeller whether it failed or not so that she or he can perform the appropriate improvements. This feature distinguishes HMC and what makes it a powerful instrument for statistical inference.

1.4 Research questions

Therefore, the immediate question that arises is *how can one employ Hamiltonian Monte Carlo to obtain robust estimates from SIR-like models fitted to time-series data?* Chapter 2 addresses this research question by showing how to estimate \mathfrak{R}_0 from an SEIR model (extends the SIR by including an *Exposed* class) fitted to an incidence report from the 1918 flu pandemic. This chapter serves two purposes. On the one hand, it allows us to explain the methodology that guides the work in subsequent chapters. Even though HMC is a powerful algorithm, it must be complemented with the appropriate workflow and visualisation tools to harness its full potential. As we will see in Chapter 3, a visualisation tool such as a scatter plot can provide valuable insights into a phenomenon. On the other hand, Chapter 2 introduces the algorithm to System Dynamics practitioners, where the likelihood slice method is a ubiquitous practice within this field, given that their proprietary software of choice provides results in a matter of seconds and after a few clicks.

Nevertheless, inferences are conditional on the model assumed (McElreath

2020), and no algorithm, however good it may be, can redress model misspecification. That is, inference methods cannot compensate for the discrepancy between the actual data-generating process and the assumed model (James et al. 2021). In fact, a critical limitation of compartmental models is that structures with different embedded assumptions can fit the data equally well, but they produce dissimilar \mathcal{R}_0 estimates (Keeling and Rohani 2011; Gostic et al. 2020). For instance, Park and Bolker (Park and Bolker 2020) showed that misspecifying the structure that accounts for the time at which infections are reported can bias \mathcal{R}_0 estimates and lead to overly narrow confidence intervals. Similarly, the work of Wearing and colleagues (Wearing, Rohani, and Keeling 2005) indicates that ignoring the latent period or formulating exponentially-distributed latent and infectious periods results in inaccurate \mathcal{R}_0 estimates. This result prompts us to investigate *“how can we mitigate biases in \mathcal{R}_0 estimates due to uncertainty in the distribution of the epidemiological delays?”* Consequently, Chapter 3 describes a simulation study that aims to identify a parameterisation of the SEIR model that produces accurate estimates in view of that uncertainty. Once we find this parameterisation, we update the estimate presented in Chapter 2.

Likewise, if the assumption of a constant transmission rate (β in Eq (1.2)) cannot be justified, it is crucial to identify a mathematical formulation that accounts for changes in contact patterns. However, the specifics of that formulation will play a prominent role in the inference of the reproduction number. Broadly speaking, there are two routes: parametric and non-parametric. In this context, parametric means that the transmission rate follows a specific shape from a single configuration. These formulations are relatively easy to fit because the inference process is reduced to the estimation of only a small number of coefficients (James et al. 2021), the type of model where HMC shines. These methods, though, make strong assumptions about the underlying phenomenon, which implies that if the specified functional form is far from the truth, the inference process can produce misleading estimates. In contrast, non-parametric formulations do not explicitly assume parametric forms, which implies that the transmission rate can take infinite shapes from a single configuration and thereby provide an alternative and more flexible approach. Nevertheless, estimates can be too sensitive to the particular data provided, a phenomenon referred to as high variance (James et al. 2021). In this context, variance entails that one may achieve a good match between actual and simulated behaviour at the expense of unrealistic corrections to certain elements in the model. Furthermore,

non-parametric formulations require a different set of more computationally intensive algorithms, such as Sequential Monte Carlo and Iterated Filtering.

We faced this conundrum during Ireland’s first COVID-19 wave, which led us to ask: *How can we improve the estimation of the reproduction number from compartmental models when the dynamics of the transmission rate are unknown?* Chapter 4 describes the process to answer this question. In short, we leverage existing approaches to propose three complementary formulations of the transmission rate. We amalgamate these formulations with an SIR-like model. These models incorporate a measurement component that accounts for reported cases, and mobility data is assumed as a proxy of the transmission rate. By following this complementary approach, we assess the trade-offs associated with each formulation and reflect on the benefits/risks of incorporating proxy data into the inference process.

Finally, Chapter 5 concludes the thesis with a summary of the main contributions, as well as a discussion of the limitations of this work and an outline of potential directions for future research.

1.4. Research questions

Chapter 2

Hamiltonian Monte Carlo in a Bayesian workflow

The work presented in this chapter was published in:

Andrade, J., Duggan, J., 2021. A Bayesian approach to calibrate system dynamics models using Hamiltonian Monte Carlo. *System Dynamics Review*, 37(4).

2.1 Introduction

From its beginnings in the mid-1950s to its modern practice, System Dynamics (SD) has been a purpose-driven approach. Namely, it is a field interested in problems to be solved, situations that need to be better understood, or undesirable behaviours that need to be corrected or avoided ([Forrester 1993](#)). To meet these goals, SD practitioners develop simulation models, formal representations often via Ordinary Differential Equations (ODE) that capture the dynamic complexity of the problem situation and from which behavioral inferences can be made ([Saleh et al. 2010](#)). The validity of these inferences hinges on the ability of the model's internal structure to adequately represent the aspects of the system that are relevant to the problem behaviour ([Barlas 1996](#)). Adequacy, nevertheless, is not merely related to the appropriateness of the equations. In order for a model to be useful, it must provide an assessment of future behavior ([Duggan 2016](#)). That is, models ought to reasonably estimate the likely impact of interventions in a system, a process

that cannot be achieved without a plausible quantification of the model’s parameters. Given the continuous nature of ODE representations, there are infinite possibilities for parameter quantification. To address this uncertainty, analysts fit models to available data in order to obtain estimates for the unknown quantities. This procedure, referred to as *model calibration*, serves a dual purpose. In addition to reducing uncertainty around the parameters, model calibration is a stringent validity test that assesses the causal claim that a particular structure accounts for an observed behaviour (Oliva 2003). It thus follows that one should reject models that fail this test.

Generally speaking, model calibration is the process of finding a match between observed and simulated behaviours via *statistical inference* (Oliva 2003). In other words, we search for plausible parameter values or *model configurations* that *accurately* account for the available data. Traditionally, within the SD field, practitioners have followed a *frequentist* approach. Following this paradigm, one employs non-linear optimisation algorithms to maximise a statistical function (often a likelihood function), which expresses how well the model fits a time series of data pertaining to an important model variable (Dangerfield and Duggan 2020). However, such optimisation routines can be inefficient for finding a match in non-trivial and high-dimensional parameter spaces (Andrade and Duggan 2020). To deal with this difficulty, SD practitioners adopt the strategy of running the optimisation algorithm from multiple starts and select the result with the largest likelihood. Unfortunately, as the number of parameters increases, so does the risk of exhausting computational resources before finding the optimal start. To further complicate matters, the Maximum Likelihood Estimate (MLE) may not even be located in regions of high probability mass in high-dimensional spaces (Betancourt 2018).

Furthermore, around the MLE, one can construct uncertainty bounds using frequentist approaches such as the likelihood ratio method (Pawitan 2013), a technique offered by SD software (*Vensim* and *Stella*). ODE models, nevertheless, often violate the assumptions implicit in the likelihood ratio method, such as identically and independently distributed (IID) normal error terms (Dogan 2007). Fortunately, the advent of powerful computational resources has been a catalyst that enabled the development of methods based on repeated random sampling to obtain numerical results (Robert and Casella 2010). These statistical simulation algorithms can be oriented to explore complex parameter spaces and lift stringent restrictions on the shape of the uncertainty bounds. Although the SD community has not ignored

these advances, a search on the SD literature suggests that the adoption of these methods has been gradual. Particularly, Dogan (2007) and Struben, Sterman, and Keith (2015) propose bootstrapping as a robust frequentist method for confidence interval estimation and demonstrate its application on a service quality model (Oliva and Sterman 2001). In essence, this method creates *new* data sets by resampling the original data, and then parameter values are estimated from each of these *new* bootstrap samples (Dogan 2007). Similarly, Ansah et al. (2017) use bootstrapping to quantify the uncertainty in the parameter estimates of a model that predicts the number of Chinese elderly with some degree of cognitive impairment by 2060.

On the other hand, Pierson and Sterman (2013) report the first use of a Markov chain Monte Carlo (MCMC) algorithm to perform inference on a System Dynamics model: “*We estimate model parameters by maximum likelihood methods during both partial model tests and full model estimation using Markov chain Monte Carlo methods to establish confidence intervals*”. Specifically, these authors estimated uncertainty bounds for twenty-one parameters of an industry-level model of airline profits. Likewise, Keith, Sterman, and Struben (2017) estimated the parameters of seven alternative models that account for product diffusion in the hybrid electric vehicle market. In these two case studies, the Markov chains are started from a point estimate obtained from non-linear optimisation routines. More recently, Ghaffarzadegan and Rahmandad (2020) inferred the value of a nine-parameter epidemiological model that describes the early infectious process of COVID-19 in Iran. All of these authors employed enhanced versions (Vrugt et al. 2009; Osgood and Liu 2015) of the Metropolis algorithm (Metropolis et al. 1953). Osgood and Liu (2015) provides a technical overview of the method, accompanied by a practical example.

Despite the benefits that statistical simulation offers for parameter inference, bottom-up implementations require from the practitioner a new mathematical and programming skillset. Therefore, a more viable strategy is the use of predefined routines provided by statistical packages. Even though these tools automate the process, their use requires the practitioner to understand what the method is trying to solve, why it works, and when and why it fails. However, the literature of parameter inference on ODE models via statistical simulation is sparse, and the notation can be challenging for practitioners with non-mathematical backgrounds, which impedes adoption within the SD community. This observation serves as the motivation for writing this chapter. Thus, the contribution of this work is two-fold. First, we introduce

to the SD field a state-of-the-art MCMC algorithm, known as Hamiltonian Monte Carlo (Neal 2011) or HMC, oriented to explore non-trivial parameter spaces such as those common to SD models. As model size and complexity grow, this method outperforms other MCMC implementations (Monnahan, Thorson, and Branch 2017; Beraha, Falco, and Guglielmi 2021) and, in some instances, non-linear routines (Andrade and Duggan 2020). Second, we frame the chapter in the context of *Bayesian* statistics, in which statements about parameters and data are given in terms of *probability* (Gelman, Simpson, and Betancourt 2017). Specifically, we draw on a practical workflow to illustrate how one can think of model calibration as the result of knowledge update in the light of new information. This workflow facilitates the interpretation and communication of results in an intuitive fashion. It should be noted that although the workflow is intuitive, model calibration is essentially a statistical procedure, and as such, key concepts like *random variables*, *continuous probability distributions* and *conditional expectations* are necessary (Blitzstein and Hwang 2019). We demonstrate this workflow’s application by fitting an epidemiological model to data using HMC. In doing so, we describe the logical process followed in each step. The chapter concludes with an overview of the insights obtained from the inference process. The model is built in Stella, and all the analysis is performed in R and Stan. The code is made freely available at https://github.com/jandraor/SDR_Bayes.

2.2 Context

As mentioned in the introduction, an SD endeavour starts with a problem. For didactic purposes, we follow a widely-analysed case study (Vynnycky and White 2010). In 1918, the H1N1 virus led to an influenza pandemic that spread over the entire world in less than six months and killed tens of millions of people (Patterson and Pyle 1991). This pandemic occurred in three distinct waves, the second wave being the deadliest. Having learned from data collection difficulties in the first wave, the United States Public Health Service organised special surveys in several localities to determine as accurately as possible the proportion of the population infected (Frost and Sydenstricker 1919). From this information, we extract the report of new cases detected in the city of Cumberland (Maryland) during the autumn of 1918 (Fig 2.1). These case counts will serve as the basis to ascertain an estimate of the disease’s degree of contagiousness, a feature commonly measured by the basic reproduction number (\mathfrak{R}_0). Simply put,

this metric is the average number of secondary infections produced when one infected individual is introduced into a totally susceptible population (R. Anderson and May 1992). In addition to other techniques (Farrington, Kanaan, and Gay 2001), one can estimate \mathfrak{R}_0 by means of compartmental models (Vynnycky and White 2010).

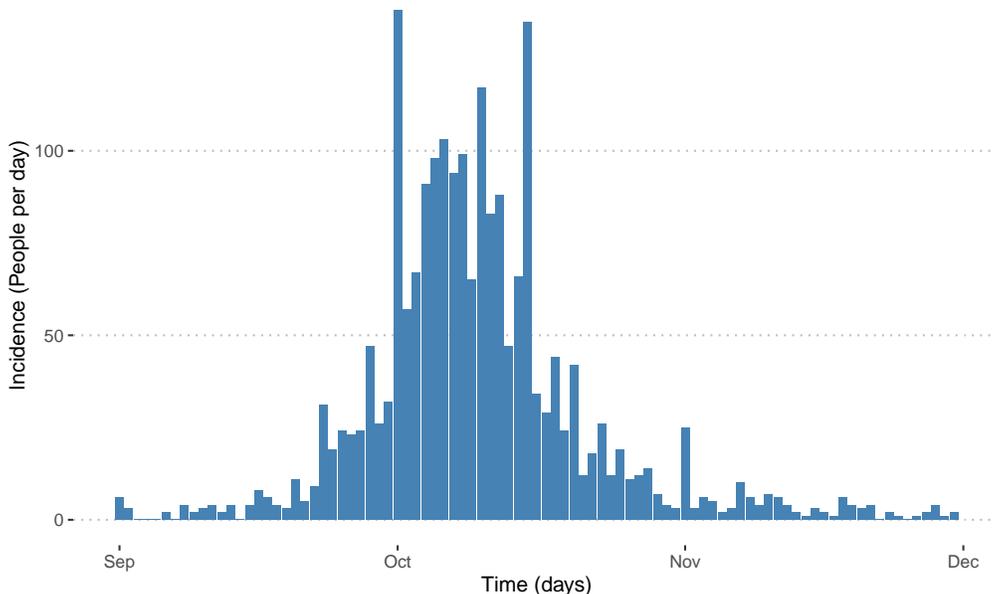


Figure 2.1: Daily number of influenza cases detected by the United States Public Health Service in Cumberland (Maryland) during the 1918 influenza pandemic, from 22 September 1918 to 30 November 1918.

A common choice for modelling the transmission dynamics of influenza is the Susceptible-Exposed-Infectious-Removed (SEIR) framework (Chowell, Nishiura, and Bettencourt 2007; Mills, Robins, and Lipsitch 2004). In this formulation (Eq (2.1)-(2.5)), $S(t)$ denotes the number of susceptible individuals at time t . Likewise, $E(t)$, $I(t)$, and $R(t)$ denote the number of exposed, infectious, and recovered individuals at time t , respectively. $C(t)$ represents the number of cumulative cases at time t . Here, we assume that the outbreak’s time-scale is much faster than the characteristic times for demographic processes (births and deaths) so that their effects are not included. Hence, it follows that the population is constant and whose size N is determined by $S(t) + E(t) + I(t) + R(t)$. Furthermore, β represents the rate of effective contacts per infected individual, σ the rate of onset of infectiousness, and γ the recovery rate. To reconcile the discrete nature of

the data and the continuous nature of compartmental models, we define the expected reported incidence (x) by Eq (2.6), where τ is restricted to non-negative discrete values. We assume that the rate of reporting ρ scales the true incidence ($C(\tau+1) - C(\tau)$). This rate reflects the fact that asymptomatic and paucisymptomatic (mild symptoms) individuals may not be *captured* by surveillance systems (Gamado, Streftaris, and Zachary 2014). From this model, \mathfrak{R}_0 can be estimated from the number of new infections caused by one infected individual in the period in which the individual is contagious (γ^{-1}), namely $\beta\gamma^{-1}$.

$$\dot{S} = \frac{-\beta S(t)I(t)}{N} \tag{2.1}$$

$$\dot{E} = \frac{\beta S(t)I(t)}{N} - \sigma E(t) \tag{2.2}$$

$$\dot{I} = \sigma E(t) - \gamma I(t) \tag{2.3}$$

$$\dot{R} = \gamma I(t) \tag{2.4}$$

$$\dot{C} = \sigma E(t) \tag{2.5}$$

$$x(\tau + 1) = \rho(C(\tau + 1) - C(\tau)) \tag{2.6}$$

2.3 Bayesian inference workflow

In an SD model, the parameters to be estimated correspond to time-independent variables and also initial conditions for stocks. To illustrate the estimation of such quantities, we follow a simplified adaptation (Fig 2.2) of a workflow from the statistical literature (Gelman et al. 2020; Gabry et al. 2019). Under this approach, we view the quantities of interest as random variables that describe our uncertainty about the actual values in the face of incomplete knowledge (McElreath 2020). Following this approach, it is possible to apply concepts of statistical inference. That is, we update our knowledge about the underlying properties that generate

the problem behaviour in light of the evidence, thus adopting a Bayesian learning perspective. To make such a process intuitive, we also draw upon data visualisation. This technique is an important tool that complements the process of model calibration, as demonstrated below.

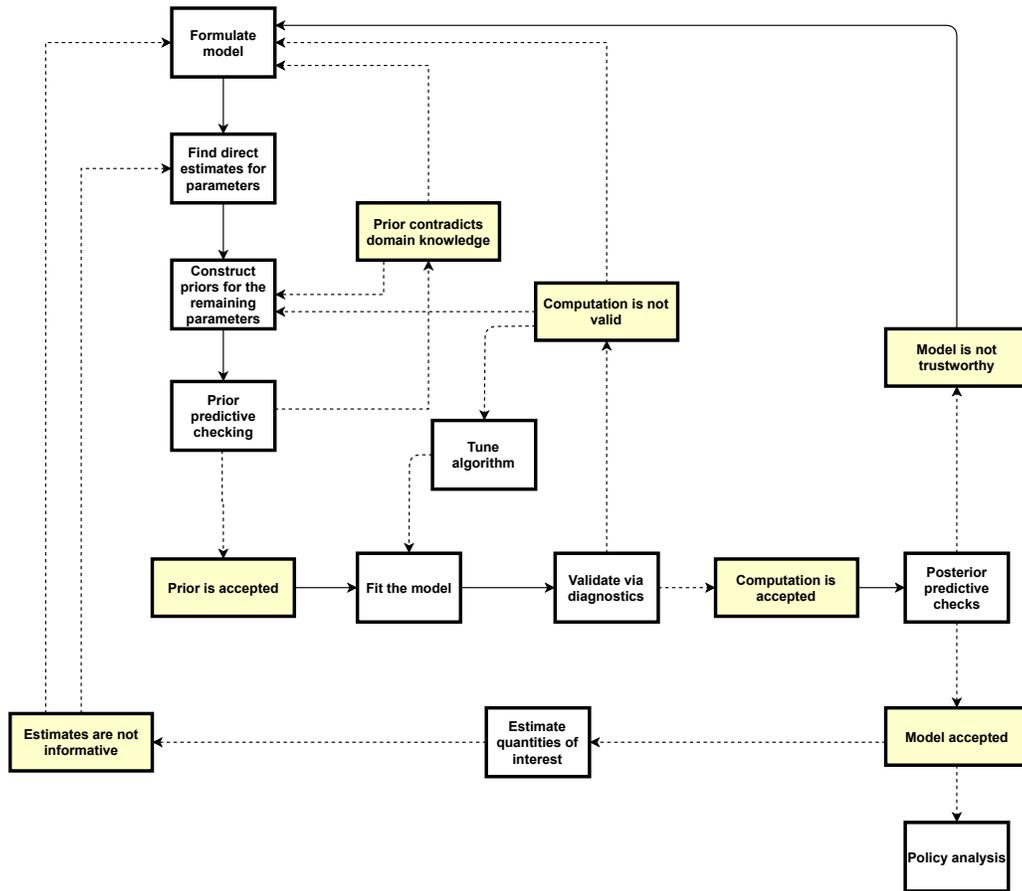


Figure 2.2: Adaptation of a Bayesian workflow (Gelman et al., 2020) to calibrate System Dynamics models. Shaded boxes indicate modeller decisions. Dotted lines indicate alternative pathways.

2.3.1 Prior information

The first step in an inference process consists of the identification of the unknown parameters in the model. In other words, whenever possible, parameters that can be directly observed or estimated from sources at the individual level should be treated as part of the known structure (Graham 1980) and removed from the process as *unmodelled predictors*. In doing so,

we mitigate the risk of model misspecification by preventing a match between actual and simulated behaviour based on unrealistic corrections to known parameter values that mask errors in the model formulation (Oliva 2003). For the remaining unknown parameters, an initial plausibility assignment should be estimated based on domain expertise, such as educated guesses, as recommended in the early days of the SD field (Graham 1980).

In the study of infectious diseases, it is common to find observational studies at the individual level that report epidemiological quantities such as the latent (σ^{-1}) and infectious period (γ^{-1}). In fact, R. Anderson and May (1992) provide estimates of such quantities for ten viral and bacterial infections, including influenza. However, measuring the average number of effective contacts by an infected person (β) remains a challenging task inasmuch as this variable encompasses individuals' social nature, the propensity of infected individuals to transmit the pathogen, and the propensity of susceptible individuals to being infected. In the same vein, continuous and exhaustive measuring of the infected population's true proportion was clearly not a viable option for the US in the early twentieth century. Consequently, we incorporate the parameters σ and γ into the model's structure, whereas β and ρ are considered the unknown time-independent variables.

In relation to the initial conditions for stocks, researchers have estimated, from serological studies in similar settings, that 70 percent of individuals were still susceptible to infection after the first wave (Vynnycky, Trindall, and Mangtani 2007). This implies $S(0) = 0.7N - I(0)$ and $R(0) = 0.3N$. For simplicity purposes, we assumed there were no exposed individuals at the beginning of the second wave. Namely, $E(0) = 0$. These assumptions leave as the only unknown the number of individuals that trigger the outbreak ($I(0)$), which is assumed as the initial value for the stock that tracks the number of cumulative cases.

Consequently, in this example, we focus on the estimation of the three parameters identified above ($\beta, \rho, I(0)$) denoted by the vector θ . For each one, we outline their plausibility before assessing the evidence (Fig 2.3). For β and $I(0)$, our domain knowledge indicates that they should be non-negative. Further, we suppose that these two quantities concentrate at low values considering the slow progression at the outbreak's start (Fig 2.1). This formulation does not discard values away from such concentration. However, we assign them small plausibilities, measured by the height of the function. This height is known as the probability density, and it is often modelled by

standard statistical distributions that we denote by the Greek letter π . In this case, $\pi(\beta)$ and $\pi(I(0))$ are distributed according to the *lognormal*(0, 1). We choose this distribution to reflect our belief that these parameters should be positive and relatively small. Regarding the reporting rate ρ , although we are unsure of its magnitude, we know that it should be between zero and one, and by including it in the model, we tacitly assumed that it should be far from the boundaries. If we had thought that the parameter was close to zero, we would have discarded the reported cases (C) stock. If we had thought the value was close to one, we would not have needed the reporting rate parameter. We model this assumption by $\pi(\rho) \sim \text{beta}(2, 2)$. Taking into account that no evidence suggests otherwise, we assume independence in the parameters. That is, having information about one parameter does not provide knowledge about the others. Mathematically, $\pi(\theta) = \pi(\beta)\pi(I(0))\pi(\rho)$. In statistical language, this is known as the prior distribution. Interested readers are referred to Gelman, Simpson, and Betancourt (2017) for philosophical and practical considerations about the prior distribution.

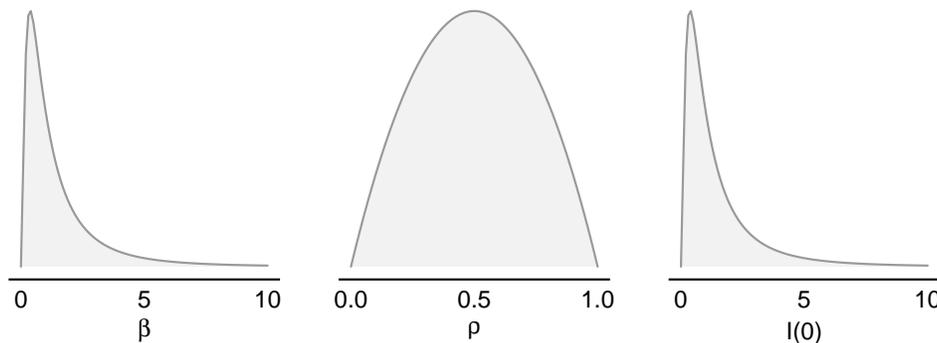


Figure 2.3: Prior distribution. Shaded areas indicate probability mass.

2.3.2 Observational model and probability of the data

$$\pi(y|\theta) = \text{Pois}(y|x(\tau)) \quad (2.7)$$

$$\pi(y) = \int \pi(y|\theta)\pi(\theta)d\theta \quad (2.8)$$

Naturally, it is expected that prior knowledge of the parameters leads to dynamics that capture the essence of the problem being studied. In other

words, if we plug the vector θ into the simulation model, the latter ought to produce outbreak-like trajectories, including the observed behaviour. Thus far, the simulation model (Eq (2.1)-(2.6)) is deterministic. That is, the model always produces the same smooth output from a given configuration. Unsurprisingly, even the most *perfect* configuration (or any other) will yield values that differ from the measurements as the model only approximates the studied phenomenon. Therefore, these differences must also be formally accounted for by a formulation, $\pi(y|\theta)$, which we refer to as the *measurement or observational model*. Since daily reported cases are non-negative discrete quantities, we formulate the observational model (Eq (2.7)) in terms of a Poisson distribution (see Appendix A.1 for a discussion on the distribution choice). Note that with the addition of Eq (2.7), a single model configuration can yield different reported incidences. Thus, when θ is fixed to a single set (θ_i), the observational model is known as the *sampling distribution*, $\pi(y|\theta_i)$.

In an ideal and unrealistic scenario, one would generate infinite samples (θ_{sim}) from the prior distribution $\pi(\theta)$, then feed the simulation model with those samples to produce the entire universe of possible trajectories of reported incidences y_{sim} . Once the complete set of y_{sim} has been sampled, one aggregates similar trajectories to establish which behaviours are more likely to be observed than others. Formally, this is expressed by Eq (2.8), where $\pi(y)$ denotes the *average probability of the data* (McElreath 2020) or *prior predictive distribution* (Gelman et al. 2013). Although generating infinite samples is infeasible, one can draw a finite number of samples to reason about the model’s behaviour conditioned on current knowledge. Accordingly, we draw 500 random samples ($\theta_1, \theta_2, \dots, \theta_{500}$) from the prior distribution to generate an equal amount of trajectories (y_1, y_2, \dots, y_{500}). We present the results in Fig 2.4). Here, we notice that large swathes of samples generate outbreak-like behaviours, possibly the observed data (solid points). Overall, this process is referred to as *prior predictive checking*, and it is a powerful tool for understanding the structure of models (Gabry et al. 2019). Prior predictive checking aims to answer the question “*Could this prior generate the type of data we expect to see?*” (Gelman, Simpson, and Betancourt 2017). Should none of the simulations had resulted in outbreak-like behaviours, or should these behaviours had not captured the observed data, it would have been an indication for reassessing the validity of the prior distribution or the model itself (Gelman et al. 2020).

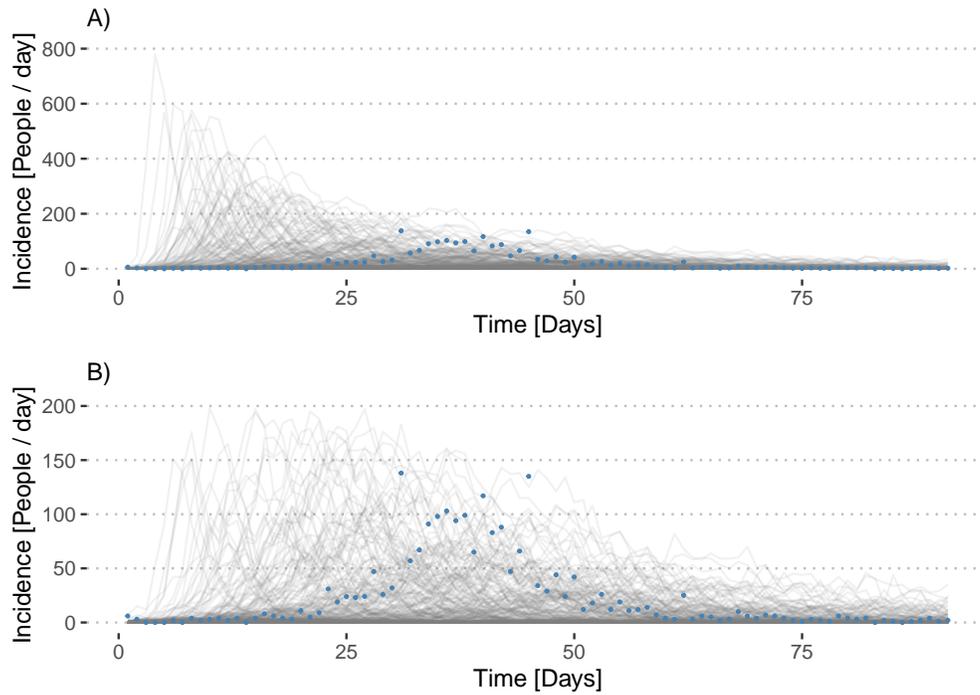


Figure 2.4: Prior predictive checks. A) Simulation of five hundred predicted incidence measurements (grey lines) from the proposed dynamic hypothesis and the prior distribution. Dots denote the actual data. B) Zoom to trajectories that may resemble the actual data. We show predicted measured incidences whose peak is lower than 200 new cases in a day.

2.3.3 Expectation

$$\pi(\theta|y_c) = \frac{\pi(\theta)\pi(y_c|\theta)}{\pi(y_c)} \quad (2.9)$$

$$\mathbb{E}[f(\theta)] = \int f(\theta)\pi(\theta|y_c)d\theta \quad (2.10)$$

Thus far, we have considered the likely behaviours over time that we could have observed. Nonetheless, the chief interest is performing parameter inference based on the available data set rather than on the infinite set of possible observations. We thus shift the focus from the universe of measurements (y_{sims}) to the observed behaviour, Cumberland's incidence data (y_c). In doing so, the density function $\pi(y_{sims})$ becomes the constant $\pi(y_c)$. Furthermore, when the observational model is regarded as a function of θ , for a fixed y , it is called the *likelihood function* (Gelman et al. 2013). This mathematical construct, $\pi(y_c|\theta)$, is the target of optimisation algorithms and a statement about the data, which quantifies the relative consistency of each model configuration with the observed data. Simply put, if θ_1 produces a larger likelihood value than θ_2 , then y_c is more likely to have occurred from θ_1 . However, in a Bayesian setting, the plausibility of a trajectory is not the desired outcome. On the contrary, the interest lies in establishing which values of the vector θ are more plausible than others given the observed trajectory (y_c), or in more formal terms, the posterior distribution of the estimated parameters, $\pi(\theta|y_c)$. Conveniently, by Bayes theorem (Eq (2.9)), we can express the posterior distribution in terms of the prior distribution, the probability of the data, and the likelihood function. Given that the posterior encodes all the information learned by our model, one could extract inferences about the data and the parameters using expectations (Eq (2.10)). As a matter of fact, prior predictive checking is an expectation, where f is the observational model averaged over the prior distribution (instead of the posterior). In the sections below, we will see that, should a solution for Eq (2.9) be available, obtaining a model fit is nothing more than the application of Eq (2.10). Regardless of the query, answers are always given in probabilistic terms. Indeed, this is the main feature of the Bayesian approach, where uncertainty is quantified with probability distributions.

2.3.4 Markov chain Monte Carlo

Consequently, from a Bayesian perspective, one approaches *parameter inference* as the process of finding the posterior distribution. It is often the case that *closed-form* solutions do not exist for such a type of formulation (Robert and Casella 2005). To address this difficulty, in the late 1940s, researchers at Los Alamos developed stochastic simulation techniques, known as Monte Carlo Methods (Robert and Casella 2011). Early conceptualisations employed *exact sampling* (Robert and Casella 2010), the generation of independent and identically distributed (IID) samples to explore the extent of the parameter space unconditionally. Regions of high probability, however, are concentrated on specific locations rather than being scattered around (Betancourt 2018). Thus, IID sampling would squander finite computational resources on low probability regions until they are eventually exhausted before reaching the target location. Aware of this, this same group of researchers enhanced the method by generating *correlated samples* from a Markov chain to approximate the equilibrium distribution of a liquid (Metropolis et al. 1953). Hence, the term Markov chain Monte Carlo (Geyer 2011). Even though further improvements in subsequent decades, such as the Metropolis-Hastings algorithm (Hastings 1970) and the Gibbs sampler (Geman and Geman 1984), broadened the method's scope, it was only until the early 1990s (Gelfand and Smith 1990), and partly due to the growth in computational power, that the mainstream statistical community widely noticed the method (Robert and Casella 2011). Since then, there has been significant growth in the number of applications to a wide range of fields, including epidemiology (Davies et al. 2020; Chatzilena et al. 2019).

A Markov chain is a sequence of random variables $\Theta_1, \Theta_2, \dots, \Theta_n$, in which each variable depends only on the previous one (Blitzstein and Hwang 2019). To iteratively draw samples or realisations, we apply a conditional probability distribution denoted by $T(\Theta_{i+1}|\Theta_i)$, also referred to as the *transition kernel*. The strength of this approach lies in the improvement achieved at the generation of each new sample. Improvement in the sense of converging to the target distribution $\pi(\theta|y_c)$. If run long enough, the Markov chain is expected to reach an equilibrium state - or stationary state- where the samples describe the posterior distribution. This approximation has the advantage that it does not impose constraints in the shape of the posterior. Under ideal conditions, the Markov chain starts from any place in the parameter space and gradually moves towards the target distribution. This initial phase is known as *warm-up*. Once the target distribution has been found, the Markov

chain explores high-probability regions (*sampling phase*), namely parameter values that have larger contributions to the observed behaviour. To obtain unbiased estimators, one discards the samples from the warm-up phase.

Although theoretically, the Markov chain will eventually reach the stationary state; in practice, this result is not guaranteed, especially for high-dimensional target distributions and distributions that exhibit non-trivial dependencies among the parameters (Betancourt 2017). Early implementations of MCMC, such as the Metropolis-Hastings and Gibbs samplers, become slow at exploring complex parameter spaces to the extent that computational resources are depleted before providing accurate estimates. This inefficiency occurs due to these algorithms' random walk behaviour to generate new samples, resulting in zig-zag movements across the parameter space (Gelman et al. 2013). For instance, Pierson and Sterman (2013) report that in the calibration of an industry-level model of airline profits, “over 1 million MCMC runs were needed to arrive at stable estimates for the confidence bounds”.

2.3.4.1 Hamiltonian Monte Carlo

According to Betancourt (2017), MCMC has benefited from an evolving interplay between statistics and physics from its inception to present developments. Conceiving a statistical system as a physical one provides an innovative way to improve computational methods. Indeed, the realisation that molecular simulation methods -in which the motion of molecules was deterministic, following Newton's laws of motion-, and MCMC could be combined, yielded a technique of wide applicable potential. In such a framework, the description of molecular motion has an elegant formalization as *Hamiltonian dynamics*, hence the term *Hamiltonian Monte Carlo* (Neal 2011). In particular, the HMC algorithm simulates the movement of a fictitious and frictionless particle (McElreath 2020) over a surface whose ruggedness is determined by the likelihood function and the prior distribution. Formally, the Hamiltonian function- the sum of potential and kinetic energies- describes such mechanics (Neal 2011). In turn, this function depends on the characterisation of each parameter in terms of *position* and *momentum*. The former is straightforward, considering that it corresponds to the fictitious particle's location in the parameter space; for the latter, the algorithm adds an artificial variable per parameter. As a result of this conceptualisation, the random-walk behaviour from early MCMC implementations is suppressed (Gelman et al. 2013), resulting in a tool that becomes efficient at traversing the complex parameter spaces. For an analytical treatment of the method,

the reader is referred to Neal (2011). Likewise, Betancourt (2018) offers an intuitive description. In the appendix A.2, we provide an example where we compare how HMC and the Metropolis algorithm explore parameter spaces.

2.3.4.2 Stan

Although HMC is a powerful method, its geometrical foundations (Betancourt et al. 2014) render *ad hoc* implementations onerous. To address this challenge, a group of researchers developed Stan (Carpenter et al. 2017), a statistical modelling platform. This tool provides an interface to perform Bayesian inference via the No-U-Turn-Sampler or NUTS (Hoffman and Gelman 2011), an HMC algorithm. NUTS takes advantage of the *warm-up phase* to identify the algorithm's configuration that best adapts to the user-supplied model for efficient parameter space explorations, resulting in significant gains in sampling speed. Furthermore, Stan supports gradient evaluation (Carpenter et al. 2015) to a broad range of distribution families and ODE solvers. In spite of this support, in some cases, SD practitioners will not be able to avail themselves of familiar builtins (such as those offered by Vensim and Stella). As a result, they will have to formulate equations explicitly, or in the case of table functions, the practitioner will have to devise parametric formulations.

In practice, Stan only requires, from the practitioner, the specification (code) of the model's equations, the prior distribution, the likelihood function, the data, and the number of draws. To these specifications, Stan runs the NUTS algorithm internally and returns a set of samples for each parameter. To run the simulation, one can directly interact with Stan through a command interface or popular statistical software such as Python and R. In this case, we choose the latter to draw upon the package *readsdr* (Andrade 2021), which automatically converts XMLE files from Stella and Vensim to Stan code.

2.3.4.3 Diagnostics

Bayesian inference via iterative simulation is performed by extracting insights from the entire collection of simulated draws from the sampling phase. Specifically, we estimate posterior probability densities and compute quantities of interest that describe the calibrated parameters, such as expected values (mean) and credible intervals. However, if the chains are not run long enough, predicted convergence to the stationary distribution may not be achieved. The resulting draws may describe partially, or even, inaccurately the pos-

terior distribution, thereby producing unreliable estimates. To address this challenge, an effective strategy consists of running at least four chains that start from different locations in the parameter space and verify that all of them converge to the same region. Accordingly, we run four chains of 2,000 iterations in this example: 1,000 allocated to warm-up and 1,000 to sampling. We employ Stan’s default initialisation strategy, which starts the chains from values drawn uniformly from the interval $(-2,2)$. If the estimated parameters are bounded, then Stan applies the appropriate transformation.

Graphically, one can inspect convergence via trace plots. These visualisation tools are time-series of the draws for a particular parameter. Here, *time* refers to the order in which the draws were sampled. In Fig 2.5A, we present the sequence of the first 100 draws for each of the three calibrated parameters. Initially, the chains traverse the parameter space before settling on a unique location. By augmenting the time frame to the complete set of samples (Fig 2.5B), it can be seen that the chains mixed; that is to say, the draws trace out a common distribution. Additionally, there is no obvious trend or change in the spread in the chains. In other words, they are stationary. These two properties suggest that the sampling procedure reached the predicted convergence. Quantitatively, the potential scale reduction factor (Gelman and Rubin 1992) denoted by \hat{R} is a useful metric to validate this assessment. This statistic compares within-chain variance (stationarity) to between-chain variance (mixing). At convergence, \hat{R} should be < 1.01 (Vehtari et al. 2021), whereas higher values indicate that the chains describe different locations of the parameter space or different trends within a single chain. In this example, all chains exhibit Potential Scale Reduction factors below this threshold (see appendix A.3 for a technical description and results).

Other diagnostics to gain confidence in the results include the Effective Sample Size (*ESS*). In general, simulation inference from correlated samples is less precise than from the same number of independent samples (Gelman et al. 2013). If the correlation among samples is strong, chains must be run for longer periods in order to obtain accurate estimations. To measure this correlation, we employ *ESS* to determine the number of independent simulation draws from the MCMC process. For reliability, this metric should be above 400 (100 per chain) per parameter (Vehtari et al. 2021), as in this example (see appendix A.3 for a technical description and result).

To conclude with MCMC diagnostics, the Hamiltonian approach also allows us to evaluate the robustness of the results. A key feature of the Hamiltonian

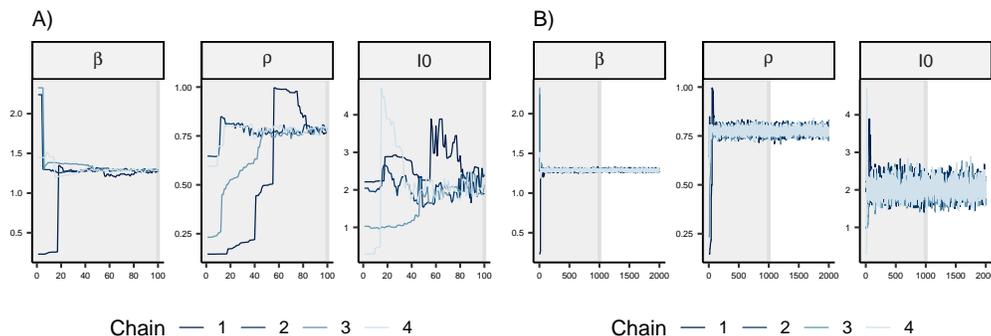


Figure 2.5: A) Early warm-up phase for our three parameters (first 100 iterations). B) Complete sequence of iterations (2000) in the warm-up and sampling phases. The shaded area indicates the warm-up phase. In both figures, we present four Markov chains per calibrated parameter through sequences of points (samples obtained from Stan) joined by lines.

function (sum of potential energy and kinetic energy) in the sampling phase is that it remains invariant along the trajectory in which the particle moves. Any divergence from its initial value indicates pathological behaviour (abnormal movements) in the chains to the extent that they cannot be trusted, and the calibration setup (SD model, prior, likelihood, algorithm’s parameter values) must be reformulated. By default, Stan reports divergences and provides ways to access which iterations encountered divergences ¹.

Given this example’s didactic scope, all diagnostics unsurprisingly return favourable results. In practical applications, however, the path to these results can be significantly less straightforward. From code bugs to structural problems, such as non-identifiability in the model (two or more parameterisations that are observationally equivalent), achieving convergence can take several iterations. To complicate matters, exploring the target distribution of differential equation models involves expensive gradient evaluations that slow down the HMC algorithm, limiting the number of debugging runs. Thus, the analyst must efficiently identify the problem’s source. To this end, Gelman et al. (2020) recommend the process of *fake data* (also known as *synthetic data* or *simulated data*). That is, feeding the simulation model with known and plausible parameter values to obtain behaviours over time similar to the real data being analysed. Then, we should check whether the same model and the inference method can recover the known values. In

¹See Stan Manual for more details

doing so, it is possible to identify strategies to address computational issues. These strategies range from model simplification and more data collection to recognising the method’s inappropriateness for the studied problem. For instance, HMC works correctly under well-defined posterior densities, and it is restricted to continuous parameters. Conversely, challenging geometries with sharp corners or multiple modes (Betancourt 2015) in the posterior distribution render the algorithm impractical, and other types of methods should be employed (Valderrama-Bahamóndez and Fröhlich 2019). We refer the reader to Gelman et al. (2020) for a comprehensive treatment of these methodological issues. Furthermore, in Appendix A.4, we draw upon *synthetic data* to illustrate the Bayesian workflow presented above (Fig 2.2) in the context of *wrong* assumptions, complex parameter spaces, and the necessity for data collection.

2.3.5 Posterior distribution

Bayesian inference is concerned with updating knowledge in the light of new evidence (McElreath 2020). Once we have gained confidence in the sampling procedure, we take the draws returned by Stan and construct probability densities. By restricting the analysis to a single parameter (marginal posterior distribution), it is possible to determine which values are plausible for the parameters after seeing the data. We can visually portray such a knowledge update process by comparing marginal prior and posterior distributions (Fig 2.6). In this graph, we observe that the concentration of probability shifted for each parameter. For instance, before the calibration, we assumed *ignorance* for the reporting fraction (ρ). On the contrary, the marginal posterior distribution indicates that 95% of the samples- or the 95% credible interval - concentrates on the region [0.74, 0.81]. In relation to \mathfrak{R}_0 , we estimate its 95% credible interval between 2.53 and 2.63, a value consistent with the estimate ([2.56-2.59]) reported by Vynnycky and White (2010). Furthermore, Mills, Robins, and Lipsitch (2004) estimate that \mathfrak{R}_0 for 1918 pandemic influenza was approximately between 2 and 3. Similarly, Vynnycky, Trindall, and Mangtani (2007) concluded that this value was in the range of 2.4–4.3 in community-based settings. Thus, during this pandemic, one infected individual could potentially infect, on average, almost three susceptible ones.

In the SD literature, it is not unusual that researchers restrict model calibration results to reports of the mean and standard deviation of the fitted

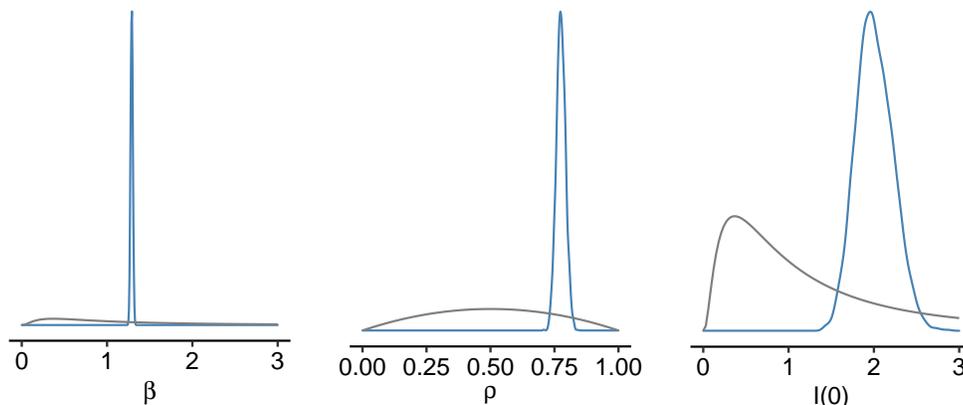


Figure 2.6: Comparison between marginal prior and marginal posterior distributions for our three parameters. Grey lines denote prior distributions. Blue lines denote posterior distributions

parameters. Although useful for descriptive purposes, we instead have focused our interest on the complete set of samples. The reader should bear in mind that in Bayesian inference, we quantify uncertainty by an entire probability distribution, which cannot be characterised by a single point estimate (we elaborate on this issue in Appendix A.5). Through the samples obtained from MCMC methods, for instance, we can extract rich information about the parameter interactions. Initially, given the lack of evidence, we assumed independence among the calibrated parameters. Nevertheless, System Dynamics models, by definition, depict problems as an interconnected confluence of factors. Based on this logic, it would be surprising that a parameter does not interact with another. To explore this property, we draw upon *pair plots*. This graphical tool displays all possible pairwise combinations (joint) of probability distributions. Moreover, we include the correlation for each combination along with the marginal distributions to gain a global perspective (Fig 2.7). This plot shows a strong interaction among the parameters to the extent of an almost perfect correlation between two parameters ($I(0)$ and β), indicating that, relatively speaking, *large* values of β are solely compatible with *low* values of $I(0)$. Consequently, independence assumptions are unwarranted. The implications of this finding are explored in the Policy Analysis section.

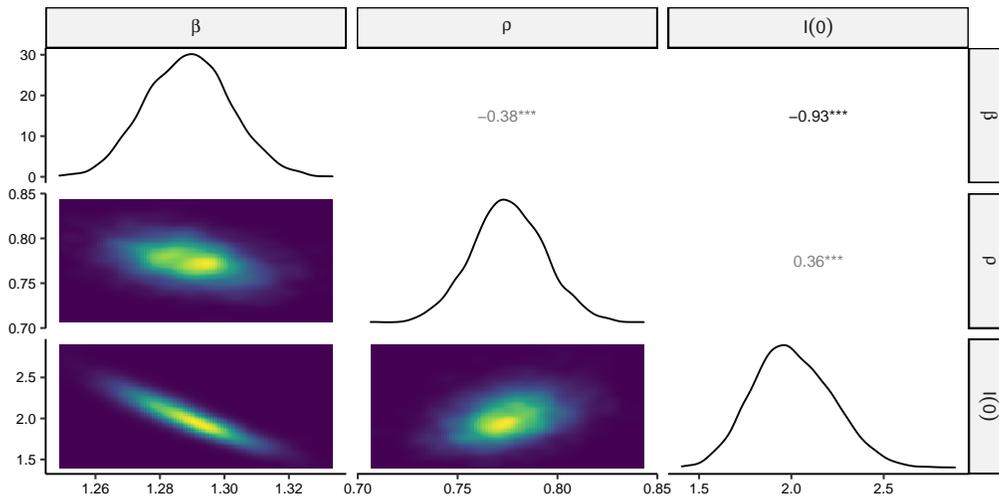


Figure 2.7: Posterior distribution for our three parameters. The diagonal shows the posterior marginal distributions. In the lower triangular part, the joint posterior distribution of each possible combination of two parameters is displayed. The upper triangular part shows the correlation among parameters.

2.3.5.1 Posterior predictive checks

$$\pi(y|y_c) = \int \pi(y|\theta)\pi(\theta|y_c)d\theta \quad (2.11)$$

The ultimate purpose of model calibration is to search for a match between observed and simulated behaviour that builds confidence in the proposed dynamic hypothesis. Following this Bayesian workflow, we frame this purpose as: “if a model is a good fit we should be able to use it to generate data that resemble the data that we observed” (Gabry et al. 2019). Notice that this statement is similar to Oliva’s quote (Oliva 2003): “Confidence that a particular structure, with reasonable parameter values, is a valid representation increases if the structure is capable of generating the observed behavior”. To provide an answer, we can use the posterior distribution to obtain predictions for the measured quantities and compare them to the observed data (Gelman and Hill 2007). Thus, in this case, $f(\theta)$ (see Eq (2.10)) corresponds to the observational model. This process is analogous to prior predictive checking, with the difference that we average over the posterior distribution. Unsurprisingly, this process is called *posterior predictive checking* (Eq (2.11)). In consequence, obtaining a model’s fit under this Bayesian paradigm is equivalent to solving Eq (2.11).

Accordingly, we generate 500 draws from $\pi(\theta|y_c)$ and insert them into the observational model to obtain predictions for the measured incidences (Fig 2.8). Qualitatively, the simulated trajectories appear to be reasonable approximations to the reference behaviour. To verify this appraisal, for each trajectory, we calculate the mean absolute scaled error or MASE, a metric of forecast accuracy (Hyndman and Koehler 2006). In practice, this procedure entails to define $f(\theta)$ as the MASE of each trajectory and average the results over $\pi(\theta|y_c)$. Considering that values lower than one indicate adequate predictive performance, and all simulated behaviours concentrate below such a threshold (see Appendix A.6), we gain support to the claim that the simulation model explored in this chapter is an adequate structure to account for Cumberland’s incidence data.

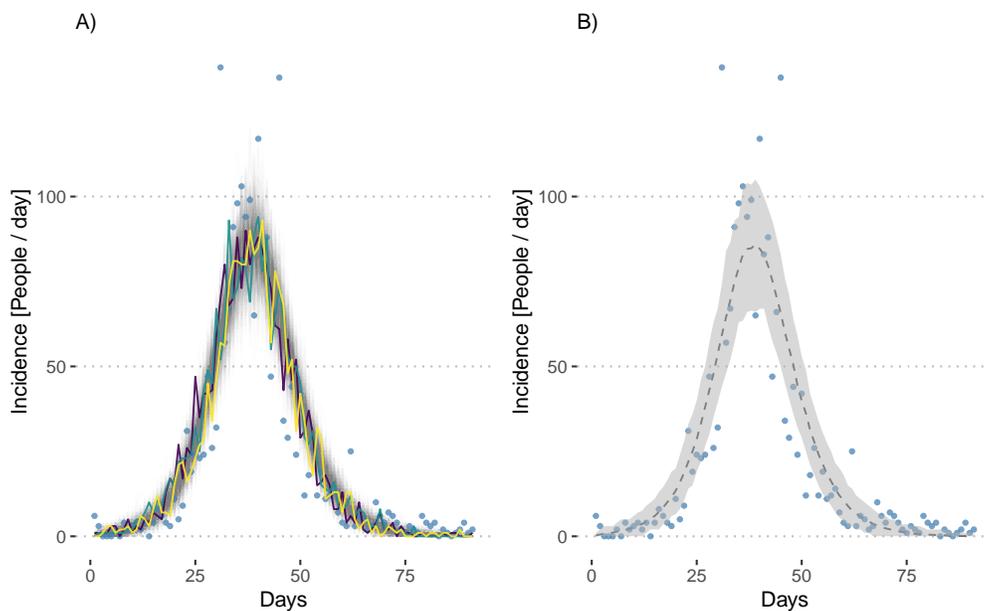


Figure 2.8: A) 500 predicted incidence measurements (grey lines) from the posterior predictive distribution (model’s fit). To obtain a single predicted measurement, we draw a sample from the posterior distribution and use it to generate a trajectory from the observational model (Eq 2.7). Three different predicted incidence measurements are highlighted in Viridis colours. Dots denote the actual data. B) Posterior predictive distribution described in terms of the mean (dashed line) and 95% credible intervals (contour).

2.3.5.2 Policy analysis

As we have seen, the usefulness of estimating $\pi(\theta|y_c)$ is not exclusively confined to find a match between observed and simulated behaviours (model calibration). Once we have gained confidence in these results, we can employ this distribution to evaluate the future dynamics in similar settings where the model is relevant. To illustrate this procedure, we simulate the model in a hypothetical situation. In particular, we are interested in predicting the dynamics of an outbreak in a city of 10,000 people under two scenarios: *unmitigated* and *intervention*. The former corresponds to the scenario where the virus is let to run unchecked until the disease runs its course. The latter describes the implementation of social distancing measures aimed at reducing the number of contacts among the population. To do so, we consider the sampling procedure. Stan returns a collection of draws for each calibrated parameter. In this case, the output forms a matrix of three columns (parameters) and four thousand rows (samples). Since the parameters exhibit correlation (Fig 2.7), we sample entire rows $\{\beta_i, \rho_i, I(0)_i\}$, where i denotes a specific row. Should the parameters be independent, we would sample separately from each column, yielding sets $\{\beta_j, \rho_k, I(0)_l\}$. We follow this procedure in situations where we cannot infer correlations from the data (e.g., report of marginal distributions).

Having established the sampling procedure, for the no intervention scenario, we feed the SEIR model with the samples and run the simulation; whereas for the intervention scenario, we multiply all β_i by a factor of 40% to describe the effect of social distancing measures implemented before the occurrence of the first case. As expected, decreasing the population's contact rate (β) translates into a slower transmission process with fewer cases. Undoubtedly, the added value of performing policy analysis from this Bayesian perspective stems from the fact that we simultaneously gauge the uncertainty in the predicted behaviours, offering a broader picture to decision-makers. Nevertheless, in this case, such uncertainty is tempered by the correlation among parameters. To visualise this, we also run the model with independently sampled draws. In Fig 2.9, it can be seen the extra uncertainty added by the independence assumption, evidenced by the excess of blue contour in comparison with that of the grey one. Notice that this application is also an instance of Eq (2.10), namely an expectation.

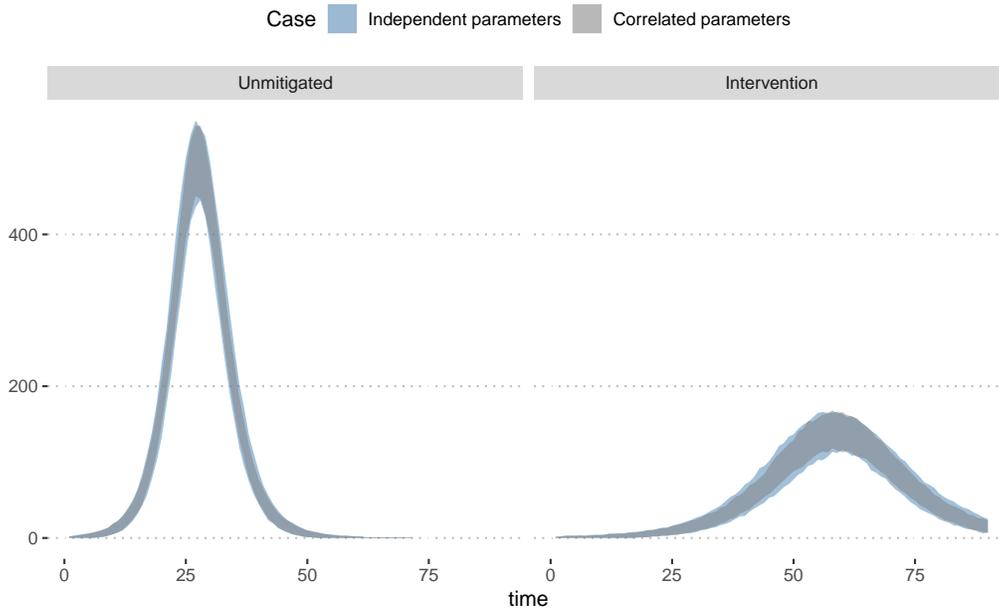


Figure 2.9: Forecast of scenarios (unmitigated and intervention) under two parameter interaction assumptions. These assumptions correspond to the correlations revealed by the calibration process and perfect independence. We simulated 500 trajectories per experiment in this two-by-two design. For each experiment, we plot a silhouette of the predicted measured incidence. The width of these silhouettes corresponds to 95 % credible intervals. Two silhouettes are superimposed per scenario.

2.4 Performance comparison

SD practitioners certainly require a clear indication that investing time and resources in applying a novel method is worth the effort. In the SD literature, we find two enhanced versions of the Metropolis algorithm: the *DREAM* sampler (Vrugt et al. 2009) implemented in *Vensim*, and a Random-Walk Metropolis -RWM- algorithm (*MCMCmetrop1R*) offered by the *MCMCPack*, the method used by Osgood and Liu (2015). We select the latter for the comparison analysis, given that it is open-source and a similar approach adopted throughout this work.

Specifically, we fit the SEIR model (presented in the *Context* section), under the conditions (priors and unknowns) described in the *Prior information* section, to Cumberland’s incidence data. This calibration is performed in six different scenarios, which differ in the number of iterations (100, 200, 500, 1000, 1500, 2000) allocated to both MCMC algorithms. The results show that HMC is computationally faster than RWM for obtaining an equal amount of samples (Appendix A.9). However, given technological implementations, it is not possible to definitively determine whether the performance differences are due to the algorithms themselves. For this reason, we compare technologically-independent metrics of convergence (\hat{R}) and efficiency (*ESS*). In Fig 2.10A, we observe that RWM requires at least 2000 burn-in samples so that all parameters reach convergence ($\hat{R} < 1.01$), a value significantly higher than the equivalent number of samples (500) required by HMC. On the other hand, the effective sample size (*ESS*) is a measure of efficiency. This metric helps us answer: *Are X samples from RWM equivalent to X samples from HMC?* The reader should recall that the *ESS* approximates the number of independent samples. We present two types of *ESS*: *bulk* and *tail*. Both metrics should be at least 400 (Vehtari et al. 2021). In Fig 2.10B, it is observed that HMC produces a higher number of *ESS* and does so more efficiently (Fig 2.10C) than RWM in all scenarios. Furthermore, HMC exceeds the 400-threshold from 500 iterations per chain, a third of the iterations required by RWM. In a nutshell, HMC converges faster, and its samples provide more information than those of RMW.

In terms of performance, we corroborate previous theoretical (Betancourt 2018) and practical studies that suggest HMC as the method of choice. For instance, Monnahan, Thorson, and Branch (2017) found that HMC outperforms the Gibbs sampler (a random-walk MCMC algorithm) in estimating the parameters of population ecology models (hierarchical and state-space)

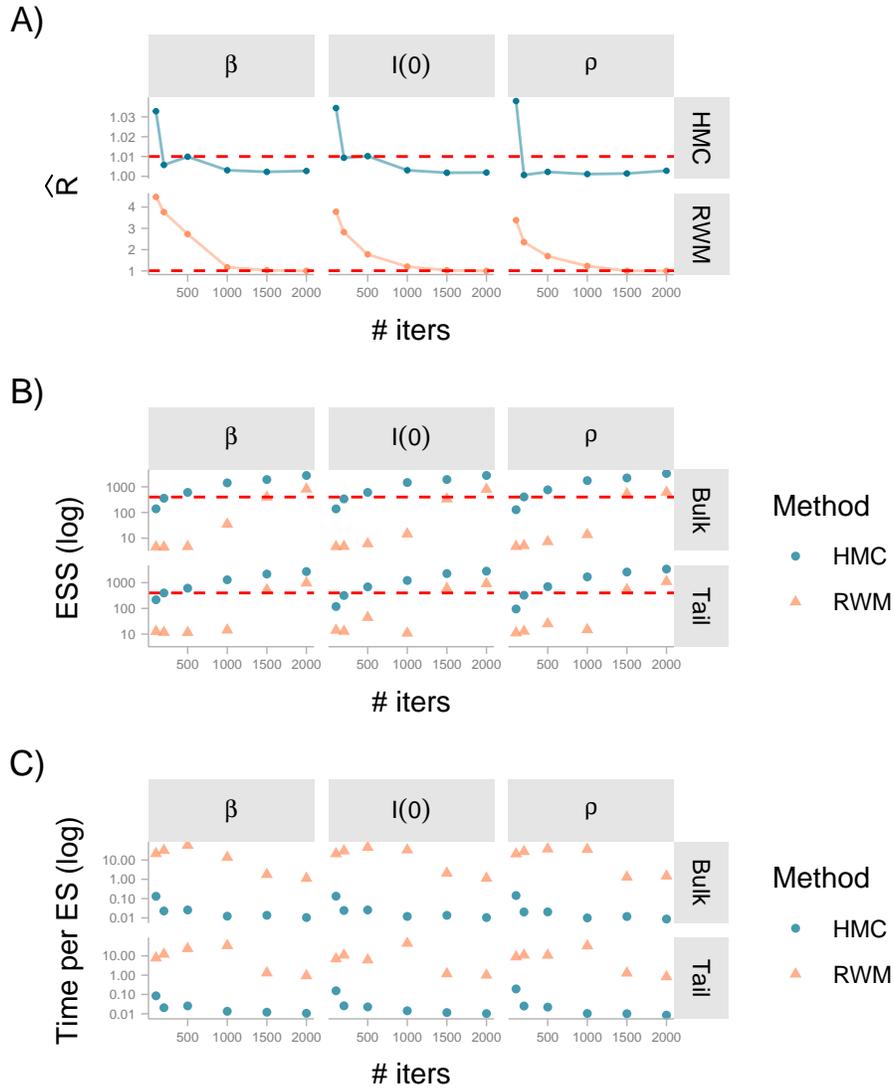


Figure 2.10: A) Potential scale reduction factor by scenario and calibration algorithm for our three parameters. Red dashed line denote the acceptance threshold (1.01). B) Effective sample size (bulk and tail) by scenario and calibration algorithm for our three parameters. Red dashed line denote the acceptance threshold (400) C) Clock time (seconds) per effective sample by scenario and calibration algorithm.

across a range of dimensions and complexity. This performance gap grew to the extent that HMC was 63 times more efficient when fitting a logistic model. Likewise, Beraha, Falco, and Guglielmi (2021) conducted a systematic study on probabilistic programming languages (PPL). The authors evaluated three PPLs on four classes of models: linear, logistic regression, mixture models, and accelerated failure time. The results from this study indicate that Stan (using the NUTS algorithm) is the “*default go-to software*” over the other two random-walk-based platforms. To the best of our knowledge, the most recent benchmark analysis on ODE models was carried out a decade ago (Girolami and Calderhead 2011), a time where the NUTS algorithm had not been developed. This observation suggests that future research endeavours should systematically explore the variations in the performance of MCMC algorithms across various ODE models.

2.5 Conclusion

In his reflection on the sixty-year history of the SD field, Sterman (2018) encouraged practitioners to “*master the state of the art and modern methods to develop, test, communicate, and implement rigorous, reliable and effective insights into the dynamics of complex systems*”. In that context, we introduce Hamiltonian Monte Carlo to the SD community to perform robust model calibration using a state-of-the-art statistical package (Stan). In doing so, we notice that valuable information about the results and the method itself is often missing in model calibration reports. Due to the established tradition of using non-linear optimisation techniques, we often find that authors limit the calibration report to the mean and standard deviation of parameter estimates. This information is complete only in the case of symmetric and independent distributions, a set of assumptions that do not hold for this simple case study. On the contrary, MCMC-based methods provide richer information about the calibrated parameters and allow the practitioner to evaluate the robustness of the estimates. In the same fashion that a model’s behaviour must be obtained from the right reasons, parameter estimates must be obtained for the right reasons as well. To communicate this process, we suggest a workflow grounded on logic and visualisation. Such a workflow is possible due to the combination of SD Software, R, and Stan. This synergy produces robust results, facilitates reproducibility, and, more importantly, enhances the comprehension of the process undertaken.

Chapter 3

The basic reproduction number and the mean generation time

The work presented in this chapter was published in:

Andrade, J., Duggan, J., 2023. Anchoring the mean generation time in the SEIR to mitigate biases in \mathfrak{R}_0 estimates due to uncertainty in the distribution of the epidemiological delays. *Royal Society Open Science*, 10(8).

3.1 Introduction

The analysis of any infectious disease's dynamics will inevitably lead to the *basic reproduction number* (\mathfrak{R}_0). Initially developed in the study of demographics ([Dublin and Lotka 1925](#)), this quantity has been interpreted in the epidemiological context as the average number of secondary infections arising from the introduction of one infected individual into a totally susceptible population ([R. Anderson and May 1992](#)). The usefulness and importance of \mathfrak{R}_0 lie primarily in its *threshold phenomenon* ([Heffernan, Smith, and Wahl 2005](#)). That is, a pathogen can invade a totally susceptible population only if $\mathfrak{R}_0 > 1$ ([Keeling and Rohani 2011](#)). Furthermore, the magnitude of \mathfrak{R}_0 gauges the transmission potential of an emerging infectious disease ([Heffernan, Smith, and Wahl 2005](#)) and the effort required to control the invading pathogen ([Wallinga and Lipsitch 2007](#)). Thus, accurate estimation of \mathfrak{R}_0 is crucial for understanding and managing infectious diseases.

Another reason for the popularity of \mathfrak{R}_0 is that one can estimate it from

epidemiological data (Kucharski 2020) using a number of methods. For diseases that allow the assumption of endemic equilibrium and homogeneous mixing, one can follow Mollison’s method (Mollison 1995) or Dietz’s approach (Dietz 1975). The former requires prevalence data, whereas the latter leverages readily available information such as age at infection and average life expectancy. On the other hand, if an infection leads to either immunity or death in a closed population, seroprevalence studies can inform the fraction of the population that acquired the disease during an epidemic, i.e. the *final epidemic size*. In their seminal paper, Kermack and McKendrick (Kermack and McKendrick 1927; Diekmann, Heesterbeek, and Britton 2013) formulated a relationship between the final epidemic size and \mathcal{R}_0 , from which the latter can be calculated.

Unlike the previous methods, which require the epidemic to reach a steady state, \mathcal{R}_0 may be determined from the intrinsic growth rate of the infected population (Heffernan, Smith, and Wahl 2005; Wallinga and Lipsitch 2007) using incidence data of the early stages of the epidemic, as long as the growth of new cases exhibits pure exponential behaviour. Alternatively, we can employ the entire report of daily case notifications if \mathcal{R}_0 is formulated as a function of a compartmental model’s parameters (Lloyd 2009; Vynnycky and White 2010; Brouwer 2022). These models can be stochastic (He, Ionides, and King 2010), semi-deterministic (Dureau, Kalogeropoulos, and Baguelin 2013; Andrade and Duggan 2022) or deterministic (Kermack and McKendrick 1927; Andrade and Duggan 2020).

These compartmental models are said to be *mechanistic* (Bretó et al. 2009), namely, structures based on a scientific understanding of infectious disease dynamics (Bretó 2018). The relevance of that mechanistic property lies in the role of the model. Rather than being a merely mathematical artefact to produce a desired output, the model also embeds a dynamic hypothesis of the underlying process that generates the observed data. Hence, the parameters, states and interactions that comprise a particular formulation represent their counterparts in the real world. If the model accurately captures the properties of the actual phenomenon, finding an adequate *configuration* (assign values to parameters) should yield a behaviour over time of infections that resembles the observed trajectory. The values of such parameters can be obtained from individual-level observations (Keeling and Rohani 2011) or via statistical inference (Bolker 2008; Andrade and Duggan 2021; Bjørnstad 2023), a process also known as *trajectory matching* or *model fitting*.

Furthermore, matching simulated and observed behaviour can be regarded as a validation test on the dynamic hypothesis that links structure to behaviour (Barlas 1996). Nevertheless, one should understand this validation step as a falsification test (Oliva 2003). That is, if the model fails to reproduce the observed behaviour, it can certainly be rejected. On the contrary, obtaining an accurate match (or fit) does not immediately validate the dynamic hypothesis inasmuch as there may be other competing hypotheses that fit the data equally well. Indeed, this circumstance impacts the estimation of \mathfrak{R}_0 from compartmental models (and the intrinsic growth rate method), where different assumptions can yield accurate fits (Wearing, Rohani, and Keeling 2005). However, estimates vary according to the specific assumptions embedded in each fitting model (Heffernan, Smith, and Wahl 2005; Gostic et al. 2020).

For instance, the choice of the distributions of the latent and infectious periods (epidemiological delays) in the deterministic Susceptible-Exposed-Infectious-Recovered (*SEIR*) framework plays an essential role in the inference of \mathfrak{R}_0 (Wearing, Rohani, and Keeling 2005). Briefly put, misspecifying the structure of such delays leads to biases in the estimates. That is, a systematic difference between true and estimated parameters. Although there are techniques (Hurtado and Kiro Singh 2019; Greenhalgh and Rozins 2021) to construct models with realistic distributions, modellers do not know exactly which distribution to incorporate in their formulation. In view of this drawback, Wearing and colleagues (Wearing, Rohani, and Keeling 2005) fitted various *SEIR* models (with different delay distributions) to a single incidence dataset to select the best structure based on a goodness-of-fit measure. Nevertheless, the results appear inconclusive. Notwithstanding that Krylova & Earn (Krylova and Earn 2013) assume their validity, no further research establishes the reliability of such an approach. This assessment immediately warrants the need for the work presented here: a systematic study oriented to determine whether it is possible to infer \mathfrak{R}_0 accurately from *SEIR* models fitted to incidence data in light of the uncertainty in the distributions of the epidemiological delays. We describe the steps of this study in the sections below. All the analysis is performed in R. The code is freely available at <https://github.com/jandraor/delays>.

3.2 Data Generating Process

3.2.1 The system (latent) component

In order to undertake a systematic study, experimenters must have access to a sizeable set of observations. In this case, multiple time series of daily case notifications of a particular disease under various conditions. Equally important, such conditions need to be known *a priori*. To meet these conditions, we leverage the mechanistic property of compartmental models and employ the *SEIR* framework as a synthetic data generator (Gelman et al. 2020; King et al. 2015). This framework has been widely applied to studying various infectious diseases, such as measles (Lloyd 2001a; Keeling and Grenfell 2002; Krylova and Earn 2013), COVID-19 (Davies et al. 2020; Gleeson et al. 2022; Andrade and Duggan 2022), and influenza (Chowell et al. 2006; Chowell, Nishiura, and Bettencourt 2007; Vynnycky and Edmunds 2008; Vynnycky and White 2010; Andrade and Duggan 2020, 2021). In this work, we restrict our attention to the simplest version of this family of models. The rationale for this decision is straightforward; conceptual models entail efficiency inasmuch as they facilitate the understanding and identification of the underlying causes of a particular result. Moreover, it is often the case that principles that stem from basic models apply to more elaborated extensions.

$$\begin{aligned}
 \dot{S}_t &= -\frac{\beta S_t \sum_{k=1}^j I_t^k}{N} \\
 \dot{E}_t^1 &= \frac{\beta S_t \sum_{k=1}^j I_t^k}{N} - i\sigma E_t^1 \\
 \dot{E}_t^2 &= i\sigma E_t^1 - i\sigma E_t^2 \\
 &\vdots \\
 \dot{E}_t^i &= i\sigma E_t^{i-1} - i\sigma E_t^i \\
 \dot{I}_t^1 &= i\sigma E_t^i - j\gamma I_t^1 \\
 \dot{I}_t^2 &= j\gamma I_t^1 - j\gamma I_t^2 \\
 &\vdots \\
 \dot{I}_t^j &= j\gamma I_t^{j-1} - j\gamma I_t^j \\
 \dot{R}_t &= j\gamma I_t^j
 \end{aligned} \tag{3.1}$$

Specifically, the *SEIR* (Eq (3.1)) stratifies individuals as susceptible (S_t), exposed (E_t), infectious (I_t), and recovered (R_t) and describes the transitions

between states ($S_t \rightarrow E_t \rightarrow I_t \rightarrow R_t$) in terms of differential equations. Susceptible individuals acquire infection, $S_t \rightarrow E_t$, through contact with infectious individuals, where the number of contacts is independent of the population size (N). Formally, one refers to this assumption as the frequency-dependent (or mass action) transmission: $\beta S_t I_t / N$. Here, β corresponds to the effective contact rate or transmission parameter. The movement of individuals from the class E_t to class R_t is modelled using a well-known mathematical procedure (D. Anderson and Watson 1980) to achieve realistic distributions (Bailey 1954; Sartwell 1995) of the time that individuals spend in states E_t and I_t , otherwise known as the latent and infectious periods, respectively. Such a procedure corresponds to the subdivision of a class into stages arranged in series. For instance, one can divide the exposed class into i stages. Newly infected individuals enter the first exposed stage, E_t^1 , pass through each in turn and become infectious upon leaving the i th stage (E_t^i). The progression between stages is assumed to occur at a constant per-capita rate ($i\sigma$), leading to an exponential waiting time with mean $\frac{1}{i\sigma}$ in each stage (Lloyd 2009). This formulation implies that the lapse between infection and becoming infectious is described by the sum of i independent exponential random variables with equal rates, a convolution resulting in a gamma-distributed random variable (Blitzstein and Hwang 2019). Therefore, the subdivision of the exposed class into various stages is equivalent to formulating the latent period in terms of a gamma distribution with mean σ^{-1} and shape i . Similarly, one can divide the infectious class into j stages to formulating a gamma-distributed infectious period.

$$\mathfrak{R}_0 = \beta\gamma^{-1} \tag{3.2}$$

Overall, we refer to Eq (3.1) as the SE^iI^jR framework. Notice that the standard $SEIR$ corresponds to the SE^1I^1R instance. Moreover, as the parameter i increases, the distribution becomes more closely centred on its mean (tighter), to the extent that if $i \rightarrow \infty$, the variance is removed. That is, in the limit, all individuals have the same latent period. An equivalent argument applies to the infectious period. No less important, as indicated by Lloyd (2009), irrespective of the values of i and j that the SE^iI^jR may take, the basic reproduction number depends exclusively on the transmission rate and the mean infectious period (Eq (3.2)). Furthermore, it is noteworthy to mention that subdividing a class is a mathematical device that allows the incorporation of additional distributions in a system of differential equations, and the number of stages may not correspond to biological features of the

3.2. Data Generating Process

infection process (Lloyd 2001a). Lastly, we assume that the disease leads to permanent immunity and that the outbreak’s time scale is much faster than the characteristic times for demographic processes (births and deaths), therefore their effects are not included. This last assumption implies that the population remains constant over the simulation period.

$$\begin{aligned} \dot{C}_t &= \rho i \sigma E_t^i \\ x_{t^*} &= C_{t^*+1} - C_{t^*}, t^* \in \mathbb{N}_0 \end{aligned} \tag{3.3}$$

Subsequently, we define the link between the SE^iI^jR and incidence data (Eq (3.3)). Based on the literature (Chowell et al. 2006; Dureau, Kalogeropoulos, and Baguelin 2013; Andrade and Duggan 2021, 2022), we posit that incidence (\dot{C}) is proportional to the rate at which individuals become infectious ($E_t^i \rightarrow I_t^1$). Such proportional effect or reporting rate (ρ) stems from the fact that individuals experience various degrees of symptom severity (Gamado, Streftaris, and Zachary 2014). In particular, individuals with low severity levels (asymptomatic and mild symptoms) may not seek health care attention, resulting in case reports that most likely miss a significant fraction of infected individuals. As opposed to the continuous nature of differential equation models, case notifications occur at discrete times. To reconcile this tension, we define the report of new cases (x_{t^*}) as the change in the total number of cases (C_t) in one-day intervals.

Furthermore, we tailor the synthetic data generator towards influenza given that this virus causes unpredictable but recurring pandemics that can have significant global consequences (WHO 2017). As a matter of fact, there have been four influenza pandemics over the past 100 years, including the H1N1 pandemic in 1918, with 50 estimated million deaths (Monto and Webster 2013). Adapting the SE^iI^jR framework to this choice involves the selection of plausible parameter values or ground truths (Talts et al. 2018). For simplicity, we restrict the synthetic data generator to eight instances: $i = \{1, 3\} \times j = \{1, 2, 3, 4\}$. These instances share constants σ , γ , β , ρ , and N , which are configured identically. In particular, we configure parameters σ and γ from the assumed values ($\frac{1}{2}$ for both) in the Cumberland case study (Vynnycky and White 2010; Andrade and Duggan 2021). Following this choice, we select a value of β that yields a basic reproduction number (2.5) within a plausible range (2-4) of pandemic influenza (Mills, Robins, and Lipsitch 2004). Regarding ρ , we choose a value (0.75) consistent with reported estimates in the literature (Vynnycky and White 2010; Andrade

and Duggan 2021). The remaining constant, N , has only a scaling effect, and any particular value (10,000 in this case) does not alter the model dynamics provided that $N = S_0 + E_0 + I_0 + R_0$, where $E_0 = \sum_{k=1}^i E_0^k$ and $I_0 = \sum_{k=1}^j I_0^k$. In relation to initial conditions, we assume that a patient zero triggers the outbreak of a novel influenza pathogen. In mathematical terms, $S_0 = N - 1$ and $I_0^1 = 1$. The remaining initial conditions of the within-host profile are set to zero.

Having delimited the SE^iI^jR framework and configured its instances, we run simulations (Fig 3.1) that illustrate the impact of the delay structure on the incidence dynamics. In agreement with the literature (Keeling and Rohani 2011; Wearing, Rohani, and Keeling 2005), note in Fig 3.1 that if we fix the latent period distribution (i) and vary that for the infectious period (j), incidence reports that stem from more tightly distributed infectious periods (larger j) reach the incidence peak earlier and end more abruptly. This difference in behaviour over time occurs despite the fact that these instances share identical \mathfrak{R}_0 and equal average latent and infectious periods. On the other hand, if we fix the infectious period (compare two lines of the same colour across panels), decreasing the latent period’s variance (increasing i from 1 to 3) produces the opposite effect. Namely, tighter latent period distributions (larger i) push forward the peak time and extend the outbreak’s duration.

3.2.2 Measurement component

$$Y_{t^*} \sim Nbin(x_{t^*}, \phi) \quad (3.4)$$

Borrowing terminology from the state-space literature (Arulampalam et al. 2002; Bretó et al. 2009), one can frame the output produced by the SE^iI^jR framework as predictions obtained from a system or *latent* component. In practice, though, continuous and smooth predictions from ODE models differ from noisy and discrete incidence reports collected by public health surveillance. Moreover, given that a system component is merely a partial representation of a more complex reality, some elements are necessarily omitted. Consequently, it is required to equip the data generating process with a structure that accounts for the discrepancies between model prediction and actual data. We refer to this structure as the measurement component. In epidemiology, one can formulate the measurement of new infections via the *Negative Binomial* distribution, considering that this function does not tie the observation mean to the variance, offering the flexibility to account for

3.2. Data Generating Process

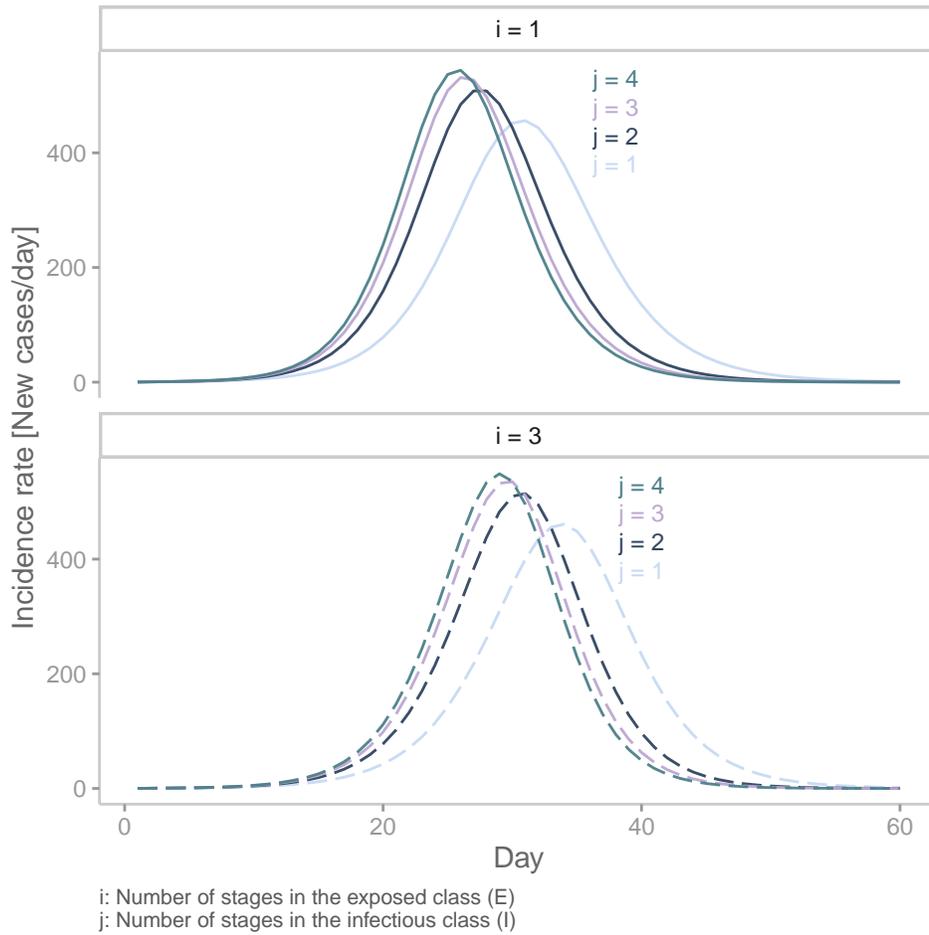


Figure 3.1: Incidence reports generated by various instances of the $SE^i I^j R$ framework. In this plot, we present two distributions of the latent period and four distributions of the infectious period. The colour of a line corresponds to a particular value of j (infectious period distribution). Solid lines indicate that the incidence report stems from an SEIR model with an exponentially-distributed latent period ($i = 1$). Dashed lines indicate that the incidence report stems from an SEIR model with a gamma-distributed latent period ($i = 3$).

overdispersion (Bretó 2018). Accordingly, we define the observation of new cases (Y_{t^*}) in terms of a Negative Binomial distribution (Eq (3.4)) specified by location (mean) and diffusion parameters. The former corresponds to the predicted incidence by the system component (x_{t^*}), whereas the latter (ϕ) modulates the concentration of measurements. Note that the inverse of the concentration parameter (ϕ^{-1}) represents overdispersion inasmuch as an increase in its magnitude leads to greater diffusion in the data.

Defining a measurement component completes the formulation of the data generating process. Consequently, we draw samples from Eq (3.4) using statistical simulation (*rnbinom in R*). For each SE^iI^jR instance, we generate 40 *noisy* time series. We perform this process for two levels (*high* and *low*) of data fidelity, a feature measured by ϕ^{-1} . High-fidelity data ($\phi^{-1} = 0$) implies that the measurement component applies only a slight distortion on the original signal (incidence). Notice that this configuration of the Negative Binomial (with no overdispersion) is equivalent to the *Poisson* distribution. Conversely, a positive value (overdispersion) of ϕ^{-1} (such as $1/3$) distorts the original signal to such an extent that one cannot easily discern the underlying incidence dynamics (low-fidelity data). We generated a total of 320 incidence reports, of which Fig 3.2 presents a sample of four representative reports (see Appendix B.1 for the complete set of synthetic data). To facilitate the communication of results, we introduce the notation D^{ij} , which indicates the origin of a given set of time series. For example, D^{14} indicates that the observed incidence was obtained from the SE^1I^4R instance.

3.3 Inference

The synthetic incidence reports described in the previous section allow us to assess the performance of various candidate models in recovering ground truths, particularly \mathfrak{R}_0 , our quantity of interest. Specifically, we fit model candidates to incidence data following a *Bayesian* approach (Gelman et al. 2013; Andrade and Duggan 2021). That is, each candidate’s unknown parameters are treated as random variables, which describe the knowledge (or uncertainty) about their actual values (McElreath 2020), expressed in terms of a probability distribution. This distribution is updated in light of new information summarised by a likelihood function. This function evaluates the compatibility between a given incidence report and multiple configurations of a model candidate (Lambert 2018). Such updating process yields the target or posterior distribution, an information device whereby we derive answers

3.3. Inference

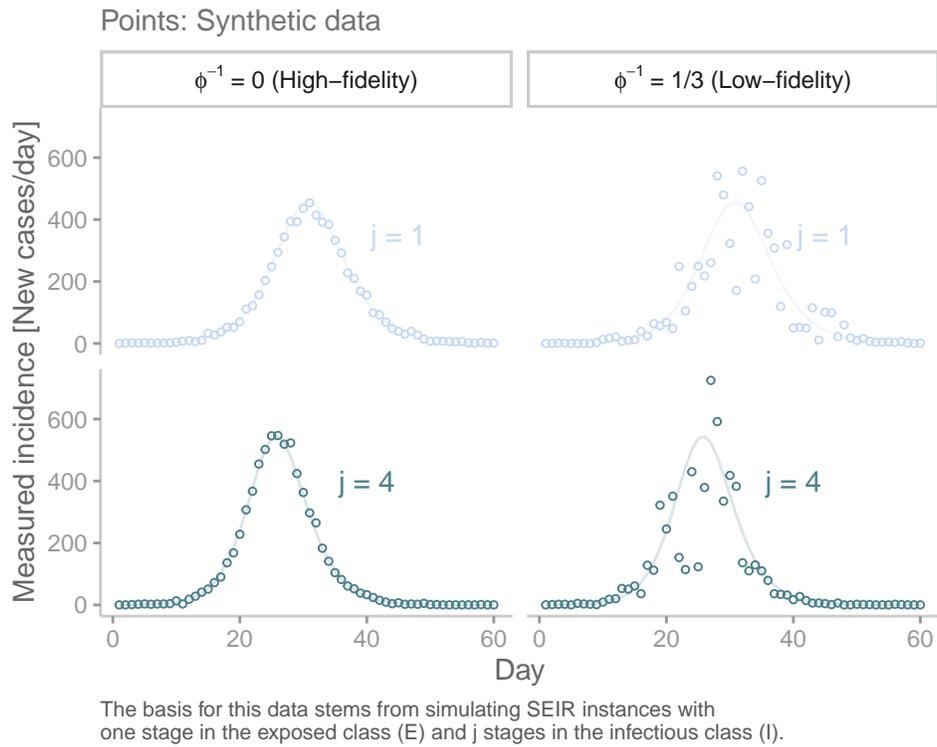


Figure 3.2: Sample of synthetic data. This plot shows four representative incidence reports (dots) obtained from the simulation of two SE^1I^jR instances (lines). To obtain each report, we sample from the negative binomial distribution.

for our inferential questions. We approximate the posterior distribution via sampling using Hamiltonian Monte Carlo or HMC (Betancourt 2018), an algorithm successfully employed to perform statistical inference from epidemiological models (Andrade and Duggan 2020, 2021, 2022; Chatzilena et al. 2019; Grinsztajn et al. 2021). This algorithm is provided by the statistical package *Stan* (Carpenter et al. 2017).

3.3.1 Three unknowns (traditional): β, ρ, I_0

For simplicity, we initially restrict the inference analysis to D^{1j} high-fidelity observations. To fit each incidence report, we postulate four instances, $j = \{1, 2, 3, 4\}$, from the SE^1I^jR framework, which share identical mean latent and infectious periods. We refer to the approach of fixing the means of the epidemiological delays to values obtained from the literature, regardless of their distribution, as the *traditional* parameterisation. Moreover, it is assumed that the measurement component is fully known. Consequently, discrepancies between estimated and actual values are ascribed to misspecification in the infectious period distribution. To avoid confusion between the origin of data and the fitting model, we denote the latter as M^{ij} . As a consequence, this design requires the estimation of 320 posterior distributions. Given this process’s computational burden, we limit the number of random variables in each model to three: the transmission rate (β), the reporting rate (ρ) and the initial number of infected individuals in stage one (I_0^1). The remaining parameters and initial conditions are considered to be known, i.e. they are fixed to their actual values. Based on this setup, we fit each candidate to a given dataset using HMC sampling, with *four* Markov chains and *1000* iterations (plus *1000* for warm-up) each, checking for convergence and effective sample sizes. The complete set of results can be found in the Appendix B.2.1.

The results presented in Fig 3.3 replicate a finding previously reported in the literature (Wearing, Rohani, and Keeling 2005; Lloyd 2001b): the existence of a subtle yet fundamental interaction between the assumed model structure and estimated \mathfrak{R}_0 . Misspecifying the infectious period distribution with a tighter distribution (higher j) generates lower \mathfrak{R}_0 estimates (Fig 3.3A). Furthermore, regardless of the assumed distribution of the infectious period, all candidate models fit the data equally well. To emphasise the importance and implications of this observation, we compare inferred and actual latent incidences in Fig 3.3B. Recall that fitting a candidate model to

a given incidence (y_t) produces a set of samples that describes the posterior distribution. Then, we use those samples to simulate the candidate’s system component, thereby generating inferred latent incidences (lines in Fig 3.3B). Then, those lines are compared to x_{t^*} , the true latent incidence (Fig 3.1). Notice that by definition, we do not have access to x_{t^*} in practical applications, but by virtue of this simulation study, such an impediment is overcome. The comparison reveals a symmetry shared among the candidate models. That is, any of these formulations can match the true latent incidence provided that β , ρ and I_0^1 are configured appropriately. It is important to remark that this symmetry is restricted to the latent incidence and does not extend to the dynamics of other states. For instance, candidates with different delay distributions that yield equivalent incidences will not reach the same long-term equilibrium, given the differences in their \mathfrak{R}_0 .

Logically, such symmetry should render the approach of comparing fit scores impractical. A fit score, such as the Maximum Likelihood Estimate (MLE), measures the consistency between a dataset and the output generated by a model. Since candidates produce equivalent output, differences among MLEs will solely reflect the stochasticity (noise) of the measurement component. We empirically verify this conjecture by selecting the candidate with the largest MLE for each incidence report (see Appendix B.2.1.1.6). We observe that M^{11} candidates attain the largest MLE in only 12 out of 20 times when matching D^{11} incidence reports. Even worse, M^{13} instances are always outperformed in fitting D^{13} datasets. Overall, no candidate passes the 60% mark. Similarly, the *mean absolute scaled error* (MASE), a metric specifically designed for evaluating the accuracy of time-series forecasts (Hyndman and Koehler 2006), indicates that candidates produce virtually identical scores when fitting any given incidence report. In light of this evidence, one can safely conclude that score comparison is not a reliable approach to determining the correct distribution of epidemiological delays from incidence data. To further complicate matters, information criteria (such as *AIC* and *BIC*) and cross-validation methods cannot assist in this task, considering that the evaluated structures produce equivalent output and share an equal number of unknown parameters.

3.3.2 Four unknowns: β, ρ, I_0, γ

The reason for such inherent symmetry is the *generation time*, the time between the infection of a primary case and one of its secondary cases

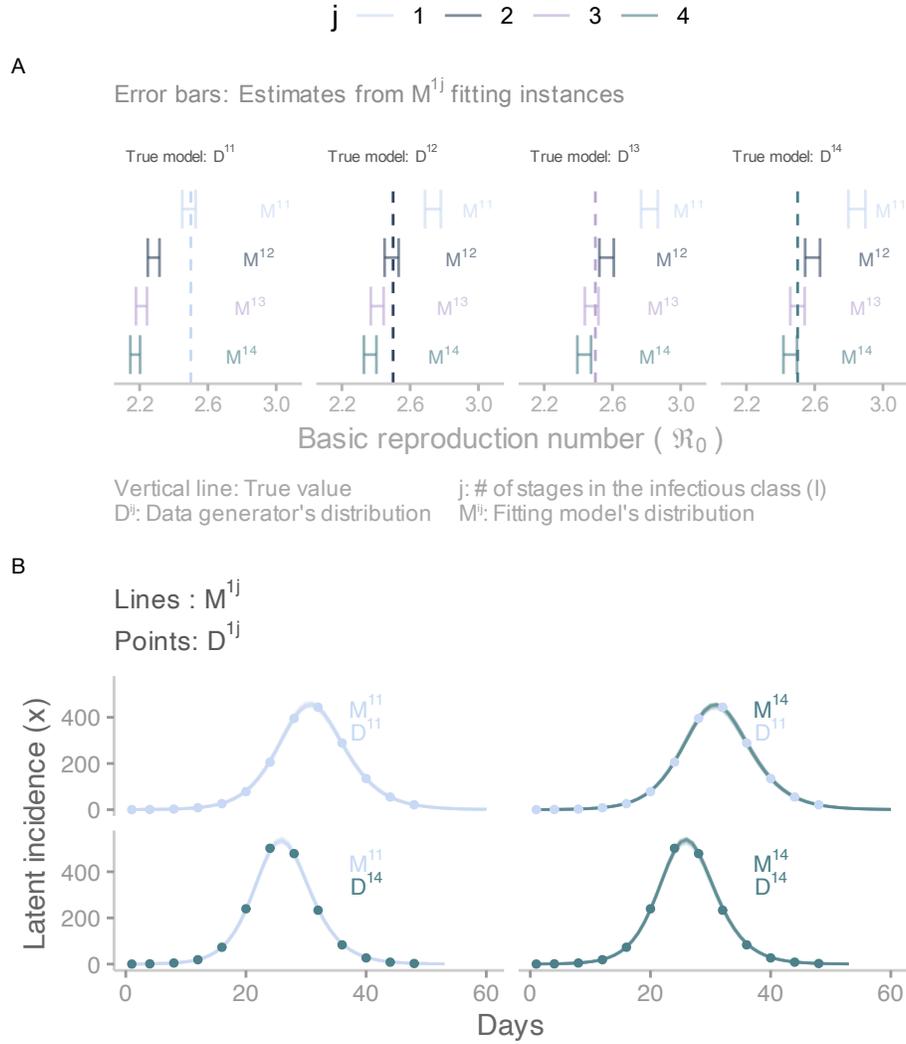


Figure 3.3: Inference results obtained from the three-unknown parameterisation. This plot shows the results of fitting model candidates to incidence reports. A) Comparison of estimates for the basic reproduction number obtained from fitting four candidate models to four incidence reports. Error bars correspond to 95% credible intervals, and the vertical line denotes the true value. B) Comparison between inferred incidence (lines) obtained from two candidate models fitted to two incidence reports (dots). Twenty time series represent inferred incidence. Given the high-fidelity data, all inferred incidences are nearly identical, giving the impression of only one line in each panel.

(Svensson 2007). This quantity’s shape, in tandem with \mathfrak{R}_0 , determines the initial dynamics of an infectious disease (Wallinga and Lipsitch 2007). Interestingly, these elements also characterise long-term behaviour. Krylova & Earn (Krylova and Earn 2013) found that *SEIR* models that account for demographic processes with different delay distributions produce equivalent dynamics of epidemiological transitions (e.g. from annual to biennial epidemic cycles) if they share identical \mathfrak{R}_0 and mean generation time (τ). An analytical expression for this last quantity can be obtained using the method described by Svensson (Svensson 2007). In particular, for the *SEⁱI^jR* framework, τ can be expressed (Eq (3.5)) as a function of the average delays (σ^{-1} , γ^{-1}) and the infectious period distribution (j).

$$\tau = \sigma^{-1} + \frac{j+1}{2j}\gamma^{-1} \tag{3.5}$$

In this analysis, we have, until now, fixed the mean generation time on each candidate model by excluding σ and γ from the inference process. Taking note of the effects on short and long-term dynamics that produce the interaction between τ and \mathfrak{R}_0 , we now promote γ to the category of *estimated parameter* in order to explore the impact of a variable mean generation time. The reason for choosing γ as the extra parameter is based on the fact that it interacts with both quantities of interest (Eq (3.2) and Eq (3.5)). This choice implies the need for estimating four parameters per model instance. To do so, we follow the approach described in the previous section. The reader can find the full set of results in the Appendix B.2.2. Unsurprisingly, given the extra degree of freedom, all candidates fit any of the incidence data equally well. In this design, though, the match between synthetic data and fitting model’s output is achieved at the expense of less precision, although greater accuracy. *Precision* refers to the width of uncertainty intervals, and *accuracy* to whether the interval captures the actual value. To illustrate this phenomenon, in Fig 3.4, we present the results of fitting four candidate models (M^{1j}) to four incidence reports that stem from different distributions of the infectious period (D^{1j}). Here, we see that the range of \mathfrak{R}_0 widened (Fig 3.4A) compared to that presented in the previous section (Fig 3.3A).

Undoubtedly, the primary insight from allowing γ to vary is the unravelled interaction between \mathfrak{R}_0 and τ . We visualise this interaction by plugging samples of β and γ into Eq (3.2) and Eq (3.5) to obtain an approximation of the expected values of \mathfrak{R}_0 and τ . When these two quantities are displayed on a scatter plot (Fig 3.4B), a linear relationship appears, regardless of the

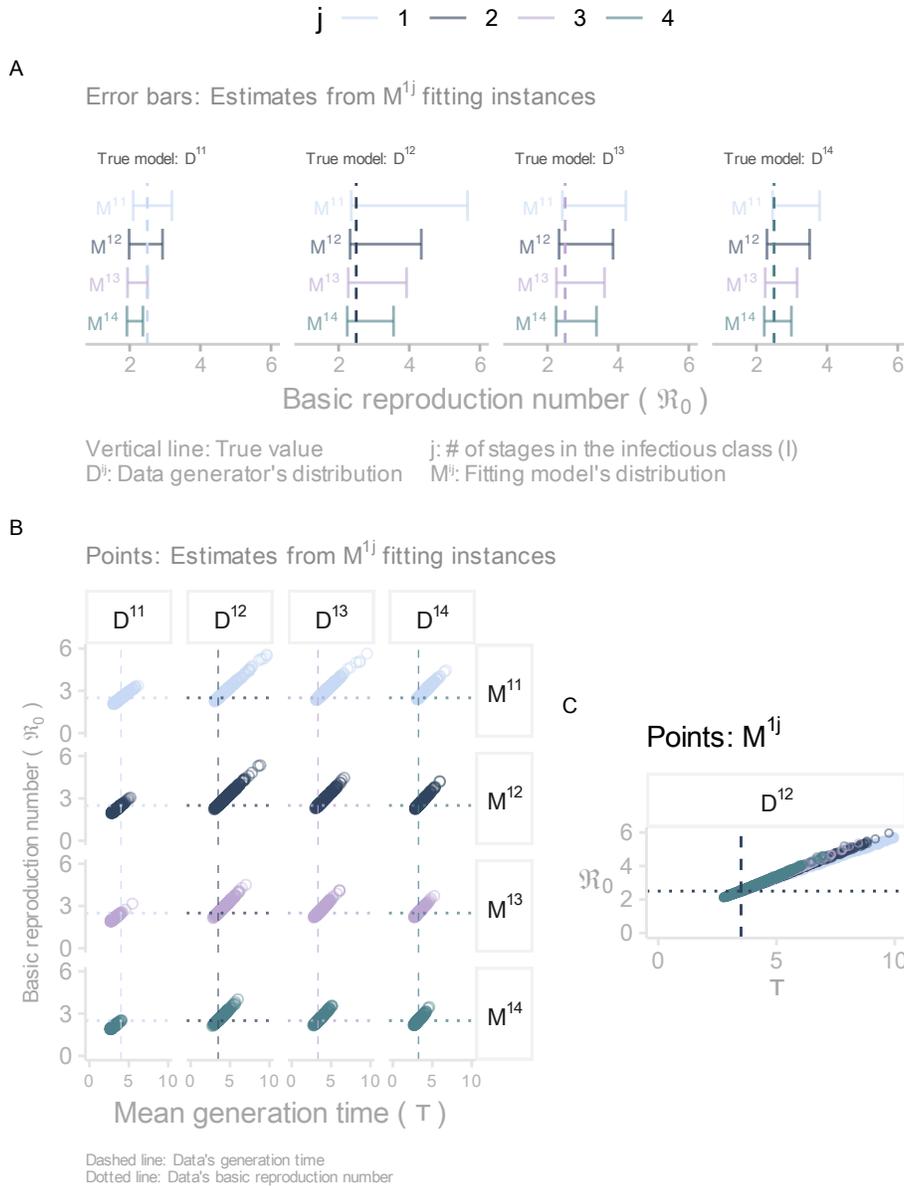


Figure 3.4: Four-unknown parameterisation. A) Comparison of \mathfrak{R}_0 estimates obtained from four candidates fitted to four incidence reports. Error bars correspond to 95% CrI, and the vertical line denotes the true value. B) Linear relationship between \mathfrak{R}_0 and τ estimated from posterior distributions (represented via samples) obtained from four candidates fitted to four incidence reports. For each sample, we compute the predicted \mathfrak{R}_0 and τ (dots). C) This plot collapses the second column in B into a single panel.

data's origin or the fitting model's structure. The interpretation of such linear association indicates that for a given fitting model, infinite pairs of \mathfrak{R}_0 and τ yield equivalent incidence dynamics. However, by virtue of their linear relationship, each value of τ corresponds to exactly one value of \mathfrak{R}_0 .

3.3.3 Three unknowns (alternative): $\mathfrak{R}_0^{-1}, \rho, I_0$

More importantly, the linear relationships shown in Fig 3.4B reveal an intriguing insight. Notice that irrespective of the structure (M^{1j}) fitting data of any origin (D^{1j}), the true values of \mathfrak{R}_0 and τ as a pair (the intersection between the dotted and dashed lines) are subsumed into any of the linear associations. This observation implies that the true \mathfrak{R}_0 can correspond only to the right τ . Therefore, it could be possible to accurately estimate \mathfrak{R}_0 from a model whose mean generation time is fixed to the true underlying value, but the shape of the epidemiological delays may differ from that of the data generating process. To test this hypothesis, we reformulate the SE^iI^jR framework so that τ becomes a parameter of every model instance. Consequently, we combine Eq (3.2) and Eq (3.5) into Eq (3.6), which expresses β as a dependent variable of four parameters: j, σ, τ , and \mathfrak{R}_0 .

$$\begin{aligned}\beta &= \frac{j+1}{\mathfrak{R}_0^{-1}2j(\tau - \sigma^{-1})} \\ \gamma &= \beta\mathfrak{R}_0^{-1}\end{aligned}\tag{3.6}$$

Parameter j is based on the fitting model's structure, whereas σ and τ are fixed to the true values that produced the incidence reports. For instance, a D^{12} report stems from a structure whose σ and τ are equal to 0.5 and 3.5 (applying Eq (3.5)), respectively. Therefore, an M^{14} candidate fitting this report has j, σ , and τ fixed to 4, 0.5, and 3.5, respectively. An immediate consequence of this procedure is the need to constrain γ in order to maintain logical consistency. Accordingly, we define γ as a function of β and \mathfrak{R}_0 (Eq (3.6)). This approach is analogous to fixing γ to an arbitrary value that yields the desired τ . Such a value may not correspond to that of the data generating process. Lastly, the remaining parameter, \mathfrak{R}_0 , is subject to inference. We opt to estimate its inverse for a practical reason. Taking into account the threshold phenomenon and the fact that all incidence reports exhibit outbreak-like behaviour, any estimated value of \mathfrak{R}_0 must fall within the interval $(1, \infty)$. It then logically follows that its inverse (\mathfrak{R}_0^{-1}) spans

over the range $(0, 1)$. This transformation permits the inference algorithm to operate in a much smaller parameter space, which enhances sampling efficiency.

We subsequently incorporate the redefined components (β and γ) into the SE^iI^jR framework to produce an alternative set of four candidate models with three unknowns: \mathfrak{R}_0^{-1} , ρ and I_0^1 . Similarly as before, we estimate the posterior distribution for each candidate fitted to an incidence report. The reader can find the complete set results in Appendix B.2.3. These results once more highlight the intrinsic symmetry of SEIR formulations. Specifically, provided there is an adequate configuration, any candidate structure can accurately match the observed incidence despite differences in the infectious period distribution. Nevertheless, this alternative parameterisation exhibits a distinctive and crucial feature: the estimation of \mathfrak{R}_0 is less sensitive to the assumed distribution of the infectious delay. To support this claim, we present in Fig 3.5 the results of fitting the four alternative candidates to four incidence reports of dissimilar origin. Here, it can be seen that all candidates recover (via 95% credible intervals) the underlying true \mathfrak{R}_0 , notwithstanding the origin of the data or the fitting model.

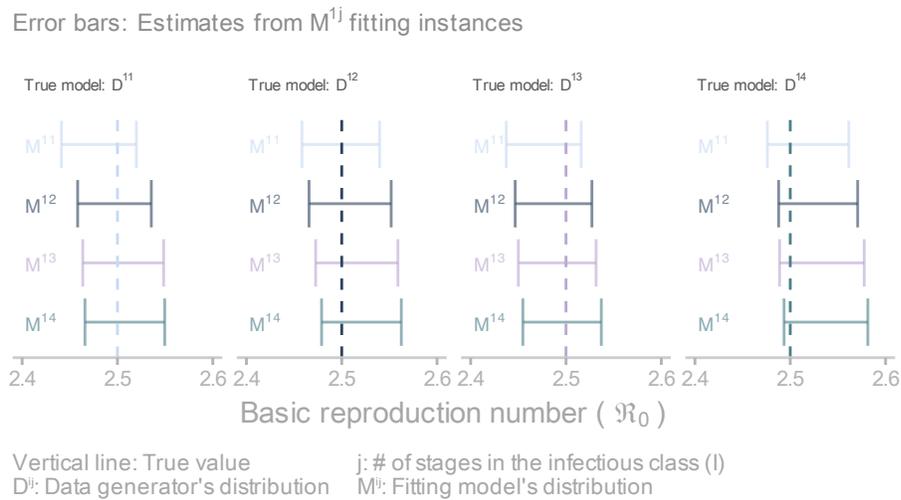


Figure 3.5: Inference results obtained from the three-unknown alternative parameterisation. This plot compares estimates for the basic reproduction number obtained from fitting four candidate models to four incidence reports. Error bars correspond to 95% credible intervals, and the vertical line denotes the true value.

Recovering the underlying \mathfrak{R}_0 is not exclusive to this sample of four datasets

but is generalised across the 80 high-fidelity D^{1j} datasets. To summarise this insight, we borrow a concept from the *frequentist* tradition. Such a concept known as *coverage* (Freedman, Pisani, and Purves 2007) means that if one collects a large number of samples from the same process and constructs the corresponding confidence intervals, then a certain percentage of the intervals will contain or cover the true parameter. This percentage is given by the confidence level. For instance, if one fits a model to 100 datasets and estimates an equal number of confidence intervals at the 95% significance level, then 95 of those intervals will cover the true value. Admittedly, it is implicitly assumed that our 95% credible intervals (obtained from posterior distributions) are proportional to 95% confidence intervals. Indeed, estimated intervals for \mathfrak{R}_0 and ρ conform to this concept (see Table B.3 in Appendix B), where minor deviance is justified by the fact that coverage is defined asymptotically (infinite measurements). However, asymptotics does not account for the large deviance observed in the estimates of I_0 . We explain this inconsistency in the section below where I_0 becomes more prominent.

To conclude this section, we report the analysis of the low-fidelity datasets (right column in Fig 3.2). The reader can find the results in Appendix B.3. Overall, we obtain similar insights in comparison to those derived from the high-fidelity datasets. In the absence of structural differences, it is unsurprising that the effect of larger noise in the signal (overdispersion) results in greater uncertainty in parameter estimates. This decrease in precision (wider credible intervals) can obscure or accentuate features of the inference process. On the one hand, overdispersion masks biases in estimates. For instance, noisier measurements cause I_0 estimates from the *alternative* parameterisation to conform to the expected coverage, which should not occur based on the results obtained from the high-fidelity datasets. On the other hand, overdispersion exacerbates identifiability issues. Under the *four-unknown* parameterisation, some \mathfrak{R}_0 estimates reach values up to 40. This result is a reminder that choosing an adequate number of unknowns is not a trivial decision. Setting more unknowns than the data can tolerate renders models unidentifiable. In this context, unidentifiability occurs because the incidence data does not provide enough information to update the prior distribution of γ . As discussed above, many values of γ are consistent with the observed incidence, an insight that holds for both levels of data fidelity. Finally, we note that overdispersion estimates are robust to the choice of the infectious period distribution.

3.3.4 Misspecifying the latent period distribution

Thus far, we have conducted the inference process assuming that the latent period distribution (i) is known. Lifting this constraint would strain our computational resources, producing a four-fold increase in the pool of candidates fitting a single report (assuming $i, j \in \{1, 2, 3, 4\}$). Instead of undertaking such costly exploration, one could leverage the fact that the mean generation time depends solely on the mean latent period rather than its particular distribution (Eq (3.5)). To test this idea, we compare the estimates obtained from candidate models with the *right* and *wrong* latent period distribution. We illustrate this process with the $80 D^{3j}$ low-fidelity ($\phi^{-1} = 1/3$) datasets. For each dataset, we fit eight candidates M^{ij} from the traditional three-unknown parameterisation, where the latent period distribution can take the wrong ($i = 1$) and the right ($i = 3$) values, and the infectious period distribution varies as before, namely, $j \in \{1, 2, 3, 4\}$. The reader can find the complete results in Appendix B.4.

To facilitate the presentation of the results, we first focus on candidates M^{13} and M^{33} fitting one D^{33} incidence report. Fig 3.6A shows that both models predict similar, although not identical, latent incidence dynamics. Further inspection reveals that the slight difference in the predicted incidence due to dissimilar latent period distributions does not lead to variation in \mathfrak{R}_0 estimates. To corroborate this assessment, we expand the analysis to the eight candidates matching the same incidence report. The right-hand side of Fig 3.6B shows that \mathfrak{R}_0 estimates are sensitive to variation in the structure of the infectious period but are indifferent to the latent period distribution. In compliance with the literature, the more dispersed latent period ($i = 1$) leads to an earlier incidence peak compared to the tighter distribution ($i = 3$) in the context of identical \mathfrak{R}_0 .

Nevertheless, the mechanism that enables models with heterogeneous distributions to produce analogous incidence dynamics remains unexplained. The left-hand side of Fig 3.6B, which displays I_0 estimates, provides the first hint. This plot shows that instances with the wrong latent period distribution (M^{1j}) systematically underestimate (via 95% credible intervals) the actual value (vertical line). To explain this phenomenon, we draw on a broader view of the posterior distribution. It is commonplace to restrict inference analyses to one parameter at a time (i.e. marginal distributions), neglecting the information provided by the full posterior distribution. To redress this shortcoming, we visualise the full distribution via pair plots. Specifically, Fig

3.3. Inference

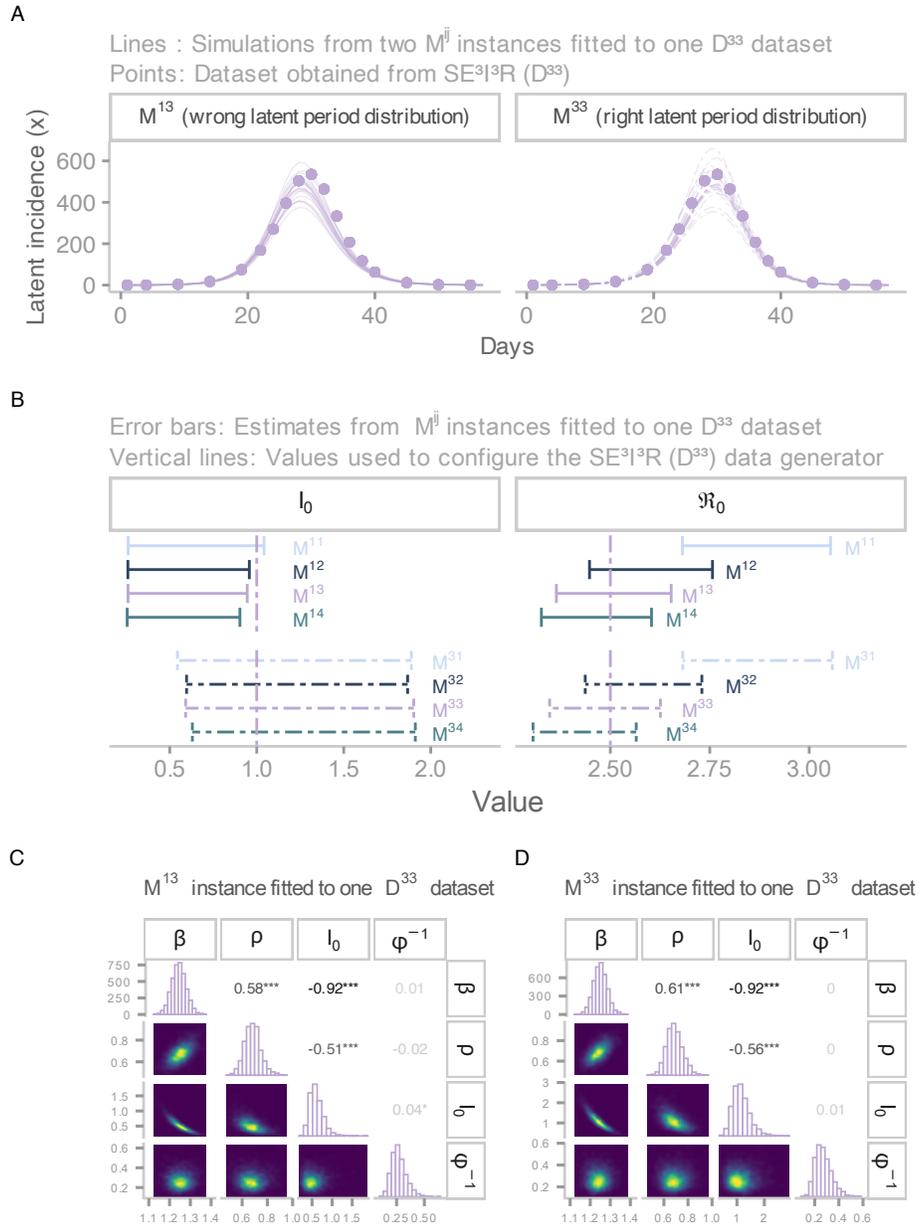


Figure 3.6: Comparing estimates from candidates (three-unknown parameterisation) with the wrong and correct latent period distributions. A) Incidence fit from two candidate models matching a D^{33} report. B) Comparison of I_0 and R_0 estimates by fitting model. Error bars correspond to 95% CrI, and vertical lines denote true values. C) Posterior distribution from an M^{13} candidate fitting a D^{33} report. The diagonal shows the marginal distributions, the lower triangular part represents each possible pairwise conditional distribution, and the upper triangular part represents the correlations among parameters. C) Posterior distribution from M^{33} candidate fitting a D^{33} report.

3.6C corresponds to the summary of the posterior distribution obtained from fitting M^{13} to one D^{33} incidence report. The upper triangular elements of this plot indicate that the three estimated parameters are strongly correlated. Especially β and I_0 , or more compellingly, \mathfrak{R}_0 and I_0 . Recall that the basic reproduction number is directly proportional to β . Therefore, although the mean generation time determines which \mathfrak{R}_0 corresponds to the observed incidence, I_0 (and ρ to a lesser extent) regulates the flexibility of \mathfrak{R}_0 to reach such a desired value. Interestingly, I_0 provides such a degree of flexibility that unrealistic adjustments in its estimates allow us to equate dissimilar model structures. Notice that the only discernible difference between Fig 3.6C and Fig 3.6D (M^{33} fitted to D^{33}) is seen in the marginal distributions of I_0 . In fact, this phenomenon explains the *failure* of the alternative parameterisation to recover the true value of I_0 .

In view of these symmetries, it is not unreasonable to expect that candidates from the four-unknown and the alternative parameterisations, too, are indifferent to the latent period distribution once I_0 and ρ correct for any misspecification. To verify this premise, we fit the parameterisations mentioned above to the D^{3j} low-fidelity incidence reports. As anticipated, the inference results indicate that the four-unknown parameterisation (Appendix B.4.3) uncover the linear association between τ and ρ due to the unidentifiability of γ . Likewise, the alternative parameterisation (Appendix B.4.4) recovers the true \mathfrak{R}_0 irrespective of the formulation of the epidemiological delays. Furthermore, we replicate these results using the D^{3j} high-fidelity datasets (see Appendix B.5).

3.3.5 Sensitivity analysis

So far, model candidates have been amalgamated with the appropriate measurement component. In this section, we explore the implications that can arise from ignoring overdispersion. That is, equipping model candidates with a Poisson measurement component. We perform such exploration by inferring \mathfrak{R}_0 from M^{1j} candidates (alternative parameterisation) fitted to the D^{3j} low-fidelity datasets discussed in the previous section. As expected, the results indicate that employing the Poisson distribution (see Appendix B.4.5) leads to overconfident (too precise) and biased (inaccurate) estimates in the context of overdispersion. We observe these features with narrow uncertainty intervals that do not cover the true value. This result implies that the wrong choice of the measurement component can offset any gains in accuracy due

3.3. Inference

Table 3.1: Scenarios

Scenario	\mathfrak{R}_0	τ_e	\mathfrak{R}_0 recovered? (Hf)	\mathfrak{R}_0 recovered? (Lf)
1	2.5	4	Yes	Yes
2	2.5	8	Yes	Yes
3	2.5	13	Yes	Yes
4	9.0	4	No	Yes
5	15.0	4	No	Yes

Note: Hf: High-fidelity. Lf: Low-fidelity

to the alternative parameterisation.

On the other hand, the synthetic data used for the analysis presented in the previous sections stems from models configured to identical \mathfrak{R}_0 and similar mean generation times (variation due to the infectious period distribution). Naturally, one wonders whether the usefulness of the alternative parameterisation holds in other conditions. To answer this question, we repeat the workflow described in this chapter for additional scenarios of τ and \mathfrak{R}_0 . For simplicity, we restrict this sensitivity analysis to datasets derived from models with an exponentially-distributed latent period (D^{1j}). Additionally, we equip the fitting candidates with the appropriate measurement component. The complete set of results is presented in Appendix B.6. We present these results in terms of scenarios (Table 3.1). For instance, the base case scenario, *Scenario 1*, corresponds to data generated from SE^iI^jR configured to $\mathfrak{R}_0 = 2.5$ and $\tau_e = 4$ (results presented in Section 3.3), where τ_e serves as a scenario identifier and denotes the mean generation time obtained from an exponentially-distributed infectious period ($j = 1$).

For *Scenario 2*, we increase the reference mean generation time ($\tau_e = 8$), while keeping \mathfrak{R}_0 at 2.5. First, we focus on the high-fidelity datasets. Overall, the greater the divergence between the fitting model’s infectious period distribution and the distribution that generated the data, the greater the loss in accuracy (lower coverage). To provide an example, the 95% credible intervals constructed from M^{14} candidates fitting D^{11} incidence reports only attain coverage of 30% for \mathfrak{R}_0 . Closer inspection, though, reveals that such accuracy loss is more statistical than practical. To support this statement, we calculate the average relative difference between the actual and estimated \mathfrak{R}_0 , finding that misspecification of the infectious period distribution leads to a maximum average relative error of 2%. In contrast, we would obtain

discrepancies up to 15 % if we adopted the traditional approach. Simply put, it is costlier to misspecify the mean generation time than the mean infectious period. Furthermore, such slight differences in the alternative parameterisation are erased by overdispersion. That is, overdispersion masks minor misspecification in the process component. Moreover, in *Scenario 3* ($\mathfrak{R}_0 = 2.5$, $\tau_e = 13$), we observe that further increasing of the mean generation time does not lead to significant drops in the coverage of \mathfrak{R}_0 under both levels of data fidelity. In a nutshell, it is reasonable to suggest that the alternative parameterisation is robust to various levels of the mean generation time.

Conversely, we cannot maintain the same assertion for various values of \mathfrak{R}_0 . Indeed, Fig 3.4C provided the first hint. This plot shows that the straight lines do not overlap as \mathfrak{R}_0 reaches relatively high values. Consequently, in scenarios 4 ($\mathfrak{R}_0 = 9$, $\tau_e = 4$) and 5 ($\mathfrak{R}_0 = 17$, $\tau_e = 4$), we test the implications of larger transmissibility levels. The results indicate that as we increase the underlying \mathfrak{R}_0 for generating the data, the equivalency among fitting models dissipates and misspecification in the infectious period distribution leads to biased estimates of \mathfrak{R}_0 . The size of such bias is proportional to the misspecification of the infectious period and the underlying \mathfrak{R}_0 . This feature is primarily seen in the estimates derived from high-fidelity datasets, where coverage levels are low, and the average relative error between actual and estimated values cannot be overlooked. However, when we examine the posterior distributions obtained from fitting the low-fidelity data, it is seen that, once again, overdispersion masks misspecification in the process component, as evidenced by the high coverage levels. This is not to say that overdispersion is a desired feature in the data, but rather to emphasise that its presence hinders the attainment of precise estimates. Undoubtedly, having this understanding is of practical importance, given that it allows us to discern the necessary effort in data collection and model improvement.

3.4 Application to Influenza A

Leveraging the knowledge gained from the synthetic data, the last step in this work consists of exploiting the relationship between the basic reproduction number and the mean generation time to update the \mathfrak{R}_0 estimate of an outbreak of the 1918 influenza pandemic. The reader can find the full set of results in Appendix B.7. In particular, we focus on an outbreak that occurred in the city of Cumberland (Maryland) during the autumn of 1918, for which the U.S. Public Health Service organised special surveys

(Frost and Sydenstricker 1919) to determine the proportion of the population infected. Previous studies (Vynnycky and White 2010; Andrade and Duggan 2021) employed the default heuristic of adopting an *SEIR* with exponentially-distributed epidemiological delays whose means were configured to values reported in the literature. Moreover, in these studies, the *SEIR* was coupled with the Poisson distribution resulting in a 95% CI [2.5–2.6] for \mathfrak{R}_0 . However, adopting a more realistic measurement component, such as the Negative Binomial distribution, produces lower and wider estimates: 95% CI [2.2, 2.4]. Further, if we jettison the assumption of an exponentially-distributed infectious period for a more realistic distribution, such as the gamma distribution, we obtain even lower estimates. For instance, a gamma-distributed infectious period with four stages (*SEI⁴R*) returns a 95% CI of [2.0, 2.2]. As noted earlier, the estimates obtained from this default heuristic or traditional approach are sensitive to the uncertainty in the infectious period distribution. On the contrary, when we fix the mean generation time in the *SEI^jR* (alternative parameterisation) to a value (2.85 days) obtained from the literature (Wallinga and Lipsitch 2007; Hirotsu et al. 2004), we derive nearly identical \mathfrak{R}_0 estimates (95% CI [2.0, 2.1]) regardless of the infectious period distribution (Fig 3.7). Notice that this estimate is similar to that obtained from the *SEI⁴R*, bolstering the fact that the actual infectious period is far from being exponentially distributed.

3.5 Conclusion

The misspecification of various assumptions within the *SEIR* framework can negatively impact the estimation of \mathfrak{R}_0 . In recognition of this risk, we ran a simulation study comprised of approximately 1000 synthetic datasets and 8000 model fits, whereby we identified the relative influence of some of those assumptions. Specifically, we found that fixing the mean generation time to a reliable estimate is of paramount importance. In contrast, one can be more lenient on the specification of the latent period distribution and the mean infectious period provided that other estimated parameters (I_0 and ρ) redress the misspecification. We leveraged this knowledge to formulate an alternative parameterisation that is more robust to the uncertainty of the epidemiological delays. However, there is a caveat with this alternative formulation. Although it exploits a local symmetry (incidence dynamics) of the *SEIR* framework, such symmetry does not extend to the other states of the system. Therefore, the usefulness of the alternative parameterisation

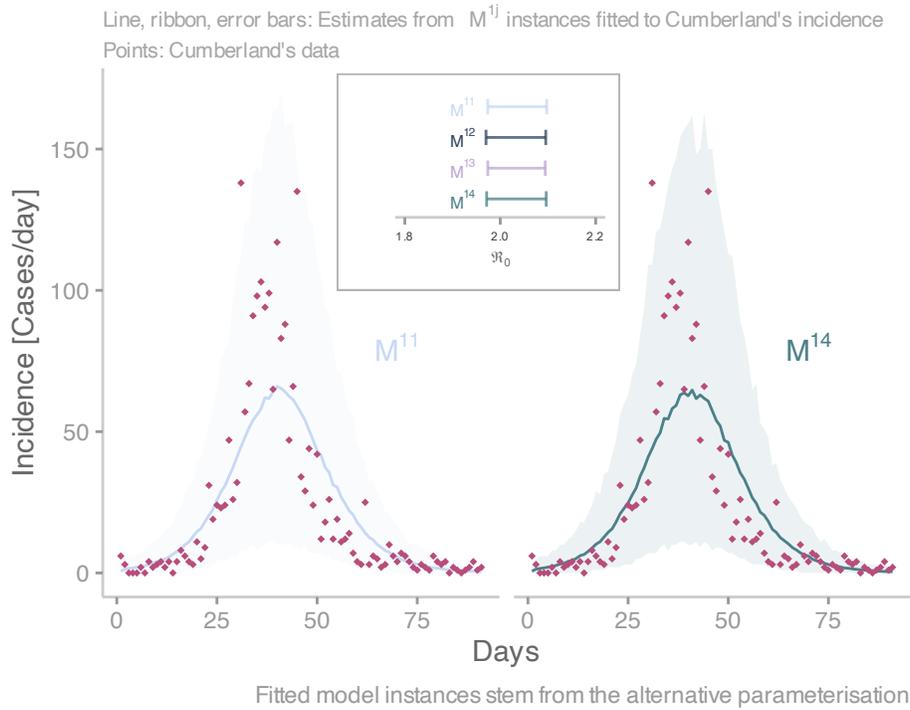


Figure 3.7: Application of the three-unknown alternative parameterisation. This plot shows the estimates obtained from fitting four candidates (M^{1j}) to the daily number of influenza cases (rhombi) detected by the U.S. Public Health Service in Cumberland (Maryland) during the 1918 influenza pandemic, from 22 September 1918 to 30 November 1918. Ribbons correspond to 95% credible intervals of the predicted reported cases by two candidate models. The solid line denotes the median. The box inside the plot shows the estimates for the basic reproduction number by fitting model. Error bars correspond to 95% credible intervals.

3.5. Conclusion

is confined to the estimation of \mathfrak{R}_0 , and it is not a substitute for other kinds of analyses. For instance, if, on the contrary, our variable of interest were I_0 , we would obtain unreliable estimates. Furthermore, the alternative *SEIR* with exponentially-distributed delays will be as overoptimistic as its traditional counterpart in predicting the critical vaccination proportion or the effectiveness of an imperfect VIH treatment in the context of within-host dynamics (Lloyd 2001b). Therefore, the alternative parameterisation is a mitigation strategy in the absence of complete information. Furthermore, its usefulness is abated by highly transmissible pathogens (Section 3.5). Nevertheless, biases in the estimates due to large \mathfrak{R}_0 are only detected with high-fidelity data. That is, data with little or no overdispersion.

Despite the significant computational effort of simulation analyses, a single study cannot offer overarching statements. Further work is required to test the validity of these insights in stricter or more elaborated contexts. For instance, we assumed the complete availability of the incidence time-series throughout this study. This assumption restricts the validity of the approach to retrospective analyses. However, other situations exist where only fewer incidence measurements are available to the modellers, such as the early phase of a pandemic response. Hence, it remains to be seen the effect of various levels of data availability on the performance of the suggested approach. Furthermore, for simplicity, we ignored age-related effects in the dynamics of the infectious disease as well as process stochasticity (demographic and environmental) and time-varying contact rates. We expect that future research builds on the findings provided by this study and addresses the aforementioned challenges to construct ever more reliable inference approaches.

Chapter 4

Time-varying transmission rate

The work presented in this chapter was published in:

Andrade, J., Duggan, J., 2022. Inferring the effective reproductive number from deterministic and semi-deterministic compartmental models using incidence and mobility data. *PLOS Computational Biology*, 18(6).

4.1 Introduction

Since early 2020, SARS coronavirus 2 (SARS-CoV-2) has spread throughout the seven continents, causing a COVID-19 pandemic of catastrophic consequences, including the loss of millions of lives and jobs. In the early days of the pandemic, given the absence of vaccines and the lack of effective therapeutics, governments primarily relied on non-pharmaceutical interventions (NPIs) to reduce the transmission of SARS-CoV-2, thereby lowering the death toll. Although effective in preventing deaths ([Flaxman et al. 2020](#)), NPIs such as mobility restrictions and stay-at-home orders impose a burden on society with economic and psychological costs ([Douglas et al. 2020](#)). In addition to this, the effectiveness of these interventions wanes over time as compliance progressively diminishes. Following these considerations, policymakers strive to find an adequate balance between the interventions' severity and acceptable transmission levels. In this decision-making process, the effective reproduction number plays a crucial role. Briefly, the effective reproduction number, \mathfrak{R}_t , is the time-varying average number of secondary cases caused by a primary case at a calendar time t ([R. Anderson and May 1992](#); [Nishiura and Chowell 2009](#)), and it is a theoretical indicator of the

course of an infectious process (Vynnycky and White 2010). Above the *epidemic threshold* ($\mathfrak{R}_t > 1$), each infectious person leads to more than one secondary infectious person, and the disease is (re)emerging (Brett et al. 2020); below that threshold, there is limited secondary transmission. In practice, policymakers can use \mathfrak{R}_t in two ways. First, as a guide to assess in near real-time whether the interventions are succeeding ($\mathfrak{R}_t < 1$) or whether it is required to increment the response’s strength (Nishiura and Chowell 2009). Second, in retrospective analyses to assess how policy decisions, population immunity, and other factors have impacted transmission at specific points in time (Gostic et al. 2020).

Generally speaking, \mathfrak{R}_t is the result of a combination of intrinsic (decline in susceptible individuals) and extrinsic (change in contact patterns due to the implementation of control measures) factors (Nishiura and Chowell 2009), for which there are no readily available measurements. One, therefore, must resort to statistical methods to obtain an approximation of this epidemic indicator. On one end of the spectrum, we find widely applicable and context-independent empirical methods such as the two-step Bayesian procedure proposed by Cori and colleagues (Cori et al. 2013; Thompson et al. 2019) and the likelihood-based estimation procedure proposed by Wallinga and Teunis (Wallinga and Teunis 2004). At the other end of the spectrum, we can infer \mathfrak{R}_t from compartmental models calibrated to incidence data (Andrade and Duggan 2020), which is the focus of this chapter. In addition to serving as vehicles to obtain estimates, these mechanistic models are based on a scientific understanding of infectious disease dynamics (Bretó 2018), which one can interpret as a dynamic hypothesis of the underlying process that produces the observable behaviour patterns. This feature implies that fitting a compartmental model to data also tests a hypothesis that links structure to behaviour (Oliva 2003). It thus follows that parameter estimates derived from this procedure have an interpretation in the real world. Notwithstanding these advantages, \mathfrak{R}_t estimates from compartmental models are sensitive to data availability and assumptions in the model structure (Gostic et al. 2020).

One such assumption is the transmission rate’s dynamics. In the context of the COVID-19 pandemic, the assumption of a constant transmission rate is rendered unrealistic, apart from a few days in the initial phase of the outbreak (Dehning et al. 2020; Davies et al. 2020). The rationale is that under the imminent surge of cases, governments implemented NPIs at early stages to reduce the number of contacts among the population. Modellers thus are required to describe formally the changes in the transmission rate over

time. For instance, in measles studies (Bretó et al. 2009; He, Ionides, and King 2010; Keeling, Rohani, and Grenfell 2001), it is not unusual to assume *term-time forcing* structures (X. Liu and Stechlinski 2012), where the contact rate experiences sudden changes in time (e.g., because of school holidays). Other approaches have adopted *smoothly-varying functions* (X. Liu and Stechlinski 2012) to model the transmission rate in tuberculosis outbreaks (L. Liu, Zhao, and Zhou 2010). In COVID-19 analyses, the transmission rate has been described as episodes of constant contact rates separated by change points where a transition occurs (Dehning et al. 2020; Linka, Peirlinck, and Kuhl 2020; Duggan et al. 2024). These are likely once-off models, more appropriate for retrospective analyses, whose formulations are not designed to incorporate new data that account for policy changes (unless the structure is modified).

Nevertheless, ascertaining which deterministic formulation is the most adequate is far from straightforward. Its search involves several *trial-and-error* iterations and model comparisons until a satisfactory structure is found. If one aims for near real-time estimates, random-walk formulations offer a flexible device to *uncover* the underlying transmission rate dynamics (Endo, van Leeuwen, and Baguelin 2019). This type of structure does not impose stringent constraints on the transmission's rate shape, facilitating the incorporation of new data without structural modifications. This approach has been applied to studying an influenza pandemic (Dureau, Kalogeropoulos, and Baguelin 2013; Endo, van Leeuwen, and Baguelin 2019) and Ebola outbreaks (Funk et al. 2018; Camacho et al. 2015). Although random-walk models yield fits to incidence data, the match between observed and simulated data may be achieved at the expense of large uncertainty bounds. Moreover, under this framework, the inference of time-independent parameters requires burdensome computational efforts. More recently, the extensive research provoked by the COVID-19 pandemic prompted researchers to use non-traditional sources of data to infer the transmission rate. In particular, mobility data has been assumed as a proxy for the changes in the transmission rate (Davies et al. 2021). In doing so, the dynamics exhibited by the transmission rate have an inherently plausible explanation (changes in human behaviour measured by mobile devices) so that models can more easily incorporate new incidence measurements. However, it should be mentioned that this approach entails a stringent assumption wherein one tacitly assumes a perfect correlation between changes in mobility data and effective contact patterns. Thus, discrepancies between actual and assumed transmission rates may result in

unnecessary corrections to the estimates of other unknown parameters.

Consequently, this chapter aims to draw upon the strengths of the approaches described above to formulate a complementary process for estimating \mathfrak{R}_t from compartmental models. Specifically, we build three structures or Data Generating Processes (DGP) that accounts for Ireland’s first COVID-19 wave. Two DGPs incorporate stochastic features in the transmission rate, whereas the other formulation is exclusively deterministic. These structures are complementary in the sense that the results obtained from one DGP inform the subsequent one. Below, we describe each DGP in detail, the inference process to obtain estimates for \mathfrak{R}_t and other unknown quantities (Fig 4.1), and finally, discuss the results. All the analysis is performed in R, mainly supported by the statistical packages *pomp* (King, Nguyen, and Ionides 2016) and *Stan* (Carpenter et al. 2017). The code is freely available at https://github.com/jandraor/time_varying_beta/tree/thesis.

4.2 Results

4.2.1 Context

By the end of February 2020, more than sixty countries had detected at least one case of COVID-19 (Ritchie et al. 2020), including Ireland, where the first confirmed case was announced on *the 29th of February*. Twelve days after this event, the Irish Government ordered the closure of all schools, colleges, and childcare facilities, followed by a stricter *stay-at-home* mandate implemented on *the 27th of March*. These interventions resulted in low incidence and mortality rates, which allowed easing the restrictions from mid-May. In Fig 4.2A and 4.2B, respectively, we present the number of daily (y_d^1) and weekly cases (y_w^1) detected from the first report up to the point where the restrictions began to be lifted, a period that we refer to as the *first wave*.

In a nutshell, stay-at-home orders and similar measures aim to restrict the movements of a population so that the risk of exposure to a transmissible pathogen is reduced. Impractical several years ago, the advent of smartphones has permitted us to gauge patterns in population mobility in real-time. For instance, since *the 13th of January, 2020*, Apple has provided an index that quantifies the level of mobility by transportation type (driving, transit, and walking). Apple generates this data by counting the number of requests made to *Apple Maps* for directions. Fig 4.2C shows Ireland’s daily driving mobility levels during the first wave (y_d^2), and Fig 4.2D, the value at the

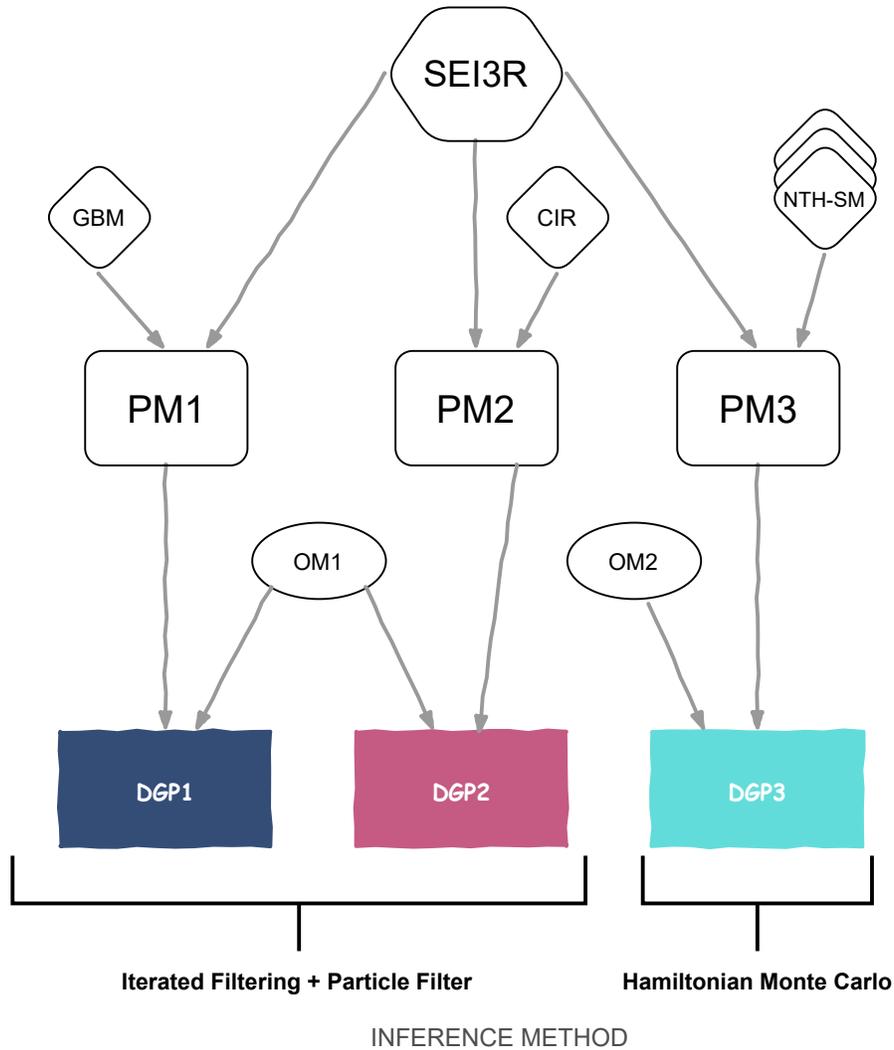


Figure 4.1: Schematic diagram of the data generating processes (DGPs) explored in this chapter. This diagram aims to portray the DGPs as the ensemble of two components: a measurement or observational model (ellipse) and a process model (rounded rectangle). For instance, DGP1 is the amalgamation of the measurement model OM1 and the process model PM1. The process model is in turn the ensemble of two structures: a within-host profile (hexagon) and a time-dependent transmission rate (rhombus). Whereas all process models share a common within-host profile (SEI3R), they differ in the formulation of the transmission rate: Geometric Brownian Motion (GBM), Cox-Ingersoll-Ross (CIR), and nth-order exponential smoothing (NTH-SM). The inference method employed on each DGP depends upon the nature of the process model (Iterated Filtering + Particle Filter for stochastic structures and Hamiltonian Monte Carlo for deterministic ones).

4.2. Results

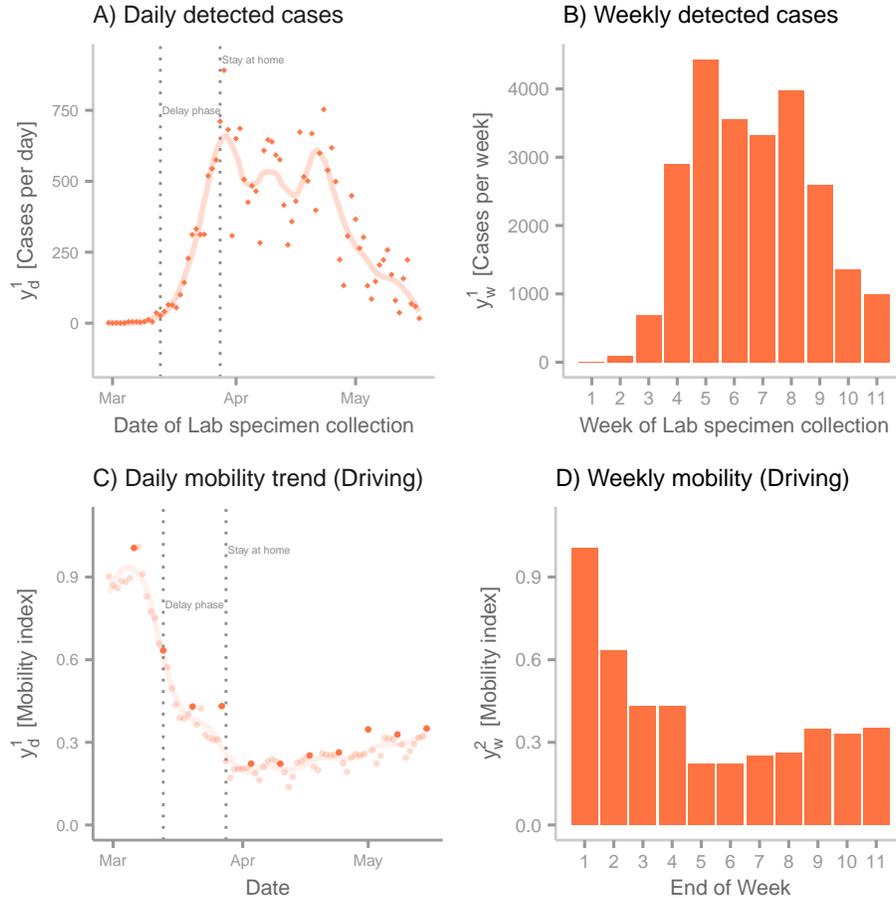


Figure 4.2: Incidence and mobility data. (A) Daily number (rhombus-shaped points) of COVID-19 cases detected during Ireland’s first wave, from the 29th of February 2020 to the 17th of May 2020. The x-axis indicates the date in which the infected individuals were swabbed. The line represents the smoothed trend (via LOESS method) from the data (B) Weekly number of COVID-19 cases detected in during Ireland’s first wave. The x-axis indicates the number of weeks since the first case was detected. (C) Apple data for Ireland from the 29th of February 2020 to the 17th of May 2020. Points represent the normalised amount of daily requests for driving directions. These indexes are normalised to the value on the 28th of February 2020. We highlight points every 7 days. These highlighted points are used to calibrate DGP1 and DGP2. The line represents the smoothed trend (via LOESS method) from the data. (D) Normalised amount of daily requests for driving directions at the end of each week starting from the 29th of February 2020. These bars correspond to the highlighted points in C.

end of each week (y_w^2) from *the 29th of February 2020*. This dataset, along with the incidence reports, will serve as the basis to calibrate the proposed compartmental models below.

4.3 State-Space Models

$$X_t \sim p_{X,t}^\theta(x_t|x_{t-1}) \quad (4.1)$$

$$Y_t \sim p_{Y,t}^\theta(y_t|x_t) \quad (4.2)$$

One can frame the inference process for compartmental structures following the terminology provided by state-space models (SSM) (Chopin and Papaspiliopoulos 2020), also known as Partially observed Markov process models (Ionides, Bretó, and King 2006). Through an *SSM*, one conceives a DGP as a generative probabilistic model that consists of two discrete-time Markovian mechanisms. The first mechanism (Eq (4.1)) describes the evolution over time of the system’s latent states (X), where X_t is drawn conditionally on the previous state of the latent process (X_{t-1}) according to the density $p_{X,t}^\theta(x_t|x_{t-1})$. Therefore, the DGP is a Markov chain (Blitzstein and Hwang 2019), as the state of the latent variable at time t depends only on its previous state and the distribution from which it comes. In the literature, Eq (4.1) is often referred to as the *latent process model* (Bretó et al. 2009) or the *system model* (Arulampalam et al. 2002). Intuitively, this formulation corresponds to the set of causal assumptions (*dynamic hypothesis*) that explains a phenomenon of interest in terms of states and transitions (rates). The process model may be defined in continuous or discrete time (Ionides, Bretó, and King 2006), but only its distribution at discrete times is considered ($X_t, X_{t+1}, X_{t+2}, \dots, X_{t+h}$), where $t \geq 1$ and h is an integer. For simplicity, we assume that X_0 is known.

In epidemiology, it is commonplace to represent the process model via compartmental structures in which individuals are categorised according to their infection status (Keeling and Rohani 2011). We refer to this categorisation as the *within-host profile*. Formally, one can employ a system of differential equations to build such compartmental models. The reader should recall that any system of ordinary differential equations $\frac{dx}{dt} = f(x)$ is Markovian. Here, we adopt the *SEI3R* profile (Davies et al. 2020; Gleeson et al. 2022), an extension of the SEIR framework. Under this profile, we stratify individuals

as susceptible (S_t), exposed (E_t), infectious, and recovered (R_t). We further disaggregate the infectious class by medical status, resulting in three compartments: *preclinical* (P_t), *clinical* (I_t), and *subclinical* (A_t) (see *Materials and Methods* section for the complete description). The three DGPs presented in this chapter share the SEI3R profile (Fig 4.1).

On the other hand, the exact state of the population at any given time is generally not observable and must be inferred from available data via statistical inference (Rasmussen, Ratmann, and Koelle 2011). It is thus necessary to formally relate (Eq (4.2)), at each discrete time ($t \geq 0$), latent states to *noisy* measurements via a *measurement* or *observational* model (Arulampalam et al. 2002), where each Y_t is drawn conditionally on the most recent state of the latent variable, according to the density $p_{Y_t}^\theta(y_t|x_t)$. This work draws on incidence and mobility data to formulate such measurement models.

4.3.1 DGP1 - Geometric Brownian Motion

$$\beta_t = \zeta Z_t \tag{4.3}$$

$$\frac{dZ}{dt} = \alpha Z_t dW \tag{4.4}$$

$$dW \sim Normal(0, \sqrt{dt}) \tag{4.5}$$

Thus far, we have not yet defined the time-varying effective contact rate or transmission rate (β_t). When defined, this component is integrated with the SEI3R profile to form a process model (Fig 4.1). For this and the other two DGPs, we formulate β_t as the product of two components (Eq (4.3)). Here, ζ denotes the transmission rate's initial value. Namely, $\beta_0 = \zeta$. From this definition, it follows that Z_t represents the transmissions rate's change over time relative to its initial value, where $Z_0 = 1$. In relation to Z_t dynamics, we initially opt for a flexible approach to build this first process model (PM1). Specifically, we define $\frac{dZ}{dt}$ in terms of Geometric Brownian Motion (GBM) with no *drift* (Eq (4.4)-(4.5)), an approach adopted in previous studies of influenza and Ebola (Dureau, Kalogeropoulos, and Baguelin 2013; Endo, van Leeuwen, and Baguelin 2019; Funk et al. 2018; Camacho et al. 2015). This stochastic structure is a model for the change in a random process, dZ_t , in relation to the current value, Z_t , where the proportional change $\frac{dZ}{Z_t}$ follows Brownian

motion (Wiersema 2008). That is, normal distributed random *jumps* (dW) moderated by a volatility parameter (α). We do not imply that the actual transmission rate follows a random walk. In fact, the expected value of Z_t is constant over time (Z_0); strictly speaking, a *martingale* (Wiersema 2008). In practice, however, we use this structure as a *scaffold* to obtain some idea of the non-linear structure of the process without committing to a particular form of non-linear model (Priestley 1980). This procedure resembles the use of smoothing splines to estimate coefficients that are allowed to vary as smooth functions of other variables (Hastie and Tibshirani 1993). Although not a requirement for this work, smoothing splines also have a Bayesian interpretation under certain conditions (Kimeldorf and Wahba 1970). In particular, we use the GBM structure to generate non-negative random walks from an initial value (Fig 4.3A displays a set of possible trajectories). The main benefit from random walks is that at each time t , we propose several possible paths that the transmission rate may take and then use the available data to determine their plausibility (Endo, van Leeuwen, and Baguelin 2019). In doing so, we *unravel* the dynamics of the effective contact rate. Formally put, we approximate $p(x_t|y_{0:t})$, the *filtering distribution* (Chopin and Papaspiliopoulos 2020) (see *Materials and methods*).

As noted above, the measurement model is the link between the process model and the data whereby one quantifies (through likelihood densities) the relative consistency of each set of parameter values, or model configuration, with observations. This quantification allows us to perform inference on time-varying and time-independent parameters. Thus, any misspecification in the measurement formulation can lead to overly confident conclusions (Bretó 2018) or biased estimations. In light of its importance, we prevent the consequences of model misspecification by proposing and testing six candidates that account for the incidence data (y^1). Moreover, a subset of these candidates incorporate mobility data (y^2), assuming that this dataset is a proxy observation of the relative contact rate (Z_t).

Before defining each candidate, we clarify certain assumptions regarding the available datasets. On the one hand, for the incidence data (y^1), we posit that *actual* periodic (daily or weekly) symptomatic COVID-19 cases (C_t) stem from the transition $P_t \rightarrow I_t$. Our assumption implies that individuals seek the healthcare system for a diagnostic test as soon as they develop symptoms. Furthermore, under this formulation, it is implicit that *underreporting* is due to the non-identification of asymptomatic cases. On the other hand, for mobility data (y^2), we emphasise its proxy nature. Contrary to the incidence

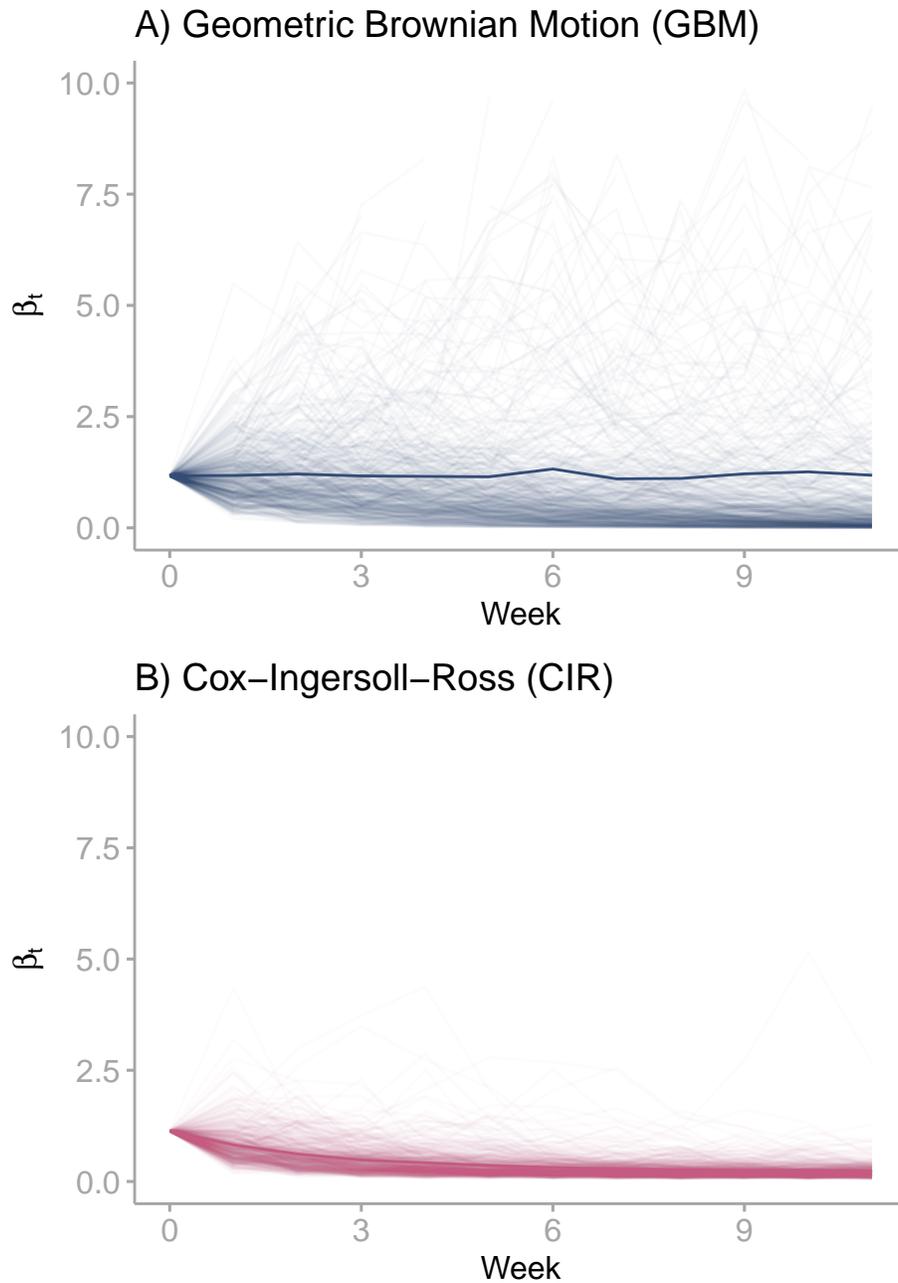


Figure 4.3: Brownian motion trajectories. (A) 200 simulations from a transmission rate described by GBM. We generate these simulations from DGP1’s Maximum Likelihood Estimate (MLE) using the Euler-Maruyama (EM) algorithm. (B) 200 simulations from a transmission rate described by the CIR model. We generate these simulations from DGP2’s MLE using the EM algorithm. In both plots, highlighted trends corresponds to the mean trajectory.

time series, it is not anticipated that models yield faithful replications of Apple’s mobility indexes. Should that be the case, we would have included this data (directly or parametrically) in the process model. However, we refrain from doing so as we deem there may be instances where the two elements are not strongly correlated. To illustrate this point, we consider the case in which the government relaxes social distancing mandates and individuals adopt a mask-wearing behaviour. Under these circumstances, the resulting increase in mobility and social contacts due to relaxed rules do not necessarily entail an equivalent effect on the effective contact rate given that individuals properly wear face coverings during their interactions. Hence, rigid structures in the process model may lead to unrealistic corrections in other parameters. As opposed to such inflexibility, we expect that the mobility data acts as a *nudge* on the transmission rate, guiding the latter towards the former only when plausible. In light of these considerations, for candidates 1 and 2, we formulate the observation of daily symptomatic COVID-19 cases (y_d^1) as independent Poisson and Negative Binomial counts, respectively. Then, we add an observational mechanism that relates Apple’s daily driving data (y_d^2) to the transmission rate’s relative level (Z_t), yielding candidates 3 and 4. Finally, even though King and colleagues (King et al. 2015) recommend that “*models should be fit to raw, disaggregated data whenever possible and never to temporally accumulated data*”, on candidates 1 and 3, we modify their periodicity from daily to weekly measurements, resulting in candidates 5 and 6. It should be noted that the use of weekly measurements has been performed previously in similar studies (Dureau, Kalogeropoulos, and Baguelin 2013; Endo, van Leeuwen, and Baguelin 2019; Funk et al. 2018; Camacho et al. 2015). We refer the reader to Appendix C.2 for the complete set of equations.

Having defined process and measurement structures, we proceed to the inference stage (Table 4.1 summarises the results). Since non-linear SSM do not allow closed-form solutions (Chopin and Papaspiliopoulos 2020) to calculate likelihood values, we must resort to simulation-based approaches such as *Sequential Monte Carlo*, also known as the *Particle Filter*. Naturally, these estimates must be robust so as to guide the inference process. By robustness, we refer to the quality that the Particle Filter returns similar likelihood values for various runs from a single model configuration. Furthermore, as with any Monte Carlo approach, it is expected that as the number of samples tends to infinity, the likelihood error (among various runs) converges to zero. To test this feature, we run the Particle Filter using the R package *pomp*,

4.3. State-Space Models

Table 4.1: Measurement model candidates

Id	Frequency	Incidence	Mobility	Converges	Fits incidence
1	Daily	Pois	No	No	N/A
2	Daily	Nbin	No	Yes	Yes
3	Daily	Pois	Yes	No	N/A
4	Daily	Nbin	Yes	Yes	No
5	Weekly	Pois	No	Yes	Yes
6	Weekly	Pois	Yes	Yes	Yes

which implements the Sequential Importance Sampling algorithm (Gordon, Smith, and Salmond 1993). In particular, through these runs, we evaluate likelihood estimates for each model candidate by varying the number of particles (samples), the integration step size, and the model configuration (see Appendix C.3). The results indicate that measurement models that account for daily incidence observations as Poisson counts lead to unstable estimates. This finding suggests model misspecification in candidates 1 and 3, which are discarded from the pool.

To the remaining candidates, we estimate their latent states. Given its strength to infer time-varying random variables in the state space, the Particle Filter is also appropriate to numerically approximate (via samples) filtering distributions (Arulampalam et al. 2002). Nevertheless, drawing relevant samples requires plausible fixed-parameter values. Here, we assume that three parameters in PM1 are unknown: the effective contact rate at time 0 (ζ), the initial value of preclinical individuals (P_0), and the volatility parameter (α). Moreover, additional parameters may be required depending upon the specific measurement model. To infer such parameters, we employ the *Iterated Filtering* algorithm (Ionides, Bretó, and King 2006; Ionides et al. 2011). This Maximum Likelihood estimation method has been designed to perform statistical inference on SSM and has been widely used to study infectious disease models (Ionides, Bretó, and King 2006; Bretó et al. 2009; He, Ionides, and King 2010; King et al. 2015; Wale et al. 2019). Briefly, Maximum likelihood via Iterated Filtering (MIF) is a modified version of the Particle Filter, in which a sequence of filtering operations converges to the Maximum Likelihood Estimate (MLE). The key feature in this procedure is the set of stochastic perturbations applied to the unknown parameters in between the sequence of filtering operations, resulting in the selection of plausible

parameter values in the light of the available data. Furthermore, the synergy between MIF and the Particle Filter permits us to calculate uncertainty bounds around the MLE. In particular, we use the *Profile Likelihood* method (Pawitan 2013) and its refined version, the *Monte Carlo-adjusted profile* (Ionides et al. 2017). Ultimately, all of this information facilitates the construction of the parameters’ likelihood surface.

For each model, we leverage its likelihood surface to draw sets of point estimates from the neighbourhood surrounding the MLE (King et al. 2015). These draws are subsequently plugged into the Particle Filter. In addition to likelihood estimates, *pomp* returns, for every run, a set of samples representing the filtering distribution at each time t . Then, we assign a weight to each run based on its relative likelihood. In doing so, we account for parameter uncertainty in the results. Finally, we summarise the results by computing weighted averages on the samples. This procedure allows us to calculate the uncertainty in the predicted latent states by the filtering distribution. The reader can find the complete set of results in Appendix C.4-C.7.

$$\frac{dC}{dt} = \eta P_t - C_t \delta(t \bmod 7) \quad (4.6)$$

$$y_w^1 \sim \text{Pois}(C_t) \quad (4.7)$$

$$y_w^2 \sim \text{Normal}(Z_t, \tau) \quad (4.8)$$

The inference process on Candidate 2 (see Appendix C.4) reveals that this model yields a filtering distribution that fits the observed daily incidence. Interestingly, although Candidate 2’s measurement model does not incorporate mobility data in its structure, the predicted relative contact rate captures the observed mobility indexes, albeit with a large degree of uncertainty. This finding supports the argument that such a dataset could be an adequate proxy for the relative contact rate. Then, one would logically expect that incorporating Apple’s data into the measurement model (as we did for Candidate 4) would diminish the resulting uncertainty in the filtering distribution. However, the results (see Appendix C.5) show that the enhanced fit on the effective contact rate stems from unrealistic corrections to the predicted incidence, rendering Candidate 4 unreliable. On the other hand, we notice that Candidate 5’s filtering distribution and parameter

estimates convey similar insights to those of Candidate 2. Therefore, the change in periodicity does not result in severe loss of information. Yet more important, the crucial feature of the weekly formulation is that it allows integrating mobility data seamlessly into the measurement model (Candidate 6). This integration is accomplished without compromising the prediction on incidence counts and simultaneously reducing the uncertainty in the relative contact rate’s fit. This behaviour differs from the unrealistic fit achieved by Candidate 4. We ascribe the resulting harmony between the two datasets to the stringency imposed by the Poisson distribution, which implicitly prioritises incidence counts over mobility indexes. In consequence, we select Candidate 6’s measurement model (Eq (4.6)-(4.8)) as the structure (OM1) that completes DGP1’s formulation (Fig 4.1).

Fig 4.4A presents a comparison between the predicted number of weekly symptomatic cases from DGP1 and observed incidence. Notice that this is a contrast between measurements (y_w^1) and a latent state (C_t). Although this approach is not generally applicable (comparing measurements to predicted latent states), in this case it is valid given that C_t corresponds to the mean of the measurement model (Eq (4.7)). In Fig 4.4A, it can be seen that this model’s filtering distribution captures the actual values in regions of high plausibility, thus yielding an accurate fit. This result helps us gain confidence in the model’s structure as an adequate dynamic hypothesis to the studied phenomenon, considering that it can reproduce the observed behaviour (Oliva 2003). Similarly, the estimated relative effective contact rate replicates to a large extent its assumed measurement values (Fig 4.4B). As expected, the filtering distribution does not capture all of the measurements (Weeks 9-11), given the proxy nature of the data. However, these results allow us to elucidate the trajectory of the effective contact rate, and in turn, the effective reproductive number (see *Material and methods* for the estimation of this quantity). It must be remarked that in the early stages of this outbreak, the dynamics of the transmission rate determined the level of \mathfrak{R}_t . This characteristic occurs when the susceptible fraction is close to one, as was the case during the first wave (HPSC 2020). In Fig 4.4C, we present the estimated \mathfrak{R}_t , where it can be observed that the behaviour change (presumably caused by mobility restrictions and people’s awareness) led to an \mathfrak{R}_t close to or below the epidemic threshold, bringing about a lowering of the incidence rate.

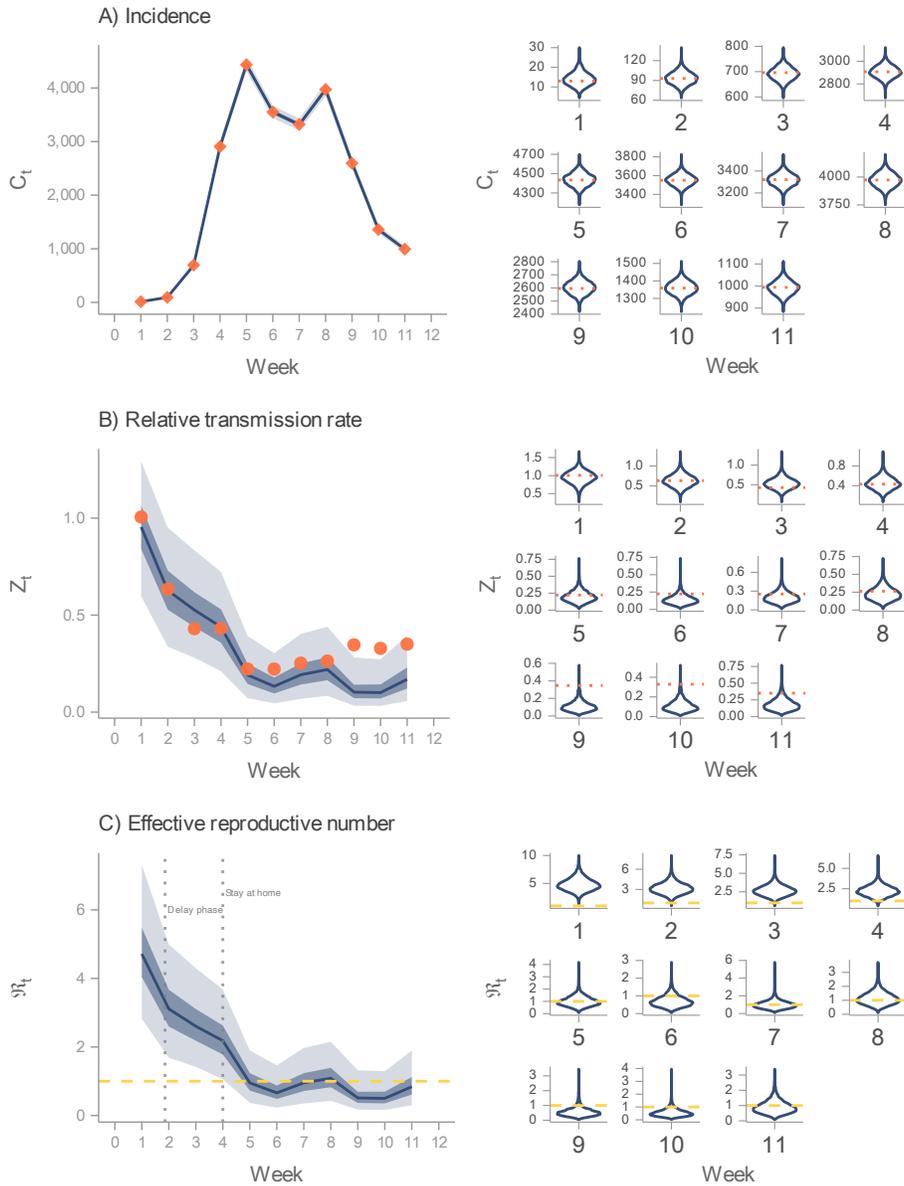


Figure 4.4: Inference on DGP1. Predicted values stem from DGP1’s filtering distribution. In the LHS, the solid line indicates the median, and the darker and lighter ribbons represent the 50% and 95% CI, respectively. (A) Predicted incidence (solid line and ribbons in the LHS; violin plots in the RHS) against actual cases (rhombi in the LHS; horizontal dotted lines in the RHS). (B) Predicted relative transmission rate (solid line and ribbons in the LHS; violin plots in the RHS) compared to mobility index (points in the LHS; horizontal dotted lines in the RHS). (C) Predicted reproduction number (solid line and ribbons in the LHS; violin plots in the RHS). Horizontal dashed lines denote $\mathfrak{R}_t = 1$.

4.3.2 DPG2 - Cox-Ingersoll-Ross

$$\frac{dZ}{dt} = \nu(v - Z_t) + \sqrt{\alpha}Z_t dW \quad (4.9)$$

The dynamics of the transmission rate (Fig 4.4B) uncovered by DGP1 exhibit a compelling pattern. The transmission rate gradually decays for several weeks from its initial value until it levels off around a determined value. In other words, a pattern that resembles *goal-seeking behaviour* (Barlas and Yasarcan 2008). Based on this recognition, we formulate the relative transmission rate in terms of the Cox-Ingersoll-Ross (CIR) model (Cox, Ingersoll, and Ross 1985). This formulation (Eq (4.9)) is a compromise between the rigidity of a deterministic structure and the flexibility offered by random walks. Under this structure, the randomly-moving quantity of interest (Z_t) is elastically pulled toward a central location or long-term *goal*, v . The strictly positive parameter ν determines the speed of adjustment. In practice, we can interpret the long-term goal as the minimum level of mobility that the restrictions can achieve and the adjustment parameter as the rate at which individuals adopt such mandates. Hence, inferring these parameters permit the characterisation of the implemented interventions, a piece of information that cannot be estimated from DGP1. The randomness in this process stems from the diffusion process (second term). That is, stochastic variations from the deterministic trend. More importantly, unlike those in the *Vasicek* and *Ornstein-Uhlenbeck* structures, this particular diffusion process precludes negative values (Wiersema 2008), a *sine qua non* to describe transmission rates. Logically, we ensemble this structure with the SEI3R profile to build the process model (*PM2*). As with DGP1, we assess the convergence of likelihood estimates obtained from the amalgamation of *PM2* and the previously defined six measurement model candidates (see Appendix C.3.7). The results reveal an identical pattern to that observed in DGP1. Therefore, it is warranted to integrate *PM2* and *OM1* (Fig 4.1) to form a DGP that we refer to as *DGP2*. In Fig 4.3B, we present simulated trajectories from this DGP, obtained from a single set of parameters (MLE).

The main objective for building DGP2 is to estimate its latent states conditional on the available data. To do so, we repeat the process applied to DGP1. Specifically, we first perform parameter inference and construct DGP2's likelihood surface using MIF and the Particle Filter. The next step consists of drawing samples from the MLE's neighbourhood to plug them into the Particle Filter. There is a slight alteration in this process, however.

Previously, we selected parameter combinations that yielded likelihood values near the MLE to construct DGP1’s neighbourhood. We then identified the bounds of these parameters to construct a four-dimensional hypercube. From this object, we obtained independent and uniformly distributed samples for each parameter. In light of DGP2’s complex parameter space, we opt for a copula (Sklar 1996) instead of a hypercube. The copula is a multivariate cumulative distribution for which the marginal probability distribution of each variable is uniform, but there is dependence (correlation) among the random variables (unknown parameters). By using the copula, we mitigate biases caused by point estimates that yield abnormal likelihood values. The reader can find the complete set of results in Appendix C.8.

Fig 4.5 displays the results obtained from the inference process carried out on DGP2. Qualitatively, the uncovered values match those obtained from DGP1. Namely, DGP2 produces an accurate fit of the incidence data (Fig 4.5A), and the inferred relative contact rate captures most of the mobility data (Fig 4.5B), resulting in a similar prediction of the effective reproduction number (Fig 4.5C). This outcome provides reassurance on the estimated transmission rate as an adequate account of the observed time series. That is, from two DGPs that differ in the transmission rate’s formulation, we estimate equivalent trajectories. DGP2, though, does not reduce significantly the uncertainty (see Appendix C.8.1.2) in the parameters (ν and v) that characterise the implemented NPIs.

4.3.3 DGP3 - Adaptive expectations

$$\beta_t = \zeta Z_t^1 \tag{4.10}$$

$$\frac{dZ^i}{dt} = \begin{cases} \frac{(v-Z_t^i)}{(\nu^{-1}/n)} & \text{for } i = n \\ \frac{(Z_t^{i+1}-Z_t^i)}{(\nu^{-1}/n)} & \text{for } i < n \end{cases} \tag{4.11}$$

The trajectories derived from the two previous DGPs (DGP1 and DGP2) suggest that it is reasonable to assume that the transmission rate’s dynamics indeed follow a goal-seeking pattern (Eq (4.10)). This conjecture is in agreement with the economic theory of *adaptive expectations*. First applied by Irving Fisher (Fisher 2006), this hypothesis posits that individuals gradually adjust their beliefs, and hence behaviour, in order to eliminate the discrepancy

4.3. State-Space Models

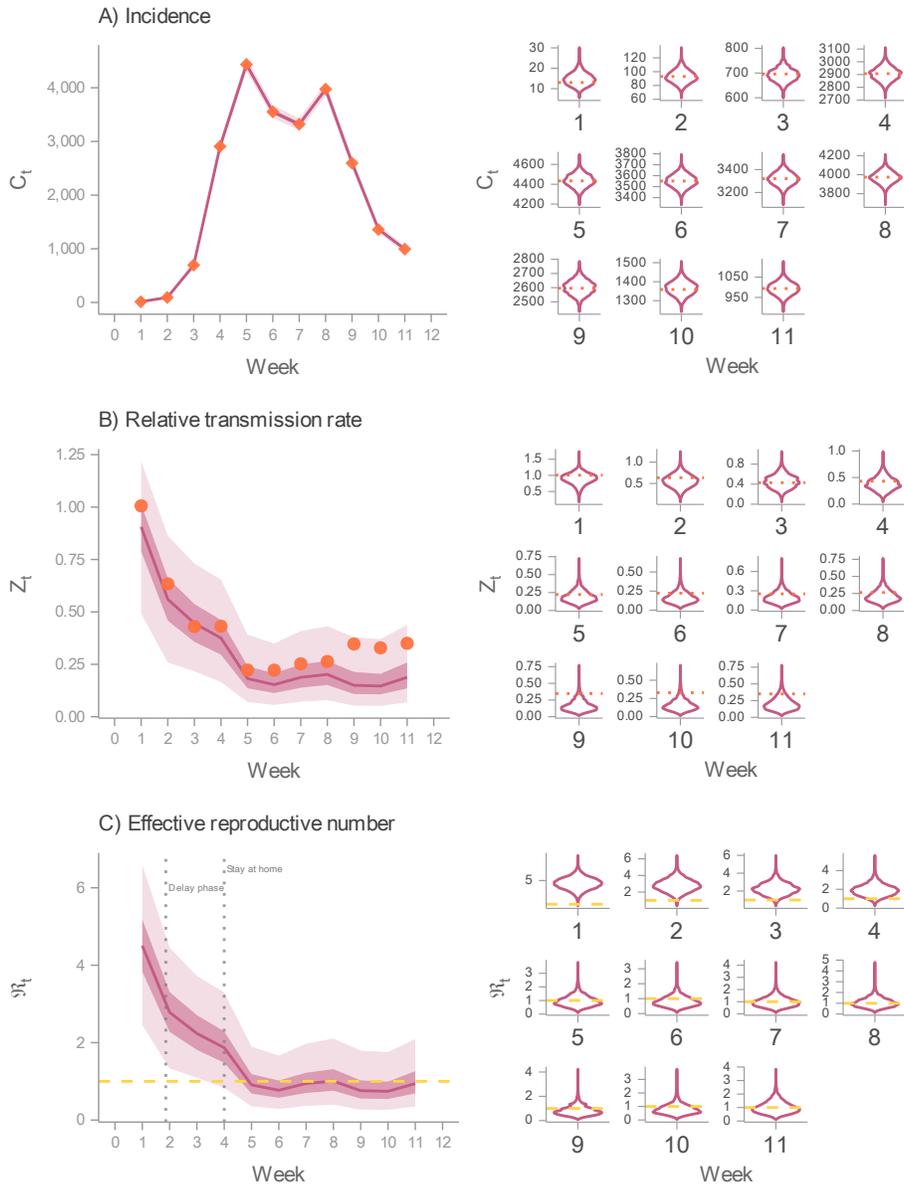


Figure 4.5: Inference on DGP2. Predicted values stem from DGP2’s filtering distribution. In the LHS, the solid line indicates the median, and the darker and lighter ribbons represent the 50% and 95% CI, respectively. (A) Predicted incidence (solid line and ribbons in the LHS; violin plots in the RHS) against actual cases (rhombi in the LHS; horizontal dotted lines in the RHS). (B) Predicted relative transmission rate (solid line and ribbons in the LHS; violin plots in the RHS) compared to mobility index (points in the LHS; horizontal dotted lines in the RHS). (C) Predicted effective reproduction number (solid line and ribbons in the LHS; violin plots in the RHS). Horizontal dashed lines denote $\mathfrak{R}_t = 1$.

between the current state and a *desired* one (Sterman 2000). In this case, such a discrepancy is the gap between individuals' behaviour at a given time t and the level of mobility that the restrictions (implicitly) aim to achieve. Mathematically, the n th-order information delay or *exponential smoothing* (Eq (4.11)) provides a formal description of such an adjustment. This deterministic formulation describes the changes in current behaviour (Z_t^1) as the result of a series of intermediate exponential adjustments ($\frac{dZ^i}{dt}$), which one can interpret as the multiple stages intervening since the Government decrees mobility restrictions to the point where individuals alter their behaviour in accordance with the new rules. The delay order (n) represents the number of stages, where the most simple case ($n = 1$), the 1st-order information delay, is equivalent to the deterministic term in Eq (4.9). On the other hand, when $n \rightarrow \infty$, the dynamics follow a *term-time forcing* pattern.

$$\frac{dC}{dt} = \eta P_t - C_t \delta(t \bmod 1) \quad (4.12)$$

$$y_1^d \sim \text{Pois}(C_t) \quad (4.13)$$

To establish the exact number of stages, we evaluate the performance of nine candidate structures ($n = 1, \dots, 9$) in explaining the available data (incidence and mobility). From this evaluation, we ensemble the selected candidate with the SEI3R profile to generate the process model (PM3) of the third DPG (DPG3) presented in this chapter (Fig 4.1). To complete DGP3's description, we formulate a measurement model (OM2) for the observed daily reported cases (y_d^1). As with DGP1 and DGP2, we assume these counts result from a Poisson distribution (Eq (4.12)-(4.13)). Moreover, OM2 does not include a structure relating mobility data to the relative transmission rate. We base this decision on the results shown in the previous sections. Since the mobility data is an imperfect predictor of the transmission rate, its inclusion in the inference process of a rigid deterministic structure may lead to *forced* model fits, resulting in undesired biases in parameter estimations. In relation to the inference process, since PM3 is deterministic, the inference of the filtering distribution becomes the estimation of DGP3's expected value. We approximate such expected value from a Bayesian perspective (Andrade and Duggan 2021; Grinsztajn et al. 2021) using Hamiltonian Monte Carlo (Betancourt 2018) via Stan. The complete set of results can be found in Appendix C.9.

To illustrate the selection of DGP3’s process model, we present the estimated expected values (fits) for each of the nine candidate structures (Fig 4.6). We depict expected values through simulated trajectories generated from one hundred draws from each model’s posterior distribution. The results indicate that all of these structures yield similar fits to the incidence data. Using the mean absolute scaled error (MASE), a metric designed to measure the accuracy of time-series predictions (Hyndman and Koehler 2006), we notice diminishing marginal gains in accuracy as the order (of the number of stages) increases. These gains, though, are so tenuous that they do not provide clear guidance about which model to choose. To further complicate matters, the lower the delay order, the higher the likelihood value. Nevertheless, when we compare the expected relative transmission rate to mobility data, it can be seen that some structures approximate better the latter than others. If we accept the premise that mobility data is a proxy (supported by the results from DGP1 and DGP2), yet imperfect, measurement of the relative transmission rate, we can then lean towards the delay order that yields the lowest MASE ($n = 4$). From this structure’s posterior distribution, we estimate, among others, the adjustment rate (ν ; mean = 0.05, sd = 0.001), the minimum level of mobility (v ; mean = 0.11, sd = 0.005), and the effective reproduction number (discussed below). Notice that the particular form of the non-linear contact rate restricts the marginal distributions of ν and v to such an extent that most of the probability mass concentrates on extremely narrow neighbourhoods. Despite this, those estimates resemble DPG2’s MLE ($\nu = 0.05$, $v = 0.19$), which help us gain confidence in the overall process.

Acknowledging that the performance metrics above (MASE and likelihood values) do not lead to an unambiguous choice, we explore the implications of selecting an alternative measurement model. As it is widely known, the Poisson distribution is a discrete probability distribution in which the observation mean equals the variance (Blitzstein and Hwang 2019). Hence, using this distribution as a measurement model imposes a stringent assumption on the observational process of incidence counts. By contrast, the Negative Binomial distribution offers a more flexible framework to account for overdispersion in daily incidence. Moreover, the Negative Binomial converges to the Poisson distribution under a specific configuration. For this reason, we test the implications of this alternative formulation. See the complete set of results in Appendix C.9.3. Indeed, the posterior distribution suggests the presence of a small amount of overdispersion in the incidence data. However, such gain in realism is achieved at the expense of a *degenerate* posterior distribution.

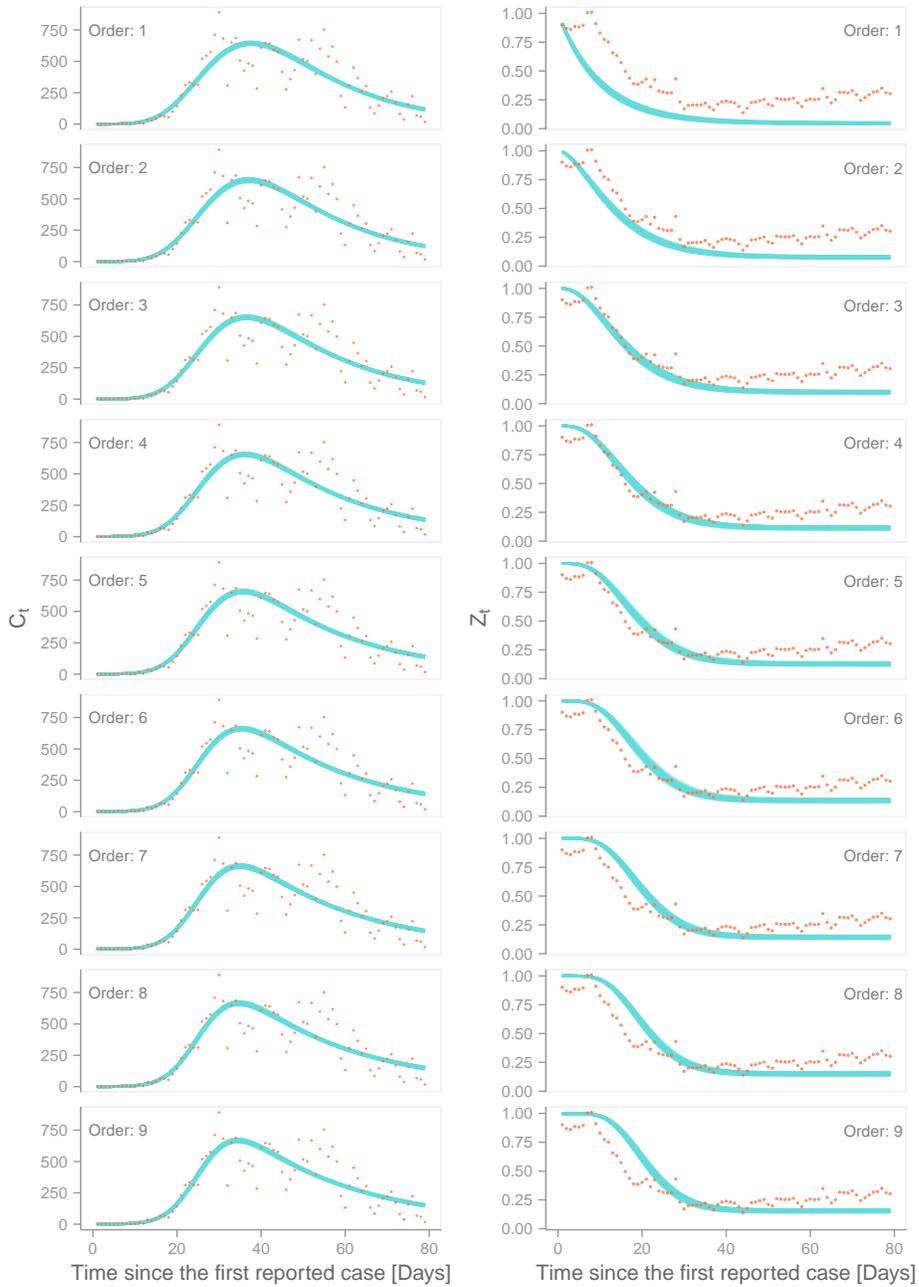


Figure 4.6: Inference on DGP3. Comparison between expected values and data. On the LHS, for each model, we show 100 overlapped simulations of the predicted incidence against daily case counts. On the RHS, for each model, we show 100 overlapped simulations of the predicted relative transmission rate against Apple’s mobility data. In this plot, we estimate the predicted values from the posterior distribution of each of the DGP3’s nine candidate process models.

Succinctly, any of the model candidates coupled with the Negative binomial distribution yields a posterior distribution of two distinct modes, even from a single unknown parameter. This kind of behaviour is not unusual in Ordinary Differential Equation models. For instance, Gelman and colleagues (Gelman et al. 2020) report a similar experience in the calibration of a simple mechanistic model of planetary motion.

In the set of bimodal distributions returned by Stan, we recognise two types of modes. One that corresponds to a region of unrealistic parameter values for which the HMC algorithm reveals pathological behaviour (divergences and low E-BFMI) (Betancourt 2018) in the sampling procedure, rendering the inference from these samples unreliable. Conversely, the Markov chains that land in the other type of mode do not trigger any warnings from Stan. Furthermore, these *well-behaved* modes are located in regions similar to those found using the Poisson distribution. Following an exploratory analysis, we find that *well-behaved* modes and the set of posterior distributions obtained from the Poisson model provide similar (although not identical) information. Overall, the choice of the Poisson distribution and the delay order (4th) is the outcome of considering as a whole the information provided by the previous DGPs, and the two explored measurement models. This assessment, therefore, implies that we envision the Poisson measurement model as an approximation that does not compromise the insights from the inference process. However, one cannot generalise this result to other applications. That is, taking the Poisson distribution as a default. On the contrary, it is imperative to test the assumptions embedded in any proposed measurement model and evaluate the trade-offs entailed by each alternative.

4.4 Discussion

Novel datasets that may assist modellers in gaining deeper insight into the dynamics of an infectious disease deserve a thorough examination. This task entails establishing adequate links between the data and a dynamical hypothesis. Far from trivial, one may derail the entire inference process by adopting a misspecified structure. For this reason, a robust approach involves the assessment of various levels of model complexity that account for the available data, which inevitably involves trade-offs (Funk and King 2020). This work highlighted the implications of committing to a particular model formulation. As seen above, DGP1 and DGP2 (DGPs with a stochastic process model) can only incorporate the mobility data if they are formulated

in terms of weekly observations. Notwithstanding that this requirement reduces the number of data points available for the inference process, the loss of information is negligible. In contrast, a rigid structure such as DGP3 (whose process model is deterministic) restricts the use of mobility data only as a discriminant criterion.

With regard to the inference of fixed parameters, DGP1's *well-behaved* parameter space yields smooth profiles with positive curvature at the MLE from which parameter uncertainty can be seamlessly calculated. Interestingly, when we amalgamate all the likelihood estimates, we obtain surfaces that resemble likelihood profiles. As a result, from three approaches (MCAP, profile, surface), we estimate similar confidence intervals. DGP2's parameter space is, on the other hand, challenging to explore. In fact, the volume of high plausibility is so tightly concentrated that some regions in the MLE's neighbourhood yield vast negative log-likelihood values. To address this issue, we iterated over several hypercube sizes and densities until obtaining profiles with positive curvature at the MLE, although not as smooth as those obtained from DGP1. Despite this hurdle, we obtain similar confidence intervals from the three quantification approaches. Regarding DGP3, given the Bayesian approach used to estimate its parameters, we refer to such uncertainty bounds as credible intervals. We obtain well-behaved posterior distributions for the nine candidate process models whose inference is backed by successful diagnostics unique to HMC. However, parameter estimates (posteriors) vary by the delay order, requiring a subjective assessment to determine which structure is more appropriate. Lastly, we consider the differences in computational burden between the inference methods (MIF + Particle Filter and HMC). Whereas performing parameter inference on DGP1 and DGP2 took roughly 14 and 20 hours, respectively; fitting DGP3's nine candidate models required 6 hours of computational time.

Likewise, the inference of the time-varying quantities deserves close inspection. DGP1 and DGP2 are more flexible than DGP3 in quantifying uncertainty. We illustrate this point with Fig 4.7B and 4.7C. Here, we notice that DGP3 generates an estimate of the relative transmission rate and the effective reproduction number with narrow uncertainty intervals in comparison with those generated by the other DGPs. This apparent precision is the result of committing to a particular form of non-linear model, which imposes a stringent constraint in the shape of the transmission rate. By choosing the 4th-order information delay structure, we implicitly discard the possibility for the other formulations to be true, reducing the uncertainty in the estimations.

However, we demonstrated that the nine delay orders account similarly for the incidence data, and to various degrees of accuracy, for the mobility data. Thus, we interpret the wide intervals generated by DGP1 and DGP2 as the uncertainty in the delay order plus the measurement error. This interpretation suggests that DGP3's plausible model candidates are subsumed under DGP1 and DGP2.

To conclude with this comparative analysis, we reflect on the role of DGPs presented in this chapter. Owing to its flexible formulation, we can employ DGP1 for both retrospective and near real-time analysis (at least for the period where demographic processes do not significantly impact the dynamics of the pandemic). In contrast, DGP2 and DGP3 formulations are context-dependent, restricted to retrospective analyses. Under this last role, we note that common patterns emerge from the three DGPs. Notwithstanding structural differences, all of them produce accurate fits to the incidence data (Fig 4.7A). Naturally, the stochastic process models replicate every feature in the data, whereas the deterministic one captures the underlying trend. Furthermore, the estimated medians for the relative transmission and the effective reproduction number (Fig 4.7B and 4.7C) tell similar stories. That is, individuals gradually decreased their movements following public health advice, which led to a decline in the transmission rate. This reduction pulled \mathfrak{R}_t below the epidemics threshold, causing the incidence rate to subside. It should be noted that this mobility reduction levels off later in Ireland's first wave, suggesting a limit on the effectiveness of the implemented policies. We interpret this limit as the minimum mobility required for running essential services.

Finally, even though the primary interest of this work has been on estimating the effective reproduction number (\mathfrak{R}_t), a by-product from this inference process is the approximation of the basic reproductive number (\mathfrak{R}_0). This widely accepted metric (Delamater et al. 2019) is defined as the average number of secondary infections produced when one infected individual is introduced into a totally susceptible population (R. Anderson and May 1992). In the context of Ireland's COVID-19 epidemic, we derive similar \mathfrak{R}_0 estimates from the three DGPs (DGP1: 95% CI[4.0 - 6.1], DGP2: 95% CI[3.9 - 6.0], DGP3: 95% CI[5.8 - 7.1]). These estimates are in close agreement with a previous modelling study on the COVID-19 pandemic in Ireland (Gleeson et al. 2022), albeit well above the initially reported $\mathfrak{R}_0 = 2.2$ value from Wuhan (Li et al. 2020); a value that has been adopted as the reference point by the World Health Organization and other research groups (Davies et al. 2020;

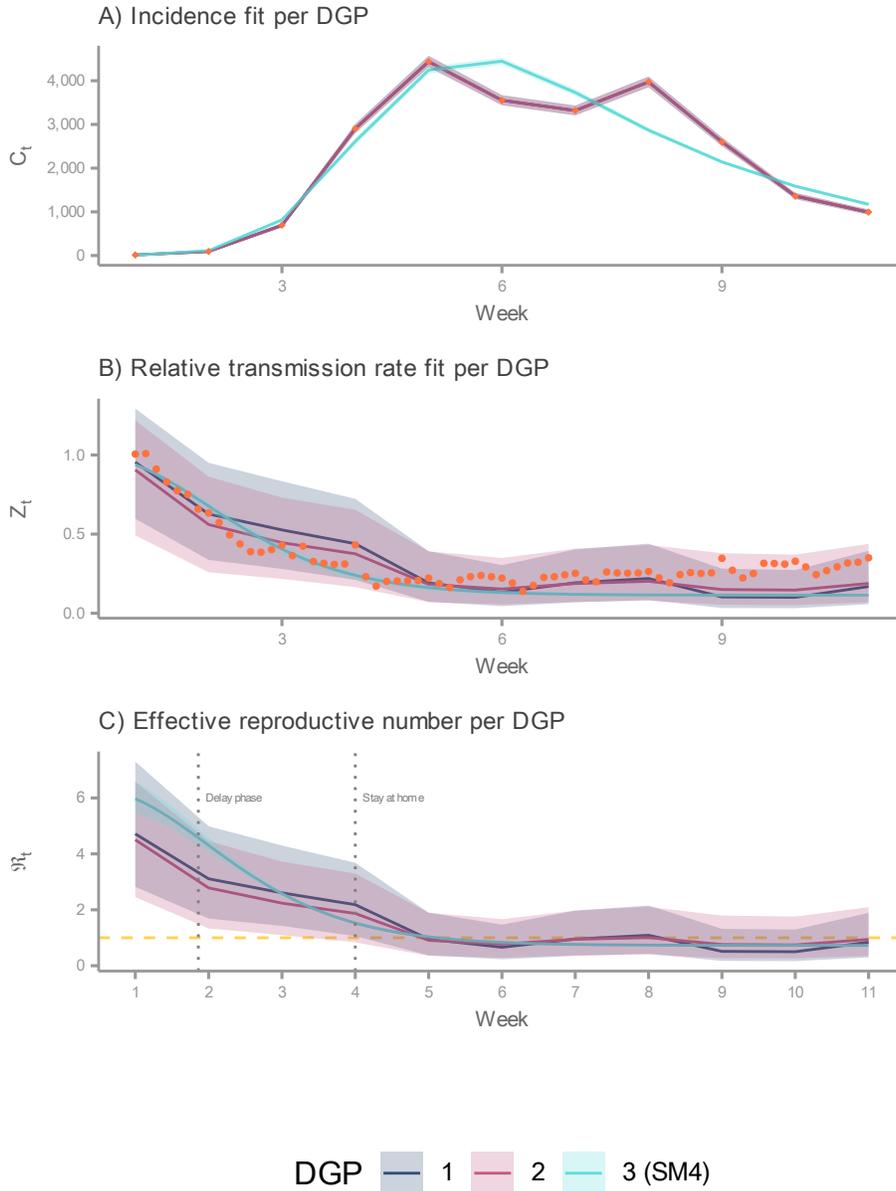


Figure 4.7: Comparison of predicted states. These values stem from the filtering distribution (DGP1 and DGP2) and the posterior distribution (DGP3). Here, DGP3's process model is the 4th-order delay. Further, solid lines indicates the median, and the ribbons represent the 95% CI. (A) Predicted incidence by DGP and weekly cases (rhombi). (B) Predicted relative transmission rates by DGP and mobility data (points). (C) Predicted reproductive numbers by DGP. The horizontal line denotes the epidemics threshold.

Petersen et al. 2020). Other streams of research, however, argue that the initial estimate was low (Katul et al. 2020), and instead, advocate for higher values (4.5 (Katul et al. 2020); 4.7 - 6.6 (Sanche et al. 2020)). Moreover, the reader should recall that \mathfrak{R}_0 is a context-dependent metric, and variations are expected due to population heterogeneity (e.g., age, spatial location, host genetics). In any case, we acknowledge the limitations that stem from the calibration of homogeneous population models, which require high \mathfrak{R}_0 values to achieve accurate fits (He, Ionides, and King 2010). To address such limitations, future research should test the impact of disaggregating (by age or location) the structures presented in this chapter. Another research avenue could explore the effect of replacing the deterministic rates in the within-host profile of these DGPs with stochastic ones that account for demographic and environmental effects.

4.5 Materials and methods

$$\frac{dS}{dt} = -S_t\lambda_t \quad (4.14)$$

$$\frac{dE}{dt} = S_t\lambda_t - \sigma E_t \quad (4.15)$$

$$\frac{dP}{dt} = \omega\sigma E_t - \eta P_t \quad (4.16)$$

$$\frac{dI}{dt} = \eta P_t - \gamma I_t \quad (4.17)$$

$$\frac{dA}{dt} = (1 - \omega)\sigma E_t - \kappa A_t \quad (4.18)$$

$$\frac{dR}{dt} = \kappa A_t + \gamma I_t \quad (4.19)$$

$$\lambda_t = \frac{\beta_t(P_t + I_t + \mu A_t)}{N_t} \quad (4.20)$$

4.5.1 SEI3R profile

This within-host profile (Eq (4.14)-(4.20)) is formulated based on the work from Davies and colleagues (Davies et al. 2020). Here, we assume that individuals are initially susceptible (S) and become exposed (E), at a rate λ , after effective contact with an infectious person (I, P, A). After a latent period (σ^{-1}), exposed individuals follow one of two paths. With probability ω , following a period (η^{-1}) of preclinical infectiousness (P), individuals develop full symptoms while transmitting the pathogen. This stage is known as the clinical infection state and lasts for γ^{-1} days. On the second path, with probability $1 - \omega$, individuals enter a subclinical state (A) with none (asymptomatic) or mild symptoms (paucisymptomatic), and are not captured by the healthcare system. Individuals on this path recover after κ^{-1} days and are relatively (μ) less infectious than their counterparts on the clinical path. Finally, individuals from both paths eventually converge to the recovered state (R), in which they are no longer infectious and are immune to re-infection. In Appendix C.1, we provide the values for fixed parameters and initial states and their respective sources.

4.5.2 Basic and effective reproductive number

$$\mathfrak{R}_0 = \zeta Z_0[\omega(\gamma^{-1} + \eta^{-1}) + (1 - \omega)\mu\kappa^{-1}] \quad (4.21)$$

$$\mathfrak{R}_t = \mathfrak{R}_0 \frac{S_t}{N_t} \quad (4.22)$$

To derive an analytical expression for the basic reproduction number (\mathfrak{R}_0) from the SEI3R profile, we employ the next-generation matrix method (Diekmann, Heesterbeek, and Metz 1990). That is, we rewrite the infected states' transitions (rates) in the form of two matrices. The first matrix \mathcal{F} corresponds to the rate of appearance of new infections in each compartment of infected individuals, and the second matrix \mathcal{V} corresponds to the rate of other transitions between compartments of infected individuals. From these matrices, we define the next generation matrix as $\mathcal{F}\mathcal{V}^{-1}$, whose largest eigenvalue (spectral radius) corresponds to \mathfrak{R}_0 (van den Driessche 2017). We obtain the spectral radius's analytical solution (Eq (4.21)) using the software system *Mathematica* (see Github repository). Following this expression, we can define (Eq (4.22)) the effective reproductive number (\mathfrak{R}_t) as the product between \mathfrak{R}_0 and the susceptible fraction ($\frac{S_t}{N_t}$).

4.5.3 Filtering distribution

The essence of the state-space approach is to estimate the state of a dynamical system using a sequence of noisy measurements made on the system. We formulate this problem in terms of a recursive filter whose purpose is to construct the state's posterior probability density function (pdf) based on all available information, including the set of received measurements (Arulampalam et al. 2002). Formally, $p(x_t|y_{1:t})$. We refer to this pdf as the *filtering distribution*, whose inference process consists of two stages: *prediction* and *update*.

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \quad (4.23)$$

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \quad (4.24)$$

The prediction stage (Eq (4.23)) draws on the *plug-and-play* property (He, Ionides, and King 2010) to generate, from simulations of the process model $p(x_t|x_{t-1})$, a vector of predictions that describe the state at time t (x_t), which are conditional on the previously estimated state ($x_{t-1}|y_{t-1}$). Then, the update operation (Eq (4.24)) uses the latest measurement to modify the prediction pdf ($p(x_t|y_{1:t-1})$). In practice, we assign weights to the prediction vector based on its plausibility, which is estimated from the measurement model $p(y_t|x_t)$. With these weights, we use the Sequential Importance Sampling algorithm (Gordon, Smith, and Salmond 1993) to produce samples that describe the filtering distribution. It is important to remark that this is a sequential process (hence the name Sequential Monte Carlo), executed every time a measurement is received. Moreover, in this simplified formulation, it is assumed that X_0 and θ (Eq (4.1)-(4.2)) are known. We refer the interested reader to (Arulampalam et al. 2002; Chopin and Papaspiliopoulos 2020) for a complete treatment of this approach.

Chapter 5

Conclusion

5.1 Research findings

Fitting compartmental models of infectious diseases to incidence data not only serves to estimate the reproduction number but also allows modellers to infer other crucial elements that contribute to the spread of a disease. This knowledge, summarised by parameter estimates and predictions of latent states, can be further employed to devise cost-effective policies that mitigate the impact of pandemic pathogens on a population. However, model calibration is a delicate procedure, and an inadequate assumption, either in the inference algorithm or model structure, can derail the entire endeavour.

Simulation-based algorithms, such as Hamiltonian Monte Carlo, are becoming the method of choice to prevent biases and overconfidence in estimates, given that their results are less reliant on asymptotic considerations. However, its mere use does not immediately grant reliable estimates. It is thus imperative to comprehend the implications of the formulated model on the inference process and, equally important, interpret correctly the algorithm's output. For this reason and in the context of this thesis, we first asked: *How can one employ Hamiltonian Monte Carlo to obtain robust estimates from SIR-like models fitted to time-series data?* This question motivated Chapter 2, which shows that extracting Hamiltonian Monte Carlo's potential requires a comprehensive workflow. In particular, a Bayesian workflow of formulation, exploration, estimation, and interpretation. These stages, in turn, can be visited more than once, depending on the proposed model and the available data.

Furthermore, each step in this workflow contributes distinctly to validating the inference procedure. Prior predictive checks allow modellers to verify whether the dynamics of the simulation model are relatively similar to those observed in the data. For instance, a SIR-like model that does not produce any outbreak-like behaviour must be reformulated immediately without running the inference algorithm. In fact, we employed this insight (Chapter 3) in formulating the alternative SEIR, which depends on the inverse of \mathcal{R}_0 . This formulation allows the definition of a prior, in terms of the beta distribution, that assigns zero probability to scenarios where the pathogen fails to invade the susceptible population.

Likewise, diagnostics play a prominent role, especially in HMC. Unsatisfactory results, such as divergences or large potential scale reduction factors, render the sampling process unreliable to the extent that the inference analysis can only proceed after a reformulated model complies with the required diagnostic standards. In Chapter 4, the deterministic candidates did not converge if coupled with the Negative Binomial distribution. As a compromise, we employed the Poisson distribution to properly quantify uncertainty at the expense of acknowledging that a slight bias may have been introduced in parameter estimates.

Nevertheless, satisfactory diagnostics are insufficient to accept the output of an inference algorithm. Posterior samples must also lead to simulations that match the observed behaviour while providing a substantial update on previous beliefs. In other words, there should be a decrease in the uncertainty. This stage was instrumental in Chapter 3 in discovering the relationship between the mean generation time and the basic reproduction number. Trace plots, scatter plots and correlation estimates enabled this discovery. In short, the model must undergo additional revisions if the inference process does not produce an accurate fit or fails to provide new information to the modellers.

Interestingly, robustness is not an exclusive characteristic of Bayesian approaches. In Chapter 4, we obtained estimates from two Data Generating Processes calibrated to incidence and mobility data following the frequentist tradition. However, before predicting the reproduction number, we verified the reliability of likelihood estimates from the Particle Filter by checking its convergence. This step discarded candidates with a Poisson measurement component for the observation of daily case counts. Furthermore, validating the match between actual and simulated behaviour eliminated from the pool a candidate that accounted for daily incidence and mobility. Lastly, we

obtained a large number of estimates along the workflow to construct the likelihood surface, which allowed us to gain enough confidence to apply the profile likelihood method in the estimation of uncertainty bounds. All in all, robustness is a quality that stems from workflow employed rather than a particular school of thought.

Nevertheless, certain weaknesses in model formulation are more challenging to detect. One that has troubled mathematical modellers over the years is formulating the appropriate distribution of the epidemiological delays. This formulation is problematic because if modellers followed the parsimony principle, which suggests choosing the simplest scientific explanation that fits the evidence, modellers would have to use the exponential distribution. However, it has been thoroughly demonstrated that this assumption is unrealistic and leads to systematic biases. Consequently, we asked: “*How can we mitigate biases in \mathcal{R}_0 estimates due to uncertainty in the distribution of the epidemiological delays?*” According to the answer provided in Chapter 3, modellers must configure the SEIR model such that the mean generation time remains fixed at its true value. This value can be directly inserted into the model as a parameter or function of other parameters.

We offer this recommendation based on the results of our simulation study. In this work, we generated synthetic data from several instances of the SEIR model, where each instance differs in the distribution of the epidemiological delays. Then, we fitted candidates from three SEIR parameterisations to this synthetic data. The first parameterisation (*traditional*) fixes the mean of the infectious and latent periods to their true values regardless of their distribution. Estimates from this parameterisation conform to the findings reported in the literature, which states that misspecifying the distribution of the epidemiological delays produces inaccurate predictions. The next parameterisation (*four-unknown*) also fixes the mean latent period but treats the mean infectious period as a random variable. The inference results reveal the linear relationship between the mean generation time and the basic reproduction number, a finding that led to the formulation of the *alternative* parameterisation. This parameterisation attains high coverage levels when recovering \mathcal{R}_0 but loses accuracy in estimating other parameters, such as the initial number of infectious individuals. We further assessed the alternative parameterisation against various transmissibility levels (\mathcal{R}_0) and mean generation times, finding that it is robust to changes in the mean generation time but sensitive to high \mathcal{R}_0 values. Of note, this sensitivity can be detected only in high-fidelity data. That is, data with little overdispersion.

In a nutshell, the alternative parameterisation is a mitigation strategy in the context of incomplete information.

However, there are no mitigation strategies in the formulation of time-varying contact rates. Instead, modellers must follow either a parametric or non-parametric route, each with its own advantages and limitations. Contemplating these choices, we posed the question: “*How can we improve the estimation of the reproduction number from compartmental models when the dynamics of the transmission rate are unknown?*” In Chapter 4, we demonstrated that incorporating mobility data into the workflow can reduce the uncertainty in the estimates. Nonetheless, this incorporation is far from trivial, given that mobility is a proxy measurement of the transmission rate, which implies that its dynamics may not always mimic those of the actual transmission rate. Therefore, integrating this additional dataset with the rest of the model requires caution.

This answer (and caveat) results from a complementary workflow that spanned three Data Generating Processes (DGPs) and both schools of thought for statistical inference. As regards the DGPs, they share the within-host profile but differ in the formulation of the transmission rate. Two of these structures were formulated in stochastic terms. First, we proposed Geometric Brownian Motion (GBM) as a non-parametric mechanism to uncover the dynamics of the transmission rate. Integrating mobility data into this DGP reduces the uncertainty in the predicted reproduction number while maintaining an accurate fit on the incidence data. Nevertheless, this integration necessitates a differentiation in the importance assigned to each measurement component, considering that adding mobility data should not have a detrimental effect on the incidence fit. For this reason, we formulated the measurement of case counts in terms of the Poisson distribution, whose stringency implicitly prioritises incidence over mobility.

Notwithstanding GBM allows the estimation of the reproduction number, its non-parametric nature does not offer further insights into the dynamics of the transmission rate. Consequently, the second DGP’s transmission rate was formulated in terms of the Cox-Ingersoll-Ross (CIR) model, a structure comprising a mixture of deterministic (parametric) and random-walk (non-parametric) components. In addition to providing similar \mathfrak{R}_t estimates to those obtained from the first DGP, fitting this DGP to incidence and mobility data provides insights into the minimum level the mobility restrictions can achieve and the rate at which individuals adopt such mandates. However,

the calculated confidence intervals for these features are imprecise (too wide), which motivated an additional formulation.

The third DGP presented in Chapter 4 corresponds to a structure whose transmission rate is described by exponential smoothing, a fully parametric formulation. Exponential smoothing, though, rather than denoting a specific formula, encompasses a family of equations known as the n th-order information delay. As in Chapter 3, committing to a particular order (distribution) produces different estimates without impacting the quality of the incidence fit. Unlike the previous two DGPs, this DGP does not include a measurement component for the mobility data, given the rigidity imposed by deterministic formulations. Even though varying the delay order produces equivalent incidence fits, it predicts different dynamics for the transmission rate. If we accept the premise that mobility data provides information, albeit imperfect, on the relative transmission rate, we can then employ this additional dataset as a discriminant criterion. It is worth mentioning that uncertainty regions calculated from these deterministic models are substantially narrower than those produced by the semi-deterministic formulations, confirming the trade-offs between parametric and non-parametric approaches.

5.2 Contribution

Beyond any particular estimate or proposed formulation, this thesis emphasises the importance of adopting a systematic approach to fit compartmental models of infectious diseases to time series data. Irrespective of the particular school of thought, model calibration is not a box-ticking exercise that validates a structure merely because it matches the data. Instead, model calibration becomes a learning process that leverages statistical methods to confront embedded assumptions in the proposed formulations with the data if performed rigorously. Namely, through a comprehensive workflow that validates each component of the inference process, including the calibration algorithm. Since SIR-like models are commonly employed for decision-making, a misleading insight from an inappropriate assumption can result in adverse consequences for a substantial portion of the population.

More concretely, a major contribution of this thesis (Chapter 3) is the discovery of the linear relationship between the mean generation time and \mathfrak{R}_0 , which occurs irrespective of the particular distribution of the epidemiological delays. In practice, this relationship implies that early in a pandemic, studies

should prioritise the estimation of the mean generation time to enable the use of compartmental models as a reliable estimator of the reproduction number. This finding suggests that precise estimation of the latent and infectious periods, in terms of their mean and distribution, may not be strictly required as long as their approximations add up to the estimated mean generation time.

Furthermore, Chapters 2 and 3 stress the relevance of visualising and interpreting parameter interactions. These interactions measured by parameter correlations (multicollinearity) can occur due to linear dependence among parameters (mathematical artefact caused by creating parameters from other parameters) or dynamic interconnection. Since SIR-like models comprise an interconnected network of variables (feedback loops), it is essentially unrealistic to expect independence among parameters. In certain cases, parameters interact to such an extent that mask errors in model formulation. Chapter 3 describes various instances where this situation occurs. For example, misspecifying the infectious period distribution leads to biases in the reporting rate (ρ), the effective contact rate (β), and ultimately, the basic reproduction number (\mathfrak{R}_0). Similarly, misspecifying the latent period distribution produces biases in the estimation of the initial number of infectious individuals, albeit it does not impact \mathfrak{R}_0 . In both cases, modellers may accept the fitted models precisely because the biases in the parameters generate simulations that are indistinguishable from those produced by the true model.

While parameter interactions can sometimes conceal errors in model formulation, they can also create a complex parameter space, resulting in unidentifiable variables. For instance, when we treated both the effective contact rate (β) and the mean infectious period (γ^{-1}) as random variables, the posterior distribution's trace plot showed that these parameters were perfectly correlated. To further complicate matters, this correlation stretched across the entire space of the mean infectious period, rendering it unidentifiable, as evidenced by the similarity between its prior and posterior distributions. However, strong parameter interactions do not immediately imply unidentifiability if this interaction is confined to a small region of the parameter space, such as the correlation between the initial number of infectious individuals and the effective contact rate.

On the other hand, Chapter 4 shows that mobility data is not a perfect predictor of the transmission rate, and its incorporation into the inference workflow can lead to unrealistic estimates if not properly treated. Given the

simplicity of the model employed, we were able to detect these shortcomings, but in more elaborated models, that may not be the case. Moreover, this work also presented a weakness in the Negative Binomial distribution as a measurement component of incidence reports. While this distribution is more flexible than the Poisson distribution in cases where the data exhibits overdispersion, the extra parameter added by the Negative Binomial distribution can create undesired parameter spaces or give too much weight to proxy measurements. In both cases, we had to revert to the Poisson distribution while verifying that it did not introduce significant biases. In short, adding one extra parameter to the inference procedure, especially in the context of compartmental models, substantially increases the complexity of the inference procedure. Therefore, the trade-offs of this decision must be thoroughly assessed.

We anticipate these findings will serve as a reference point for modellers that fit SIR-like models to time-series data. These guidelines include which information to collect, how to approach the inference procedure, and how to interpret calibration results. In doing so, the epidemiology field will benefit from reliable estimates that improve the understanding of the dynamics of infectious diseases.

Lastly, this work also produced *readsdr*, a package that translates models built in System Dynamics software into R objects and Stan code. The use of *readsdr* substantially improved the productivity of the formulation-fitting-interpretation cycle to the extent that without this package, it would have been infeasible to conduct this research, given the enormous amount of time that coding Stan scripts imposes. We anticipate this package will be a valuable resource for performing rigorous inference on compartmental models.

5.3 Limitations and future work

Whereas the recommendation of following a thorough workflow has no mitigation factors, given the scope of this research, we limited ourselves to simple models. This choice implied an exchange of realism for mathematical and computational convenience. Admittedly, the assumption of homogeneous mixing, which means everyone interacts with equal probability with everyone else (Keeling and Rohani 2011), is unrealistic inasmuch as empirical evidence has shown that mixing patterns are highly assortative (i.e., ‘with-like’)

(Mosson et al. 2008). More importantly, the reproduction number estimated from disaggregated models is generally larger than if estimated from aggregated structures (Keeling and Rohani 2011). Therefore, discarding possible heterogeneities arising from age, space, or behavioural aspects introduces biases in estimates. Consequently, future work should study the relationship between the mean generation time and the basic reproduction number in more intricate extensions of the SIR model. These elaborated models will inevitably increase the number of estimated parameters. Consequently, modellers should also devise mechanisms to analyse and understand much larger parameter spaces (mainly correlations).

In the same vein, we favoured compartmental models with deterministic transitions for computational convenience. As we saw in Chapter 4, fitting a deterministic model takes much less time than fitting a stochastic one. Nevertheless, it has been demonstrated that stochasticity can play a significant role in the dynamics of an infectious disease, even in large populations (Rohani, Keeling, and Grenfell 2002). Thus, subsequent work should study the relationship between the basic reproduction number and the mean generation time in models that account for demographic and environmental stochasticity.

As regards the inference of time-varying transmission rates, it remains to be determined which approach should be preferred and under which circumstances. For instance, one of the formulations in Chapter 4 was in terms of Geometric Brownian Motion merely because it was adopted in prior studies. However, these studies do not provide any guidelines as to what type of trajectory GBM can recover and when it should be precluded.

5.4 Final remarks

In 1976, the renowned statistician George Box quipped, *all models are wrong, but some are useful*. While it is undeniable that no model, however elaborated it may be, can capture all the features of a phenomenon, we, as modellers, cannot take this aphorism as a justification to overlook problematic assumptions in certain structures or inference methods. Thankfully, the advent of computational power has enabled the development of more efficient algorithms that expose weaknesses in model formulation. This exposure allows us to pinpoint areas for improvement, resulting in robust structures that yield reliable estimates for decision-making. It is hoped that this thesis

will usher in research that reveals further insights into the dynamics of infectious diseases.

5.4. Final remarks

Appendix A

Supplementary Information (Ch. 2)

A.1 Measurement model

In this section, we discuss the choice of the measurement model and how we use prior predictive checks to guide such a decision. Initially, we opt for the default choice, i.e., the normal distribution. This choice implies that, at every time step, the difference between the measurement and the true value (error) follows a normal distribution. Additionally, this distribution entails that the error across all time steps is similar (homoscedasticity). In other words, the magnitude of the error is indifferent to the magnitude of the true value, an assumption that may seem unrealistic. As could be expected, the normal distribution adds a new unknown, the standard deviation (δ). To test this model, we assume $\delta \sim \text{Cauchy}(0, 1)$, and simulate 500 trajectories. We notice that this configuration yields continuous unconstrained values, in contrast to Cumberland's discrete non-negative measurements. These dissimilarities hint that more suitable specifications should be found.

Next, we consider the Poisson distribution given that it has been used in the empirical treatment of count data, particularly concerning counts of events per unit of time. This distribution lifts the constraint of equal variance across measurements as the error magnitude is proportional to the true number of reported individuals at each time step. As with the normal distribution, we generate 500 simulations. As we saw above, the Poisson model adequately captures the given data.

A.1. Measurement model

Finally, to count data, the Negative Binomial distribution is an alternative to the Poisson should there be overdispersion in the data. We model such overdispersion in the Negative Binomial's scale parameter ($\phi \sim \text{Half-normal}(0, 1)$). We can see that the trajectories generated by the Negative Binomial are more dispersed than the ones produced by the Poisson distribution (Fig A.1). Nevertheless, in this case, they do not provide a more accurate representation of the data. Consequently, we adopt the Poisson distribution as the measurement model.

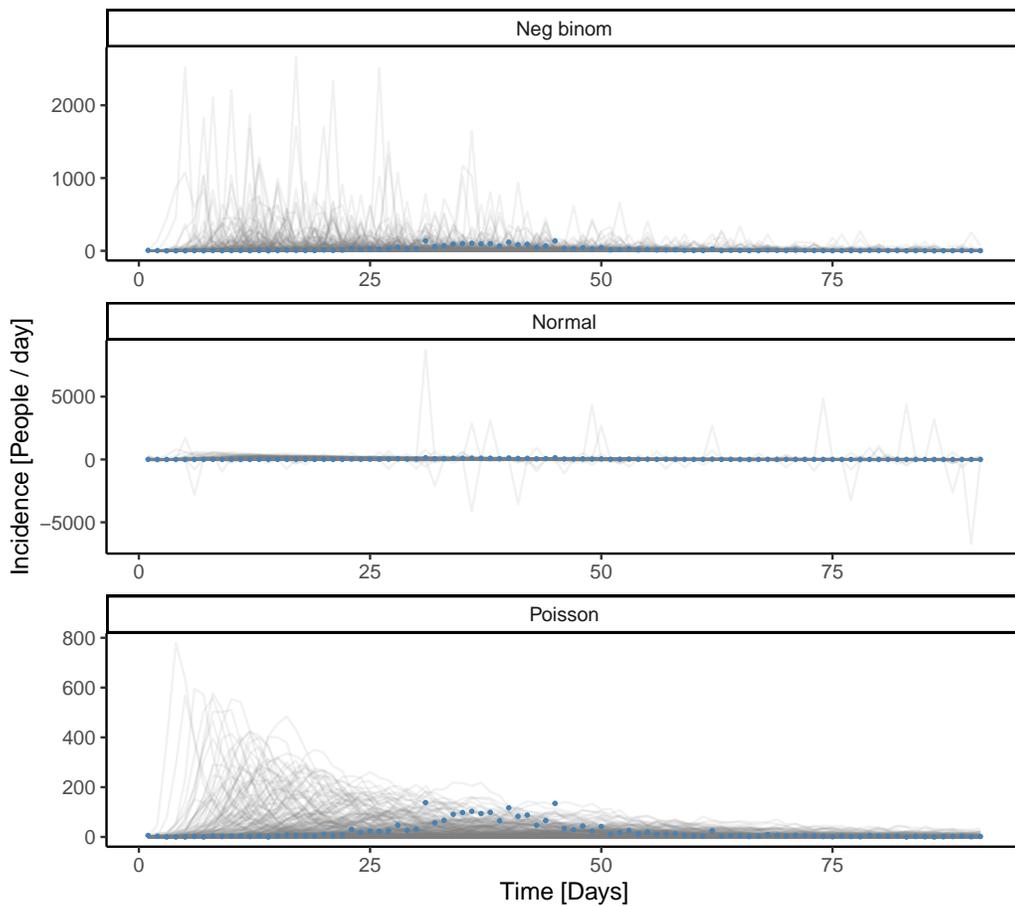


Figure A.1: Prior predictive checks by measurement component

A.2 Parameter space exploration by MCMC algorithm

$$x_i \sim \text{Normal}(\mu_x, 1) \quad (\text{A.1})$$

$$y_i \sim \text{Normal}(\mu_y, 1) \quad (\text{A.2})$$

$$\mu_x \sim \text{Normal}(0, 0.5) \quad (\text{A.3})$$

$$\mu_y \sim \text{Normal}(0, 0.5) \quad (\text{A.4})$$

With the purpose of presenting an intuitive comparison between the Metropolis algorithm (RWM) and Hamiltonian Monte Carlo (HMC), we replicate a simple example provided by McElreath (2020). We assume that we have received bi-dimensional data (Fig A.2). To model this data, we propose a formulation (Eq (A.1)-(A.4)) in which we assume that the measurements in each dimension are independent. For each dimension, we hypothesise that the measurements follow a normal distribution from a common mean and one standard deviation. Therefore, our goal is to estimate the posterior distribution of parameters μ_x and μ_y . Our prior knowledge about these parameters is represented by a normal distribution (Eq (A.3)-(A.4)). The estimation is achieved via RWM and HMC.

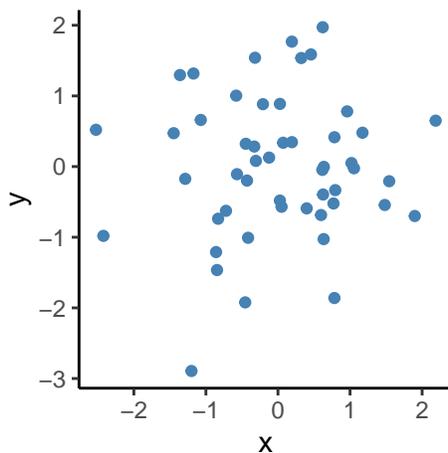


Figure A.2: Toy example data

A.3. Potential scale reduction factor (\widehat{R}) & Effective Sample Size (\hat{n}_{eff})

In Fig A.3, we depict the process of finding and exploring the posterior distribution (the dashed line represents the target distribution’s 99% bounds). In the left-hand side panels, we present the first ten samples generated by the MCMC algorithms from the starting point (red dot). Here, we can notice Metropolis’ random-walk behaviour. In spite of the samples get closer progressively to the target distribution, they follow a haphazard trajectory. In addition to this, five proposals were rejected by Metropolis’ acceptance criterion so that various samples overlap, and we can only distinguish half of them. In stark contrast, HMC samples move directly to the target distribution. Once the chain finds the target, it uses information from parameter space’s *landscape* to construct trajectories. Intuitively, we can conceive the target distribution as a bowl whose bottom represents areas of high plausibility. Here, sample generation corresponds to the process of throwing a marble inside the bowl with some momentum and let that marble explore the bowl’s *curvature* to record the position where the marble loses speed. Recall that the marble loses speed when it reaches the bowl’s bottom. Hamiltonian mechanics play its role by describing, in terms of kinetic and potential energies, the trajectories that the frictionless particle is following. We refer the reader to McElreath (2020) to complement this intuition.

Further, in the right-hand side panels, we present the first 500 draws from each method. Interestingly, they both find the target, but HMC draws spreads more evenly across the posterior distribution. In other words, HMC samples provide a more accurate description of the explored space than RWM.

A.3 Potential scale reduction factor (\widehat{R}) & Effective Sample Size (\hat{n}_{eff})

\widehat{R} is a convergence diagnostic, which compares the between- and within-chain estimates for model parameters and other univariate quantities of interest. If chains have not mixed well, \widehat{R} is larger than 1. It is recommended to run at least *four chains* by default and only using the sample if \widehat{R} is less than 1.01 (Vehtari et al. 2021). Stan reports \widehat{R} , which is the maximum of rank normalized split- \widehat{R} and rank normalized folded-split- \widehat{R} , which works for thick-tailed distributions and is sensitive also to differences in scale.

For each parameter θ , we split each chain from the *sampling phase* in two halves. That is, from **four** chains of 1000 draws each, we obtain **eight**

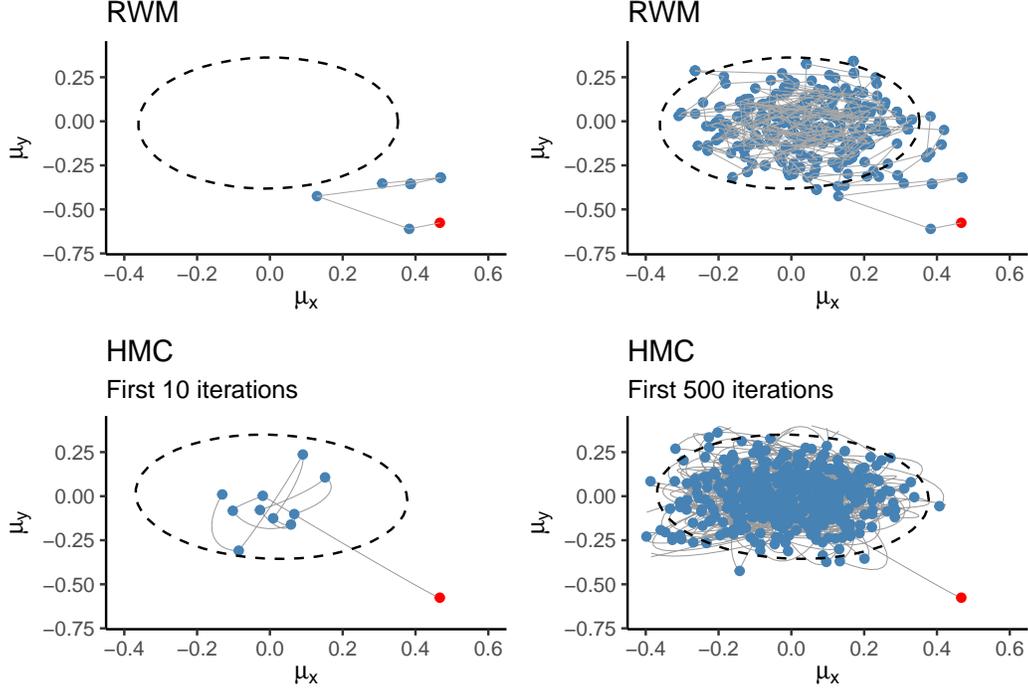


Figure A.3: Parameter space exploration

split chains of 500 draws each. Then, we label the simulations as θ_{ij} ($i = 1, \dots, N; j = 1, \dots, M$), where N is the number of samples per split chain, M is the number of split chains, and $S = NM$ is the total number of draws from all chains. We subsequently transform these simulations to their corresponding rank normalized values z_{ij} . According to Vehtari et al. (2021), we replace each value θ_{ij} by its rank r_{ij} within the pooled draws from all chains. Second, we transform ranks to normal scores using the inverse normal transformation and a fractional offset via Eq (A.5):

$$z_{ij} = \Phi^{-1} \left(\frac{r_{ij} - 3/8}{S - 1/4} \right) \quad (\text{A.5})$$

Using these normal scores, we calculate \hat{R} following the formulation proposed by Gelman et al. (2013). Initially, we compute B and W , the between- and within-sequence variances, respectively:

A.3. Potential scale reduction factor (\widehat{R}) & Effective Sample Size (\widehat{n}_{eff})

$$B = \frac{N}{M-1} \sum_{j=1}^M (\bar{z}_{\cdot j} - \bar{z}_{\cdot\cdot})^2, \text{ where } \bar{z}_{\cdot j} = \frac{1}{n} \sum_{i=1}^n z_{ij}, \bar{z}_{\cdot\cdot} = \frac{1}{M} \sum_{j=1}^M \bar{z}_{\cdot j} \quad (\text{A.6})$$

$$W = \frac{1}{M} \sum_{j=1}^M s_j^2, \text{ where } s_j^2 = \frac{1}{N-1} \sum_{i=1}^n (z_{ij} - \bar{z}_{\cdot j})^2 \quad (\text{A.7})$$

Then, we can estimate $\widehat{\text{var}}^+(\theta|y)$, the marginal posterior variance of the parameter, by a weighted average of W and B :

$$\widehat{\text{var}}^+(\theta|y) = \frac{N-1}{N} W + \frac{1}{N} B \quad (\text{A.8})$$

From Eq (A.7) and (A.8), we obtain the rank normalized split \widehat{R} :

$$\widehat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\theta|y)}{W}} \quad (\text{A.9})$$

To obtain the rank normalized folded-split \widehat{R} , we simply transform the simulations (Eq (A.10)) and then apply the procedure described above (Eq (A.5)-(A.9)).

$$\zeta_{ij} = |\theta_{ij} - \text{median}(\theta)| \quad (\text{A.10})$$

For MCMC draws, we define the estimated effective sample size as

$$\widehat{n}_{\text{eff}} = \frac{MN}{1 + 2 \sum_{t=1}^T \widehat{\rho}_t} \quad (\text{A.11})$$

This quantity requires an estimate of the sum of the correlations ρ up to lag T (the first odd positive integer for which $\widehat{\rho}_{T+1} + \widehat{\rho}_{T+2}$ is negative). The correlation at any specific lag t (Eq (A.12)) depends upon the estimate $\widehat{\text{var}}^+$ and the *Variogram* at each t (Eq (A.13)).

$$\widehat{\rho}_t = 1 - \frac{V_t}{2\widehat{\text{var}}^+} \quad (\text{A.12})$$

Table A.1: Diagnostics

Parameter	\hat{n}_{eff}	\hat{R}
β	1412.859	1.003264
ρ	1713.404	1.001340
I(0)	1460.322	1.004549

$$V_t = \frac{1}{M(N-t)} \sum_{j=1}^M \sum_{i=t+1}^N (z_{i,j} - z_{i-t,j})^2 \quad (\text{A.13})$$

We use the term *bulk effective sample size* to refer to the effective sample size based on the rank normalized draws. To ensure reliable estimates of variances and autocorrelations needed for \hat{R} and \hat{n}_{eff} , Vehtari et al. (2021) recommend that the rank-normalized effective sample size must be greater than 400 (100 per chain), conditions that hold for our estimated parameters (Table A.1).

A.4 Synthetic data

This section illustrates the Bayesian workflow described in the main text using synthetic data. First, we generate incidence data (Fig A.4) via the SEIR model and a Poisson measurement component. To generate this data, we assume a population size (N) of 5234, of which 30% have previous immunity ($R(0) = 0.3N$). Both the epidemiological delays (σ^{-1} and γ^{-1}) are set to 2 and the reporting fraction (ρ) to 77%. Furthermore, it is assumed that the outbreak is triggered by two infectious individuals ($I(0) = 2$). Lastly, the underlying basic reproduction number (\mathfrak{R}_0) governing these dynamics is set to 2.58.

A.4.1 Case 1

The purpose of this example concerns the recovery of \mathfrak{R}_0 from the calibration of the SEIR model to the synthetic data. To begin with, we classify each parameter in the model as assumed (fixed) or estimated. The latter class will be subject to inference via MCMC sampling. Initially, we consider the effective contact rate (β) and the number of initial infectious individuals

A.4. Synthetic data

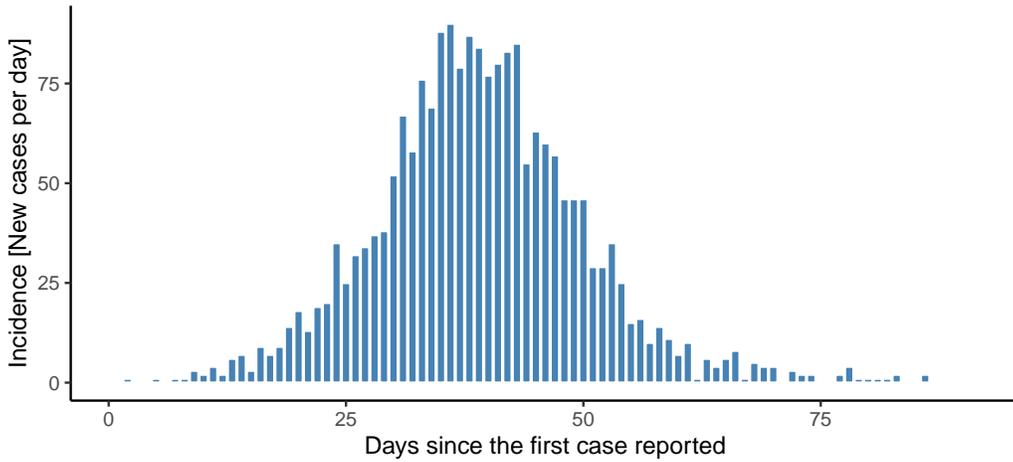


Figure A.4: Synthetic incidence

($I(0)$) as estimated. Further, we assume no underreporting ($\rho = 1$) and no previous immunity ($R(0) = 0$). For the two estimated parameters, we adopt the priors described in the main text. The remaining parameters are fixed to their true values.

A.4.1.1 Prior predictive checks

Then, we check that our prior information captures the data (Fig A.5). Based on the results below, we accept the prior.

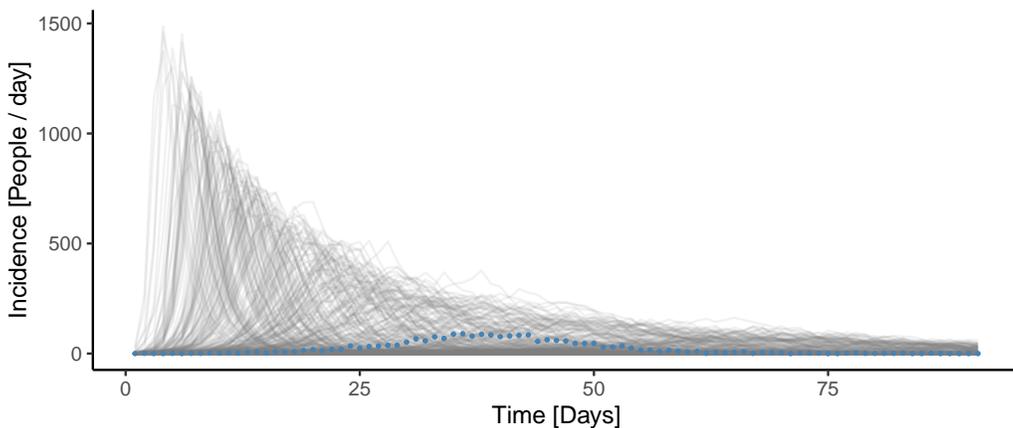


Figure A.5: Case 1. Prior predictive checks

A.4.1.2 Diagnostics

After checking the prior, we perform the inference process with four Markov chains and 1000 iterations (plus 1000 for warm-up) each, checking for convergence and effective sample sizes. These checks are satisfactory and suggest no pathological behaviour.

```
## Checking sampler transitions treedepth.
## Treedepth satisfactory for all transitions.
##
## Checking sampler transitions for divergences.
## No divergent transitions found.
##
## Checking E-BFMI - sampler transitions HMC potential energy.
## E-BFMI satisfactory.
##
## Effective sample size satisfactory.
##
## Split R-hat values satisfactory all parameters.
##
## Processing complete, no problems detected.
```

A.4.1.3 Posterior predictive checks

However, when we compare the model's fit against the data (Fig A.6), we note that the estimated trajectories (grey contour denotes 95% credible intervals and solid line denotes the mean) tend to overestimate the observed data (blue dots). Therefore, assuming no underreporting and no previous immunity produce an inadequate explanation of the observed data. For this reason, we deem the model *not trustworthy*.

A.4.2 Case 2

Since the specification in Case 1 does not yield satisfactory results, we consider a more complex model. Specifically, $R(0)$ and ρ are no longer assumed as fixed parameters, resulting in a four-unknowns model. We adopt the priors described in the main text. For $R(0)$, we choose a normal prior. The remaining parameters are fixed to their true values.

A.4. Synthetic data

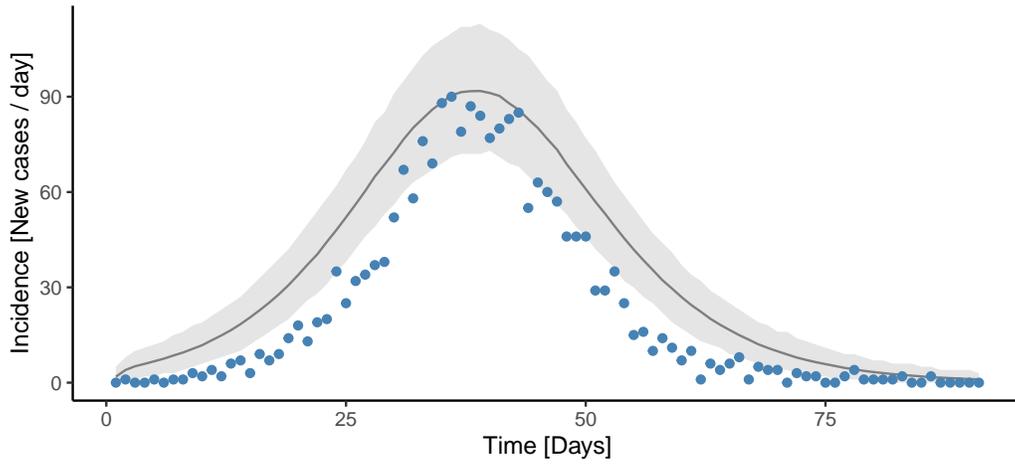


Figure A.6: Case 1 fit

A.4.2.1 Prior predictive checks

Again, we check that our prior information captures the data (Fig A.7). Based on the results below, we accept the prior.

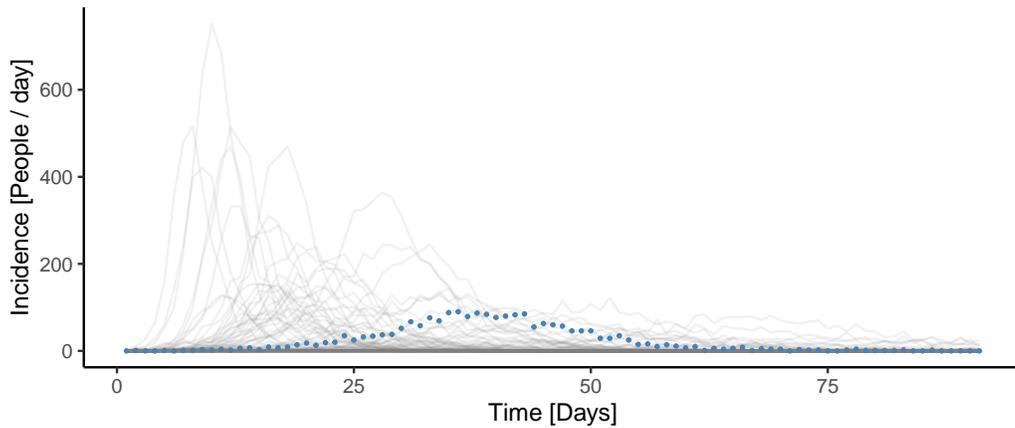


Figure A.7: Case 2. Prior predictive checks

A.4.2.2 Diagnostics (first attempt)

Following prior checks, we run four chains of 1000 iterations (plus 1000 for warm-up) each, checking for convergence and effective sample sizes. Stan diagnostics inform the presence of divergent iterations and \hat{R} values above the recommended threshold, an indication of pathological behaviour that

may lead to biased estimations.

```
## Checking sampler transitions treedepth.
## 33 of 4000 (0.82%) transitions hit the maximum treedepth
## limit of 10, or 2^10 leapfrog steps. Trajectories that
## are prematurely terminated due to this limit will result
## in slow exploration. For optimal performance, increase
## this limit.
##
## Checking sampler transitions for divergences.
## 178 of 4000 (4.45%) transitions ended with a divergence.
## These divergent transitions indicate that HMC is not
## fully able to explore the posterior distribution. Try
## increasing adapt delta closer to 1. If this doesn't
## remove all divergences, try to reparameterize the model.
##
## Checking E-BFMI - sampler transitions HMC potential energy.
## E-BFMI satisfactory.
##
## Effective sample size satisfactory.
##
## The following parameters had split R-hat greater than 1.05:
## rho, par_beta, I0, R0, x[1,1], x[2,1], x[3,1], x[4,1], x[5,1], x[6,1], x[7,1]
## Such high values indicate incomplete mixing and biased
## estimation. You should consider regularizing your model
## with additional prior information or a more effective
## parameterization.
```

Additionally, using the *bayesplot* package, we pinpoint the divergent iterations (red crosses) in the parameter space (Fig A.8).

A.4.2.3 Diagnostics (second attempt)

One strategy to address divergent transitions consists of increasing the *adapt_delta* parameter. This parameter corresponds to the target average proposal acceptance probability during Stan's adaptation period, and increasing it will force Stan to take smaller steps. Consequently, we increase *adapt_delta* from its default value (0.8) to 0.99 and re-run the sampling process. This adjustment eliminated the divergent transitions at the cost of hitting the maximum treedepth threshold. Unlike divergent transitions, this

A.4. Synthetic data

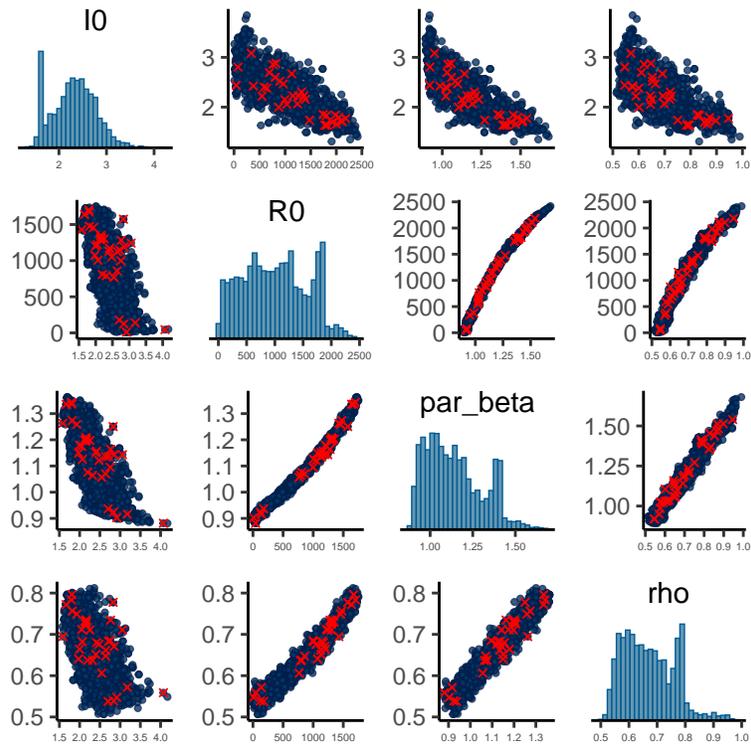


Figure A.8: Case 2.1. Pairs-plot

warning is not a validity concern but an efficiency indicator. Despite the fact that this indicator informs us about the complexity of the posterior explored, we could still use the samples for further analysis should the convergence and efficiency metrics are satisfactory. Moreover, an updated pairs-plot (Fig A.9) highlights the iterations where the algorithm hit the maximum treedepth.

```
## Checking sampler transitions treedepth.
## 323 of 4000 (8.07%) transitions hit the maximum treedepth
## limit of 10, or 2^10 leapfrog steps. Trajectories that
## are prematurely terminated due to this limit will result
## in slow exploration. For optimal performance, increase
## this limit.
##
## Checking sampler transitions for divergences.
## No divergent transitions found.
##
## Checking E-BFMI - sampler transitions HMC potential energy.
## E-BFMI satisfactory.
##
## Effective sample size satisfactory.
##
## Split R-hat values satisfactory all parameters.
##
## Processing complete.
```

A.4.2.4 Diagnostics (third attempt)

However, it is possible to remove this warning by increasing the maximum number of steps allowed per iteration. We also increase the number of iterations. This time the diagnostics are satisfactory and indicate no pathological behaviour.

```
## Checking sampler transitions treedepth.
## Treedepth satisfactory for all transitions.
##
## Checking sampler transitions for divergences.
## No divergent transitions found.
##
## Checking E-BFMI - sampler transitions HMC potential energy.
## E-BFMI satisfactory.
```

A.4. Synthetic data

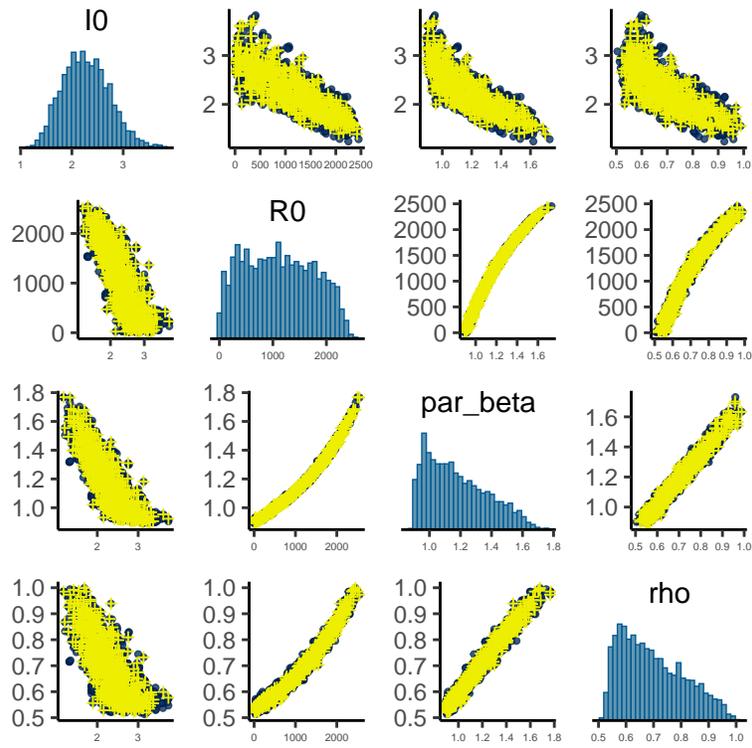


Figure A.9: Case 2.2. Pairs-plot from second attempt

```
##
## Effective sample size satisfactory.
##
## Split R-hat values satisfactory all parameters.
##
## Processing complete, no problems detected.
```

A.4.2.5 Posterior predictive checks

Then, we compare simulated trajectories against the data (Fig A.10). This comparison suggests that the updated specification allows the SEIR model to capture the dynamics of the observed incidence. Therefore, we *accept the model* for parameter estimation.

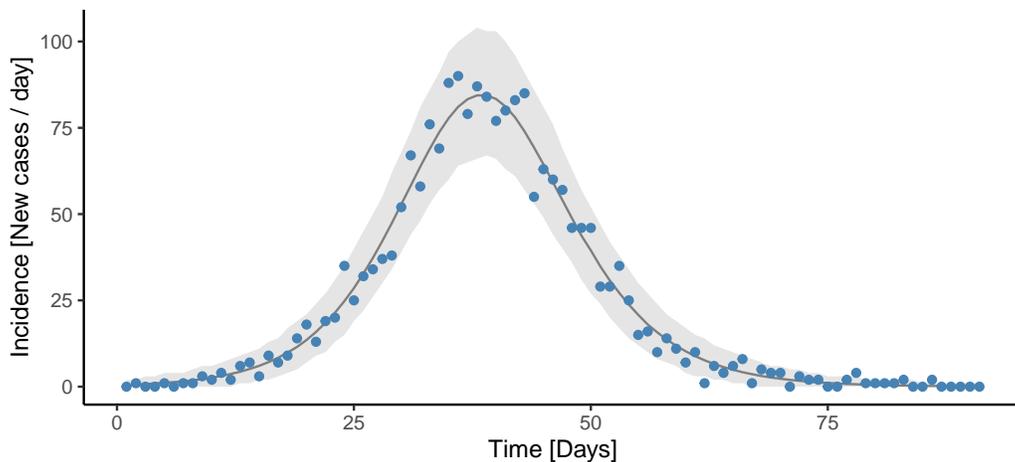


Figure A.10: Case 2. Fit

A.4.2.6 Parameter estimation

Before calculation marginal posterior uncertainty intervals, it is recommended to check for the correlations among parameters (Fig A.11). Here, we can observe this model's complexity given the large degree of correlations occurring in this parameter space. Subsequently, we estimate credible intervals for each parameters, including \mathcal{R}_0 (Table A.2).

A.4. Synthetic data

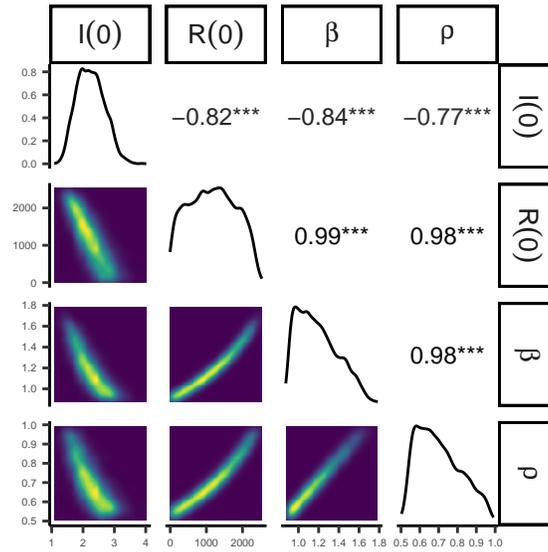


Figure A.11: Case 2. Marginal and joint distributions

Table A.2: Case 2 estimates

Parameter	Mean	2.5%	97.5%
$I(0)$	2.25	1.50	3.09
$R(0)$	1162.41	89.03	2269.57
\mathfrak{R}_0	2.38	1.83	3.19
β	1.19	0.92	1.60
ρ	0.70	0.54	0.94

A.4.3 Case 3

Although Case 2 is computationally satisfactory, the uncertainty around the estimates could be improved with more data. In this context, we assume that new evidence regarding population immunity becomes available. In particular, this new data suggests that around 30% of the population has antibodies against this virus. In light of this information, we repeat the inference process with three estimated parameters $(\beta, \rho, I(0))$, whereas $R(0)$ becomes an assumed parameter. The remaining parameters are set to their true values.

A.4.3.1 Prior predictive checks

As with Cases 1 & 2, we check that our prior information captures the data. Based on the results (Fig A.12), we accept the prior.

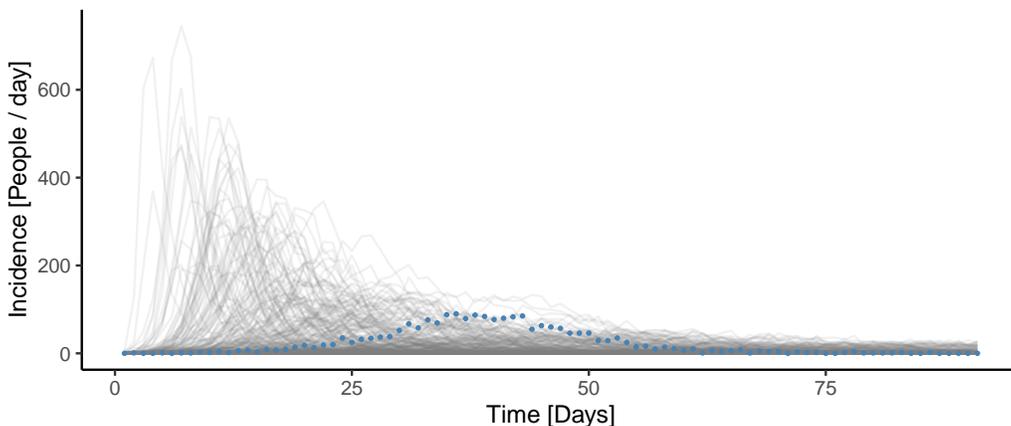


Figure A.12: Case 3. Prior predictive checks

A.4.3.2 Diagnostics

After checking the prior, we perform the inference process with four Markov chains and 1000 iterations (plus 1000 for warm-up) each, checking for convergence and effective sample sizes. These checks are satisfactory and suggest no pathological behaviour.

```
## Checking sampler transitions treedepth.
## Treedepth satisfactory for all transitions.
##
## Checking sampler transitions for divergences.
```

```

## No divergent transitions found.
##
## Checking E-BFMI - sampler transitions HMC potential energy.
## E-BFMI satisfactory.
##
## Effective sample size satisfactory.
##
## Split R-hat values satisfactory all parameters.
##
## Processing complete, no problems detected.

```

A.4.3.3 Posterior predictive checks

Comparing simulated trajectories to actual data (Fig A.13) shows that the SEIR model explains adequately the observed incidence.

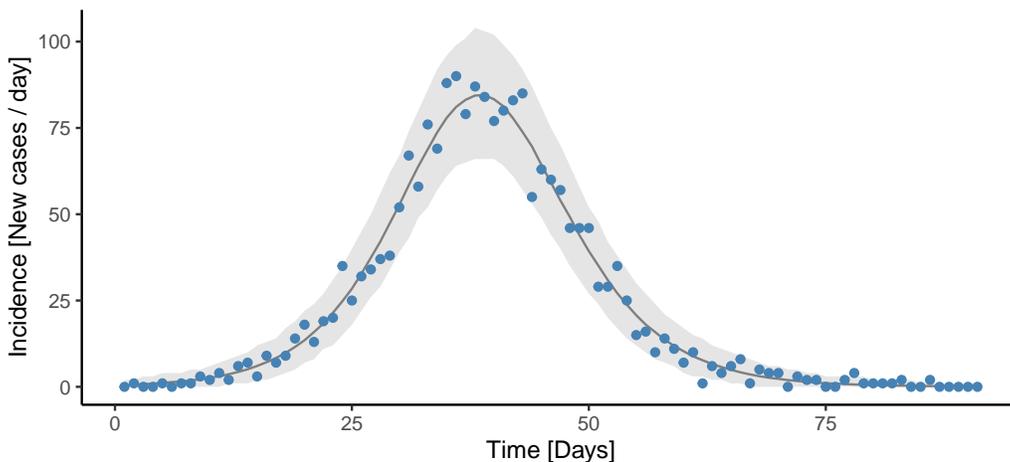


Figure A.13: Case 3 fit

A.4.3.4 Parameter estimates

As with Case 2, we check for the correlations among parameters (Fig A.14). Even though the correlation between $I(0)$ and β slightly increases, correlations between ρ - $I(0)$, and ρ - β significantly decrease in this revised specification. Finally, we estimate credible intervals, which we deem as useful information (Table A.3). Notice that these narrow intervals recover the true values.

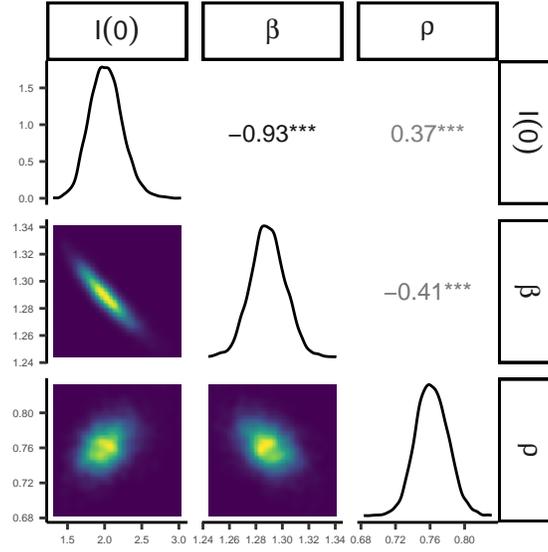


Figure A.14: Case 3. Marginal and joint distributions

Table A.3: Case 3 estimates

Parameter	Mean	2.5%	97.5%
$I(0)$	2.02	1.63	2.48
\mathfrak{R}_0	2.58	2.53	2.63
β	1.29	1.26	1.31
ρ	0.76	0.73	0.80

A.5 Point estimates vs neighbourhoods

Throughout this work, we intentionally focused on probability distributions rather than point estimates. Here, we provide the rationale for such a choice. The reader should recall that one can also approach model calibration via optimisation. That is, using non-linear optimisation routines, one finds the maximum or mode of a likelihood function $\pi(y|\theta)$, which is a point estimate. Around this estimate, we can construct confidence intervals, which are defined in terms of significance and coverage. For instance, a 95% confidence interval means that 95 out of 100 intervals constructed from 100 measurements obtained from an identical process will capture the actual value. In this context, a single measurement corresponds to the entire time series of case counts reported in Cumberland. Using the samples obtained from fitting the SEIR model to this dataset, we display the likelihood function for each parameter (Fig A.15), where the dashed line indicates the mode or Maximum Likelihood Estimate (MLE).

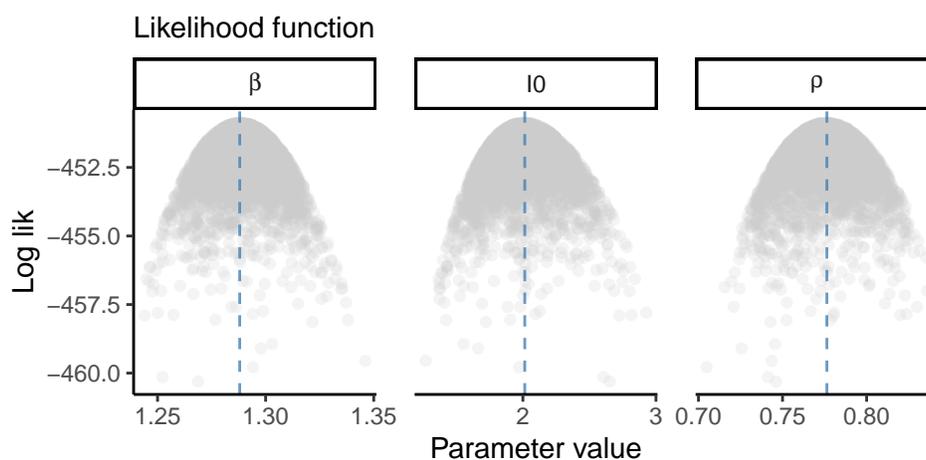


Figure A.15: Likelihood function by parameter

However, we consider whether a point close to the MLE is less important than the MLE itself. For instance, the MLE for β is 1.288... (please note that since this is a point estimate on the real line, after the last digit -8-, there is an additional infinite number of digits). Does this imply that 1.2881... is less useful given that this other point estimate yields a lower likelihood value? Certainly, this is a value judgement. To elaborate on this issue, we explore three point estimates: the mode, the mean of each marginal posterior, and a random point from the posterior distribution (Table A.4). We show these

Table A.4: Point estimates

β	ρ	I(0)	Estimate
1.289201	0.7746664	2.003492	Mean
1.288000	0.7763670	2.014230	Mode
1.265660	0.7794390	2.262070	Random

estimates as lines superimposed over the posterior distribution (Fig A.16).

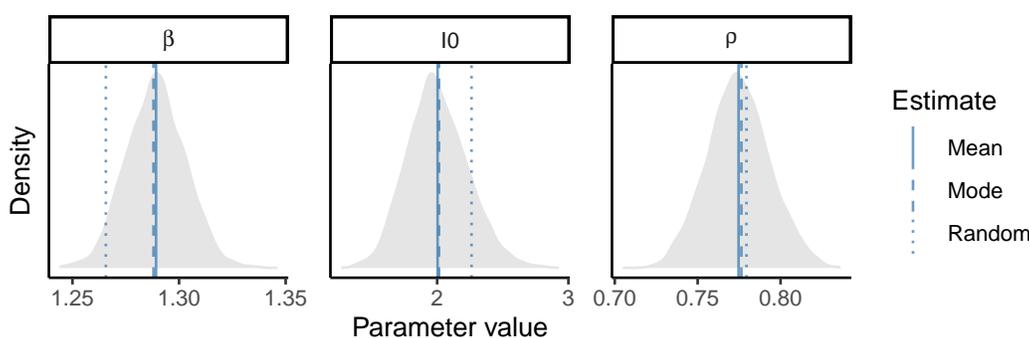


Figure A.16: Point estimates

We then use such point estimates to simulate the SD model and compare the output against the incidence data (Fig A.17). It thus follows that *usefulness* in this context is defined as how accurately the simulated behaviour matches the observed data. To measure this accuracy, we employ three different metrics: the mean absolute scaled error (MASE), the Mean Square Error (MSE), and R^2 . For the MASE and MSE, lower values indicate better performance. Conversely, high values of R^2 indicate more accurate predictions. These metrics, though, tell a contradictory story. On the one hand, the mean produces the most precise estimate according to the MASE and the MSE. On the other hand, the random point results in the most precise estimate according to R^2 . Interestingly, the mode underperforms under the three metrics (Table A.5).

On the contrary, from a Bayesian perspective, ascertaining which point estimate stands out above the others is uninteresting. The critical feature lies in their location. That is, they form a common neighbourhood, which we refer to as the posterior distribution $\pi(\theta|y)$. In this case, this distribution is continuous, so the probability of any given point is zero. Thus, from a Bayesian viewpoint, we are interested in finding which regions (collections of

points) within the parameter space are consistent with the data. This feature highlights a key difference between optimisation algorithms and MCMC methods: their targets. Whereas the former aims for the point of highest density, the latter aims for concentration of probability mass, also known as the “typical set”. In lower dimensions, they tend to agree, but as the number of parameter increases, the difference becomes evident. Colloquially, this is known as the *Curse of dimensionality*. To complement this argument, we refer the reader to Betancourt’s paper (Betancourt 2018) and Carpenter’s case study¹ to build intuition about this complex issue.

Consequently, statistics such as means, medians, standard deviations, and quantiles are merely used for descriptive purposes. Specifically, quantiles describe where in the parameter space the collection of plausible point estimates concentrate. These quantiles known as credible intervals quantify the uncertainty of the unobserved parameter values. For instance, a 95% credible interval implies that we *believe* that there is a 95% chance of finding the actual value within those bounds. Nevertheless, given that we have defined model calibration as a Bayesian inference process, which quantifies inferences by an entire distribution, no point estimate or interval characterises that distribution.

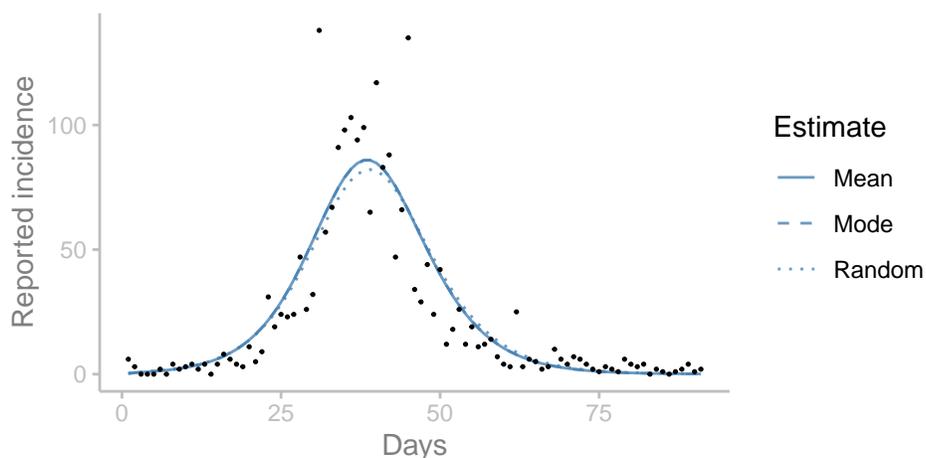


Figure A.17: Fit comparison by point estimate

¹<https://mc-stan.org/users/documentation/case-studies/curse-dims.html>

Table A.5: Performance metrics

Estimate	MASE	MSE	R^2
Mean	0.7283700	220.1650	0.8030856
Mode	0.7301842	220.7375	0.8025443
Random	0.7615599	236.3105	0.7920332

A.6 Mean Absolute Scaled Error

Hyndman offers the following definition²:

The MASE was proposed by Hyndman and Koehler (2006) as a generally applicable measurement of forecast accuracy. They proposed scaling the errors based on the in-sample Mean Absolute Error (MAE) from the naïve forecast method. Using the naïve method, we generate one-period-ahead forecasts from each data point in the sample. Accordingly, a scaled error is defined as:

$$q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|} \quad (\text{A.14})$$

where the numerator e_t is the absolute forecast error ($|\hat{y}_t - y_t|$) for a specific time. Here, \hat{y}_t denotes the predicted data from the fitted model and y_t the actual data. Further, n represents the number of data points. The denominator is the MAE from the one-step “naive forecast method”, defined as the actual value (y_t) minus the forecast value (y_{t-1}) for $t > 1$.

Thus, the mean absolute scaled error is:

$$MASE = \text{mean}(|q_t|) \quad (\text{A.15})$$

MASE is a scale-independent metric defined if there are zero values. A scaled error is less than one if it arises from a better forecast than the naïve forecast. Conversely, it is greater than one if the forecasts worse than the average one-step, naïve forecast computed in-sample.

Consequently, we estimate the MASE for each simulated trajectory from the SEIR model fitted to Cumberland’s data and present its distribution (Fig

²<https://robjhyndman.com/papers/foresight.pdf>

A.18).

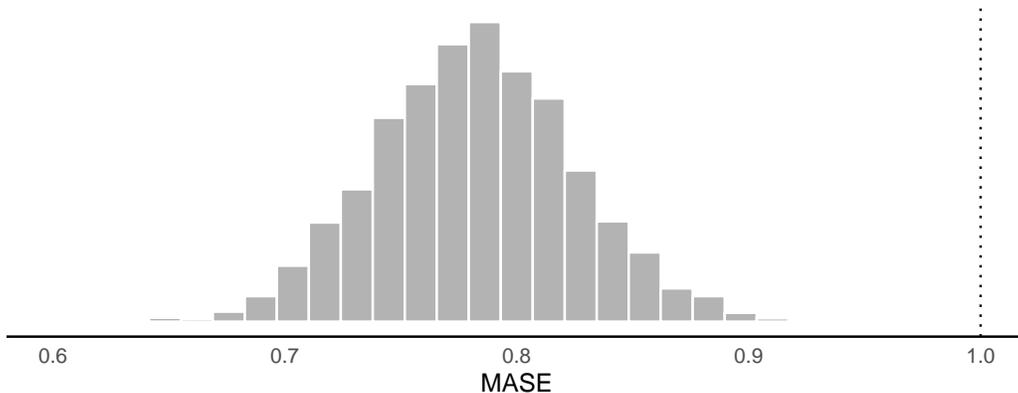


Figure A.18: MASE distribution

A.7 Performance comparison between HMC and RWM

This section complements the main text by showing differences in computational time and trace plots when fitting the SEIR model to Cumberland’s incidence data using two MCMC methods: RWM and HMC. We perform the calibration process under six scenarios, which differ in the number of iterations (100, 200, 500, 1000, 1500 and, 2000). For instance, the 100-iterations scenario indicates that we allocate 100 iterations to the *burn-in/warm-up* phase and 100 iterations to the *sampling* phase. It should be noted that we only use the draws from the *sampling phase* in trace plots. Conversely, we take into account the *burn-in/warm-up* phase to measure execution times.

A.7.1 Computational time

This metric refers to the time the practitioner awaits the results. This evaluation ignores whether the Markov chains converge to the posterior distribution. The results (Fig A.19) indicate that, in all scenarios, HMC takes less time to produce an equal amount of samples than RWM does. Nevertheless, this metric is not free of confounders. It is not known whether these differences are due to the methods per se or if, on the contrary, the observed gap stems from performance discrepancies in the technological implementations.

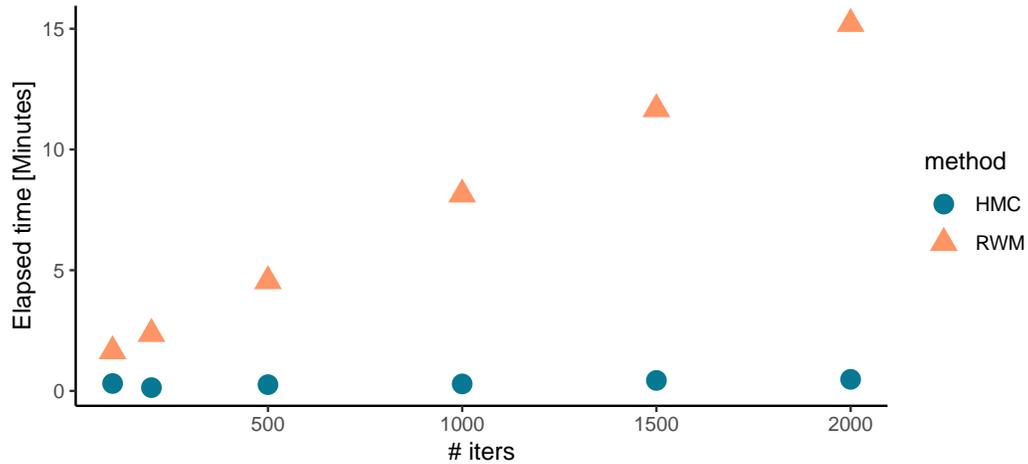


Figure A.19: Computational time by MCMC method

A.7.2 Trace plots

Before estimating diagnostic quantities (shown in the main text), it is always recommended to check trace plots so as to detect convergence issues. The trace plots in this section show the draws obtained in the *sampling phase*. As mentioned above, we allocate an equal number of iterations for the *burn-in/warm-up* and *sampling* phases. In the first scenario, MCMC samplers return 400 samples (4 chains * 100 samples from the sampling phase) that describe the target (posterior) distribution. In these graphs (Fig A.20-A.25), we observe that even from 100 warm-up samples, HMC apparently reaches convergence. Conversely, RWM seems to require at least 1500 burn-in draws for the chains to converge.

A.7. Performance comparison between HMC and RWM

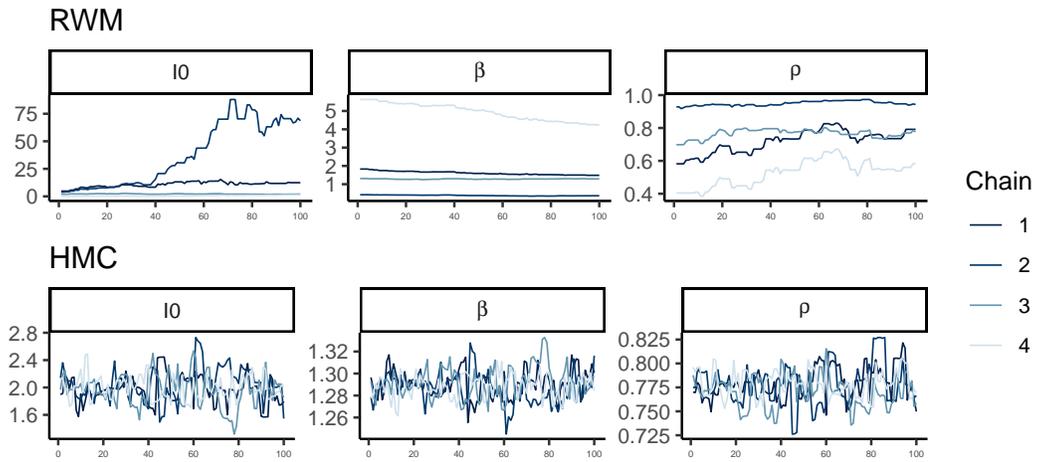


Figure A.20: Trace plots of the 100-iterations scenario by method

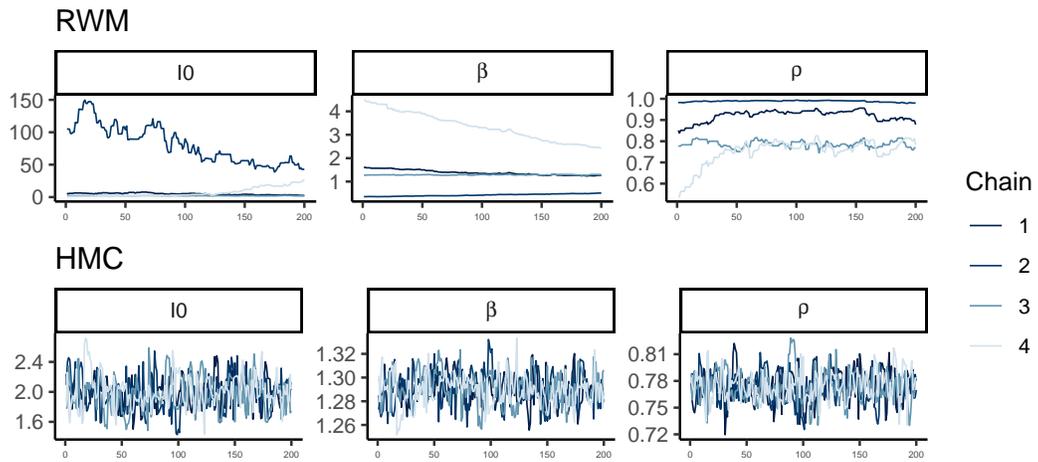


Figure A.21: Trace plots of the 200-iterations scenario by method

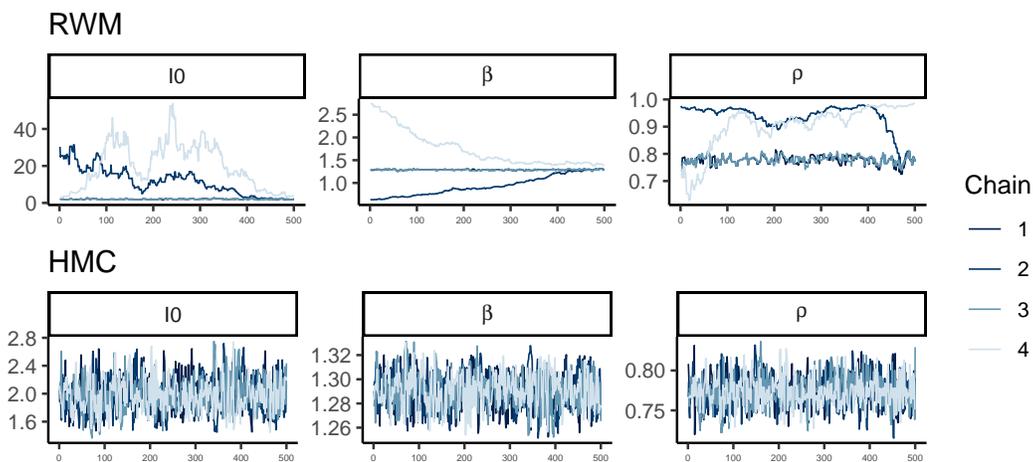


Figure A.22: Trace plots of the 500-iterations scenario by method

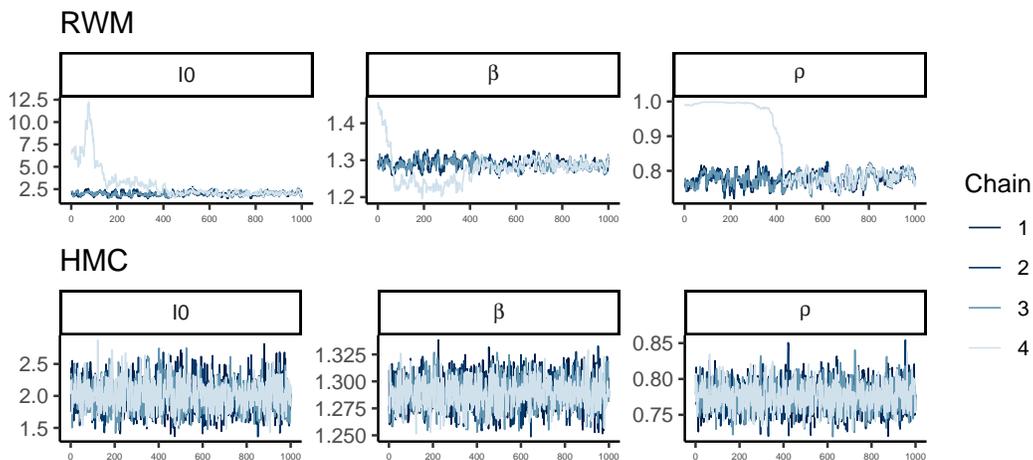


Figure A.23: Trace plots of the 1000-iterations scenario by method

A.7. Performance comparison between HMC and RWM

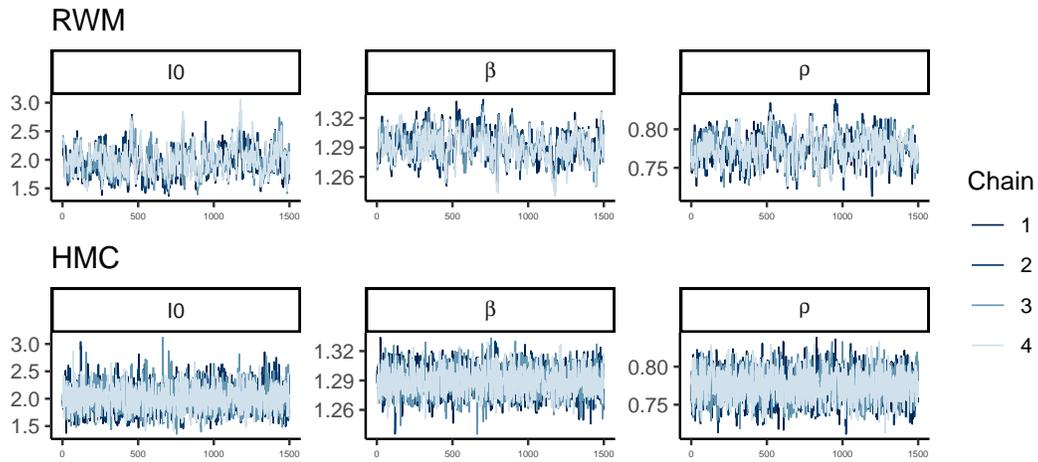


Figure A.24: Trace plots of the 1500-iterations scenario by method

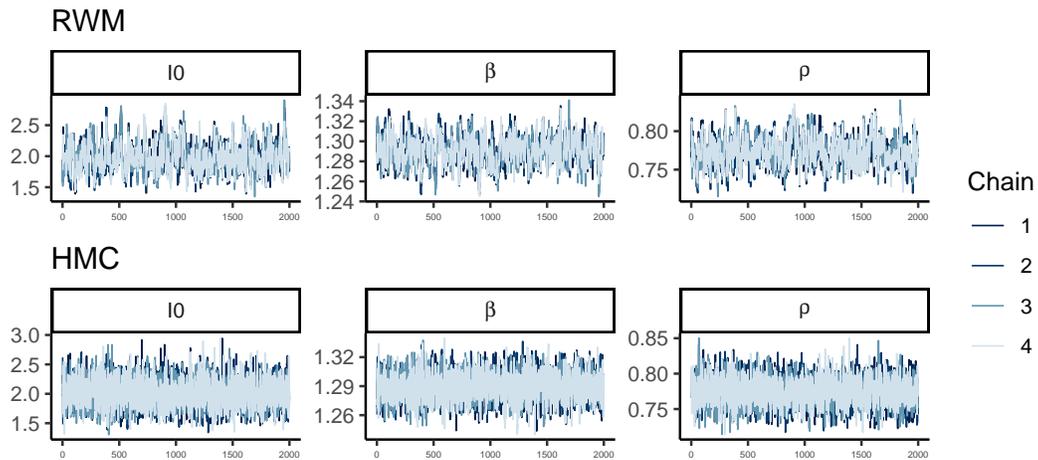


Figure A.25: Trace plots of the 1500-iterations scenario by method

Appendix B

Supplementary Information (Ch. 3)

B.1 Synthetic data

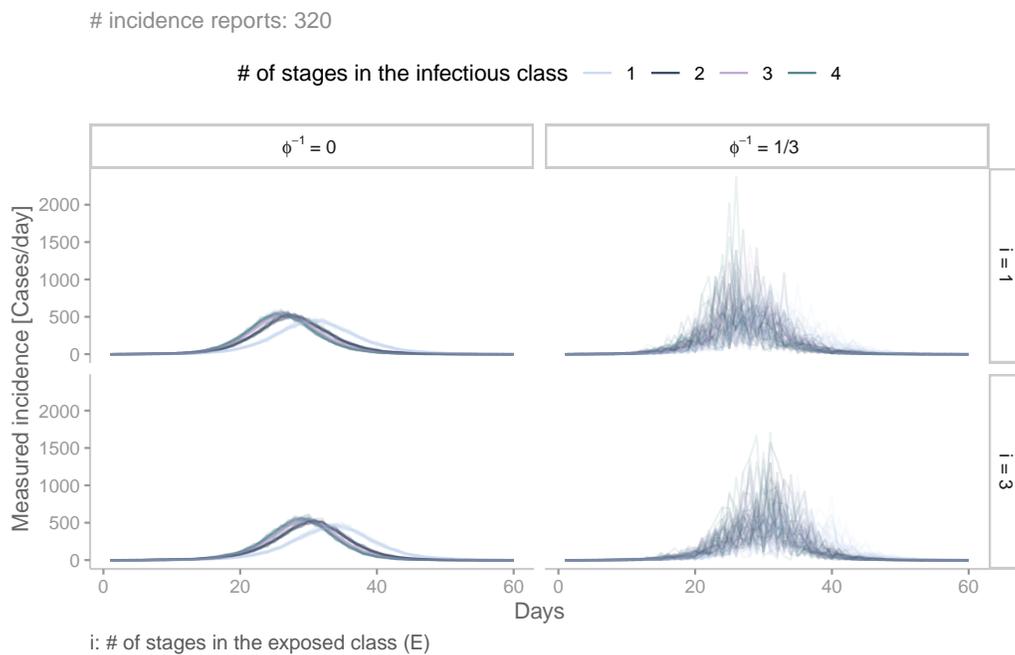


Figure B.1: Synthetic incidence data by measurement component and delay distribution. There are 80 time series per panel.

B.2 Fitting high-fidelity D^{1j}

This section illustrates the process of fitting various parameterisations of the SE^1I^jR model (M^{1j}) to **high-fidelity** D^{1j} incidence reports. We mean by *parameterisation* the decision of categorising model parameters as either *unknown* or *assumed*. For the unknown parameters, we construct prior distributions, which will be eventually updated in light of the data via HMC sampling, resulting in a posterior distribution. On the other hand, assumed parameters are fixed at their true values.

B.2.1 Three-unknown parameterisation (traditional)

We assume that candidates from this parameterisation have three unknowns: β , ρ , and I_0^1 . Fig B.2 shows their prior distributions.

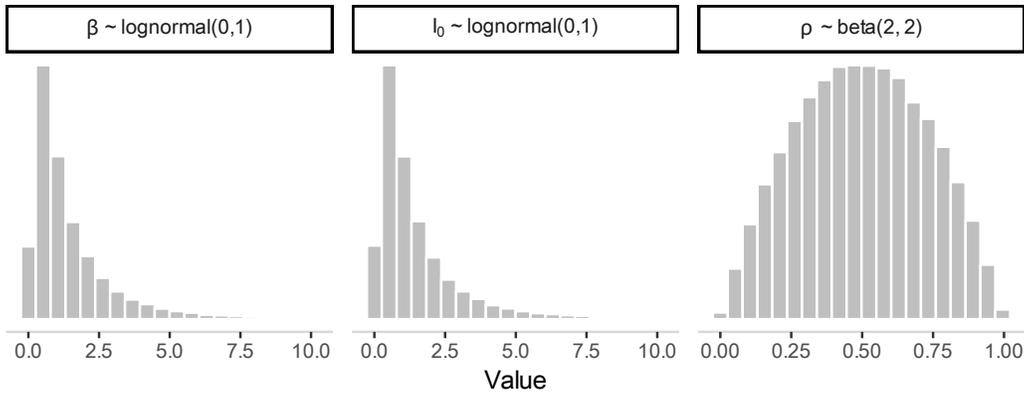


Figure B.2: Prior distribution

B.2.1.1 Inference

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) to each of the 80 high-fidelity D^{1j} incidence reports.

B.2.1.1.1 Incidence fit

Fig B.3 compares actual (points) and simulated (lines) latent incidence. For brevity reasons, we only show two candidate models (M^{11} and M^{14}) matching two incidence time-series of different origin (D^{11} and D^{14}).

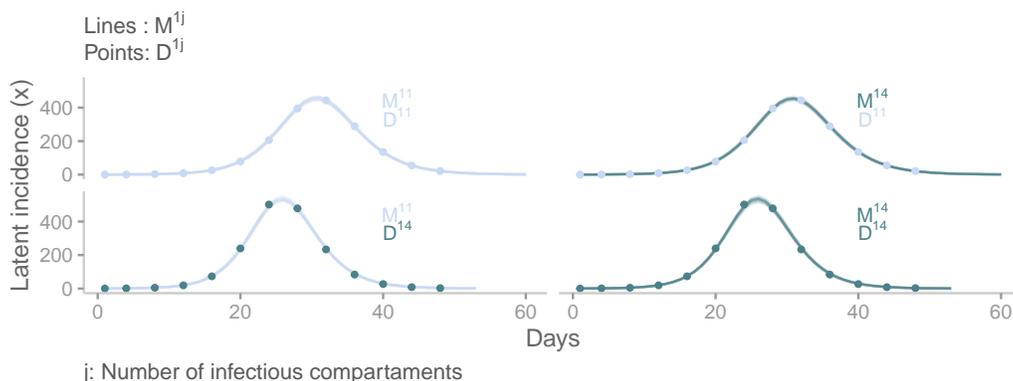


Figure B.3: Incidence fit

B.2.1.1.2 MASE

To persuade the reader that all models fit the data equally well, we draw on the mean absolute scaled error (MASE). This quantity is a measure of the accuracy of forecasts, well-suited for time-series. Therefore, we employ the MASE to compare each of the 4000 simulated latent incidences against its true counterpart (x) and observed incidence (y). By simulated latent incidences, we refer to the process of plugging the samples of a posterior distribution into an ODE model to obtain incidence trajectories via simulation. We summarise the results via histograms (Fig B.4). The left column (of panels) contains the comparison between simulated and actual latent incidences. The right column of panels displays the comparison between simulated latent incidence and the observed incidence. Overall, there is no noticeable variation in the histograms as the fitting model changes (increasing j).

B.2.1.1.3 Parameter estimates

Fig B.5 shows parameter estimates (error bars) calculated from posterior distributions (samples). These samples were obtained from four model candidates fitted to two incidence reports of different origin (D^{11} and D^{14}). Error bars correspond to 95% credible intervals. We show \mathfrak{R}_0 instead of β , given that they are directly proportional and the former is our quantity of interest.

B.2.1.1.4 Joint distribution

Fig B.6 presents an example of a joint posterior distribution. Specifically,

B.2. Fitting high-fidelity D^{1j}

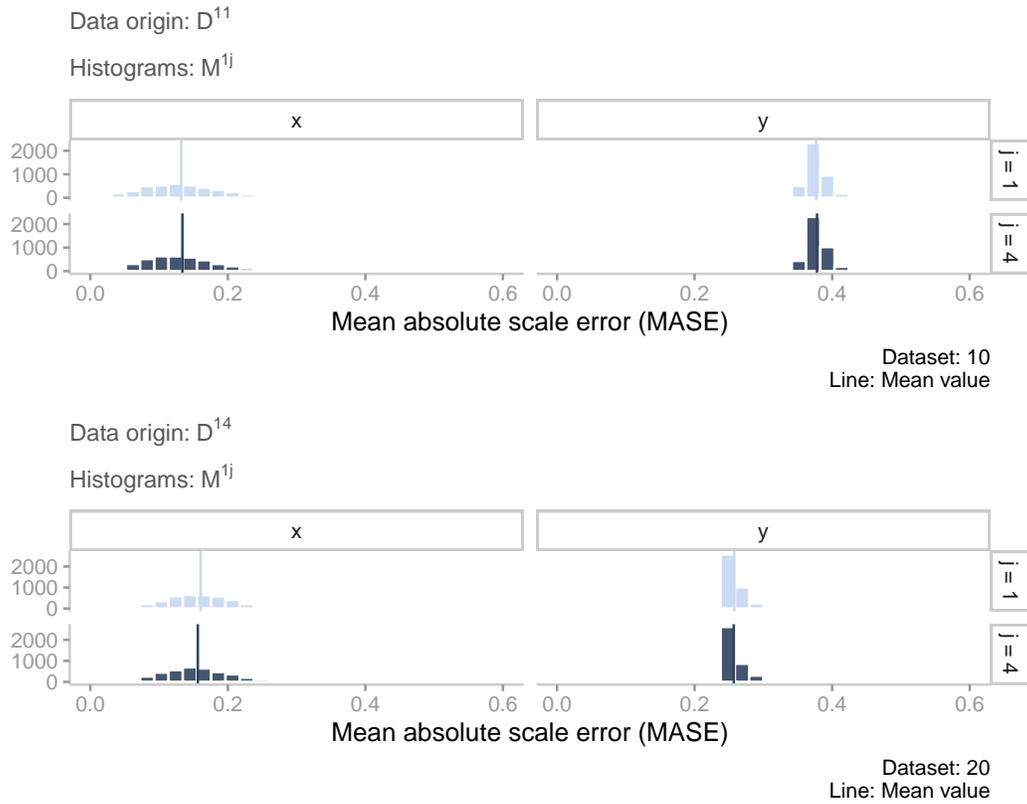


Figure B.4: Fit scores by candidate model and type of data. x denotes latent incidence, whereas y indicates observed incidence. Vertical line denotes the mean.

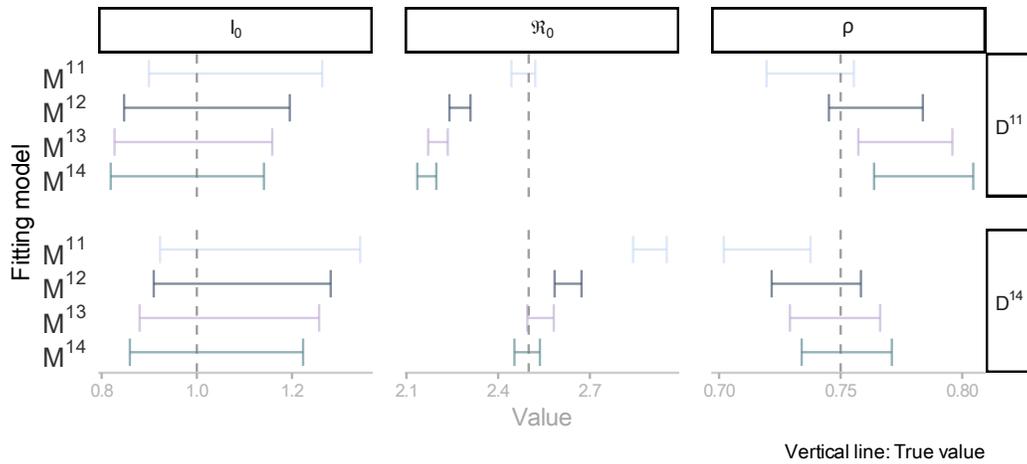


Figure B.5: Parameter estimates. Panel headings on the right-hand side indicate the data's origin.

this posterior distribution was derived from fitting the M^{13} candidate to one D^{13} incidence report. Notice the correlation among parameters.

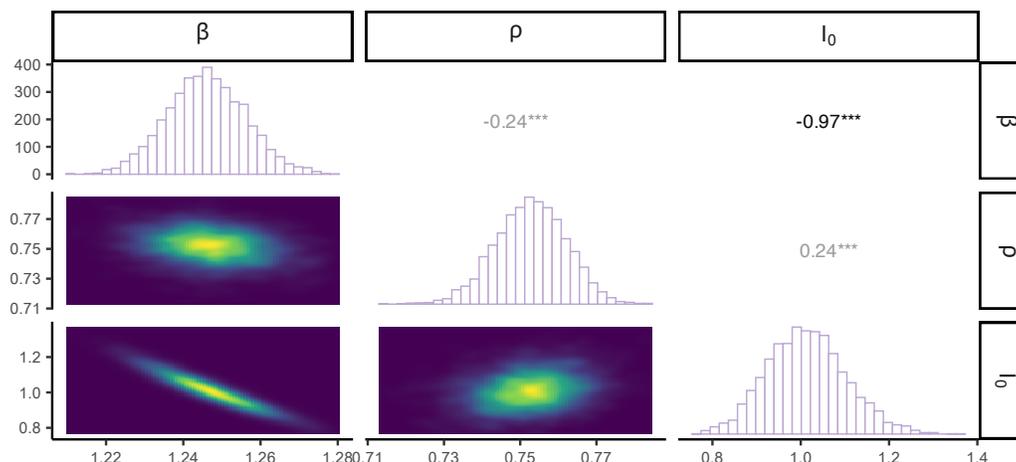


Figure B.6: Joint distribution

B.2.1.1.5 Coverage

Table B.1 summarises the accuracy of estimates by model candidate and data origin. For instance, the first row corresponds to the accuracy of candidate M^{11} when fitting 20 D^{11} incidence reports. In particular, the third column of this row indicates that 19 out of the 20 95% credible intervals cover the true \mathfrak{R}_0 value. Namely, 95% coverage.

B.2.1.1.6 MLE criterion

Each bar (column) in Fig B.7 represents the number of times that a particular model candidate attains the largest likelihood score for a given set of incidence reports. Recall that each set of reports comprises 20 time-series. For instance, the first column (left to right) in the first panel (top-left) indicates that M^{11} outperforms its competitors 12 out of 20 times (60%) in fitting D^{11} incidence reports.

B.2.2 Four-unknown parameterisation

We assume γ as an additional unknown with respect to the previous parameterisation. Fig B.8 correspond to its prior distribution.

B.2. Fitting high-fidelity D^{1j}

Table B.1: Traditional parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0
11	11	95%	100%	95%
11	12	0%	30%	95%
11	13	0%	0%	95%
11	14	0%	0%	90%
12	11	0%	15%	90%
12	12	90%	100%	95%
12	13	5%	95%	95%
12	14	0%	75%	95%
13	11	0%	5%	85%
13	12	0%	95%	90%
13	13	95%	100%	100%
13	14	35%	100%	100%
14	11	0%	0%	85%
14	12	0%	60%	90%
14	13	25%	90%	95%
14	14	85%	95%	95%

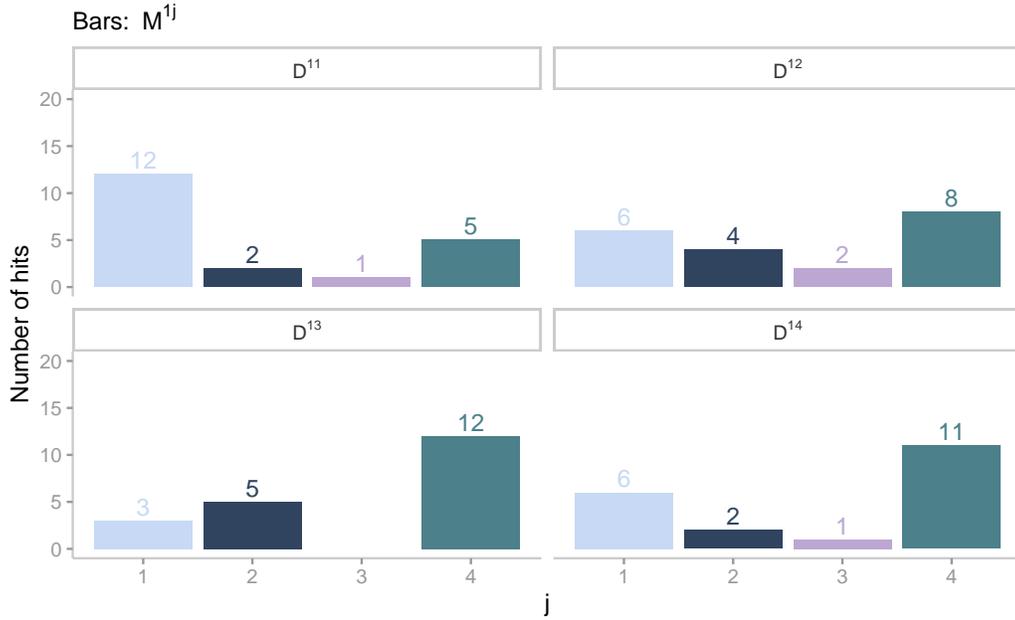


Figure B.7: Score summary

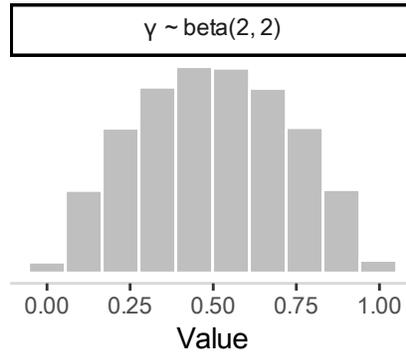


Figure B.8: Recovery rate's prior distribution

B.2.2.1 Inference

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from this parameterisation to each of the 80 high-fidelity D^{1j} incidence reports.

B.2.2.1.1 Incidence fit

Fig B.9 compares actual (points) and simulated (lines) latent incidence. For brevity reasons, we only show two candidate models (M^{11} and M^{14}) matching two incidence time-series of different origin (D^{11} and D^{14}).

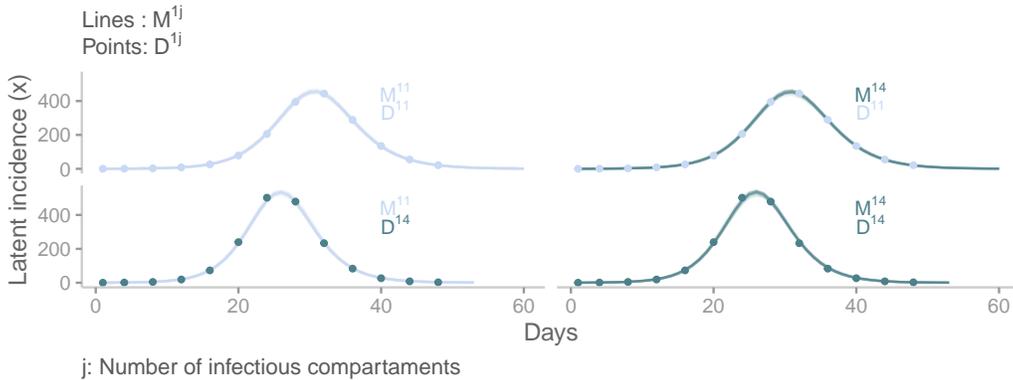


Figure B.9: Incidence fit

B.2.2.1.2 Parameter estimates

Fig B.10 shows parameter estimates (error bars) calculated from posterior distributions (samples). These samples were obtained from four model

B.2. Fitting high-fidelity D^{1j}

candidates fitted to two incidence reports of different origin (D^{11} and D^{14}). Error bars correspond to 95% credible intervals.

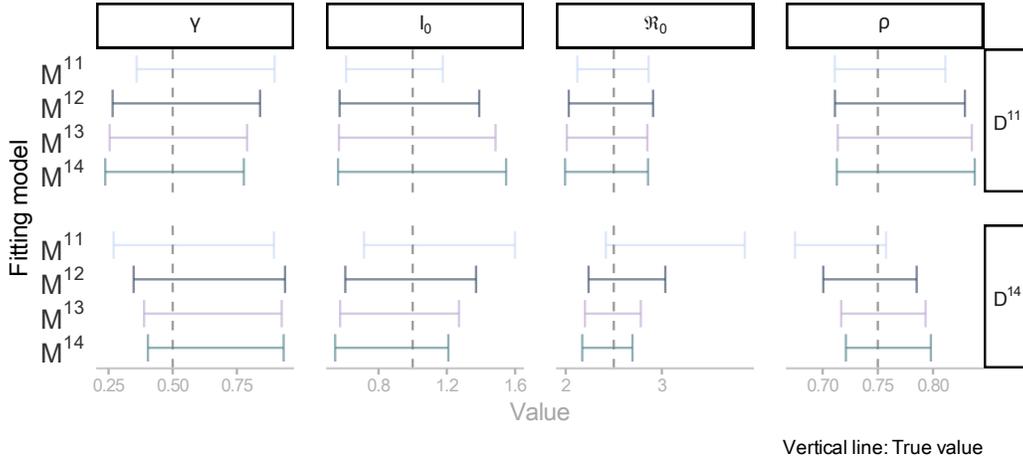


Figure B.10: Parameter estimates. Panel headings on the right-hand side indicate the data's origin.

B.2.2.1.3 Joint distribution

Fig B.11 presents an example of a joint posterior distribution. Specifically, this posterior distribution was derived from fitting the M^{12} candidate to one D^{12} incidence report. Notice the perfect correlation between β and γ .

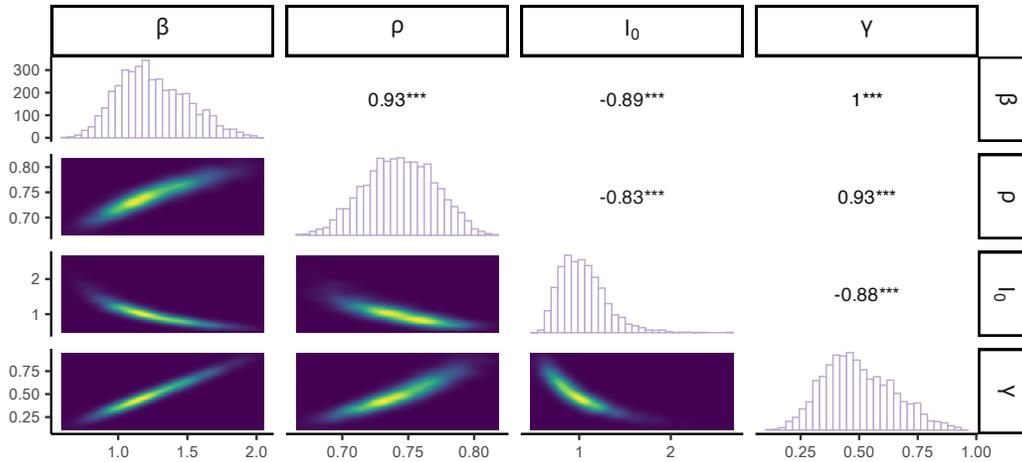


Figure B.11: Joint distribution

B.2.2.1.4 Recovery rate (γ) estimates

We argue that the four-unknown parameterisation is unidentifiable. To support this claim, we compare the recovery's rate prior against the posterior (Fig B.12) distributions obtained from four candidates fitted to one D^{33} report. Notice that the prior and the posterior are remarkably similar. In other words, the data does not provide enough information to update the recovery rate's prior distribution.

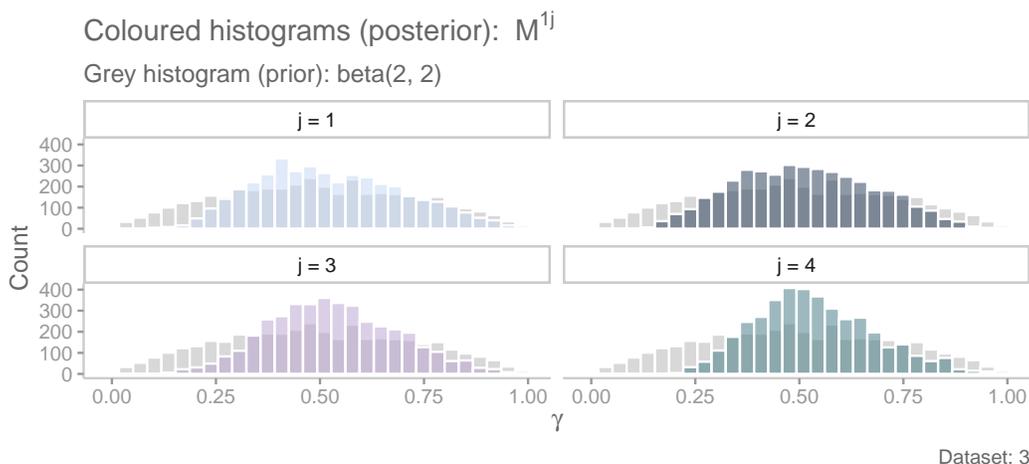


Figure B.12: Comparison between the recovery rate's prior distribution (grey histograms) against four marginal posterior distributions obtained from each candidate (coloured histograms).

B.2.2.1.5 Coverage

Table B.2 summarises the four-unknown parameterisation's coverage.

B.2.2.1.6 \mathfrak{R}_0 vs τ

Fig B.13 shows the unravelled linear relationship between the basic reproduction number and the mean generation time.

B.2.3 Three-unknown parameterisation (alternative)

The *alternative parameterisation* refers to the algebraic manipulation of the SE^iI^jR framework so as to obtain a set equations wherein the basic reproduction number (\mathfrak{R}_0) and the mean generation time (τ) are explicit parameters of the model. In doing so, β and γ become functions of other

B.2. Fitting high-fidelity D^{1j}

Table B.2: Four-unknown parameterisation's coverage table

D^{ij}	M^{ij}	\Re_0	ρ	γ^{-1}	I_0
11	11	100%	100%	100%	100%
11	12	95%	100%	100%	100%
11	13	65%	100%	70%	100%
11	14	50%	100%	60%	95%
12	11	95%	100%	95%	100%
12	12	100%	100%	100%	100%
12	13	95%	100%	95%	100%
12	14	95%	100%	90%	100%
13	11	95%	100%	95%	95%
13	12	95%	100%	95%	95%
13	13	100%	100%	100%	95%
13	14	100%	100%	100%	95%
14	11	70%	95%	80%	95%
14	12	95%	100%	90%	90%
14	13	100%	100%	95%	100%
14	14	100%	100%	100%	100%

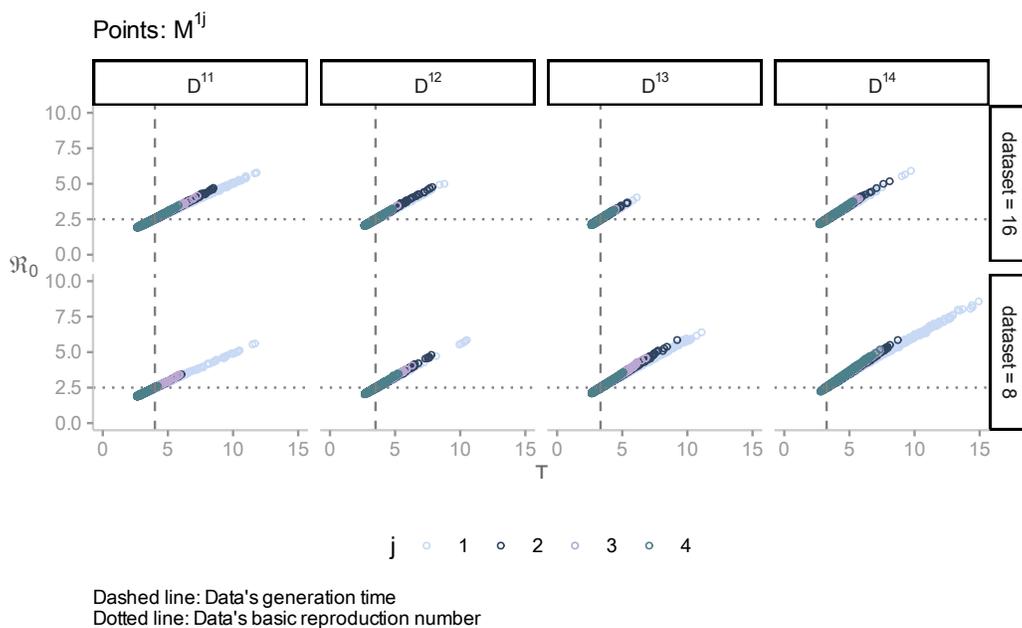


Figure B.13: Scatter plot

variables rather than parameters. In these parameterisation, we assume the basic reproduction number's inverse (\mathcal{R}_0^{-1}), the reporting rate (ρ) and initial number of infectious individuals (I_0^1) as unknowns. Since model candidates are fitting outbreak-like incidence data, it is warranted to assume that the basic reproduction number is higher than one ($\mathcal{R}_0 > 1$). This observation implies that the magnitude of its inverse (\mathcal{R}_0^{-1}) between 0 and 1. Consequently, we formulate a prior that is consistent with such a constraint (Fig B.14).

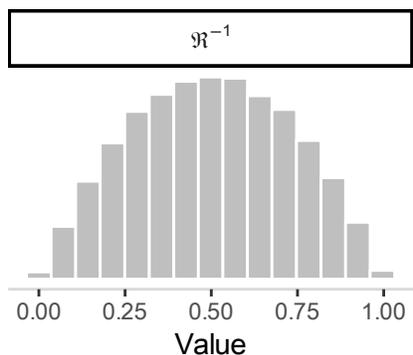


Figure B.14: Basic reproduction number's prior distribution

B.2.3.1 Inference

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from this parameterisation to each of the 80 high-fidelity D^{1j} incidence reports.

B.2.3.1.1 Incidence fit

Fig B.15 compares actual (points) and simulated (lines) latent incidence.

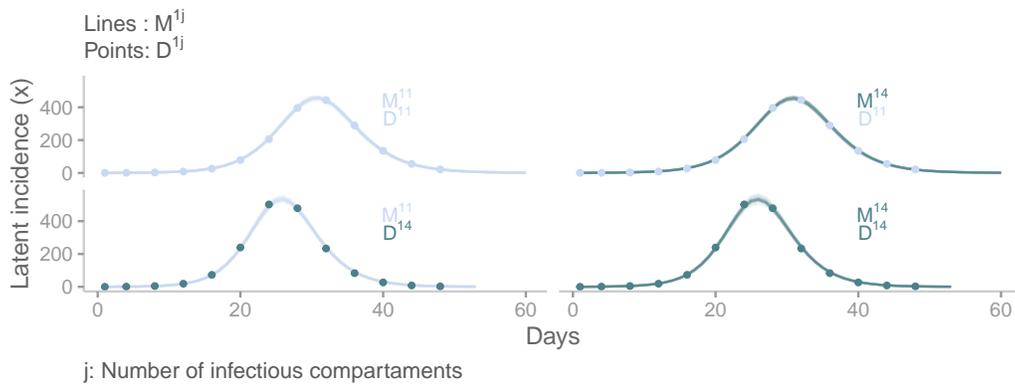


Figure B.15: Incidence fit

B.2.3.1.2 Parameter estimates

Fig B.16 shows parameter estimates (error bars) from four model candidates from this alternative parameterisation fitted to two incidence reports of different origin (D^{11} and D^{14}).

B.2.3.1.3 Joint posterior distribution

Fig B.17 presents an example of a joint posterior distribution. Specifically, this posterior distribution was derived from fitting M^{11} to one D^{11} incidence report.

B.2.3.1.4 Coverage

Table B.3 summarises the alternative parameterisation's coverage.

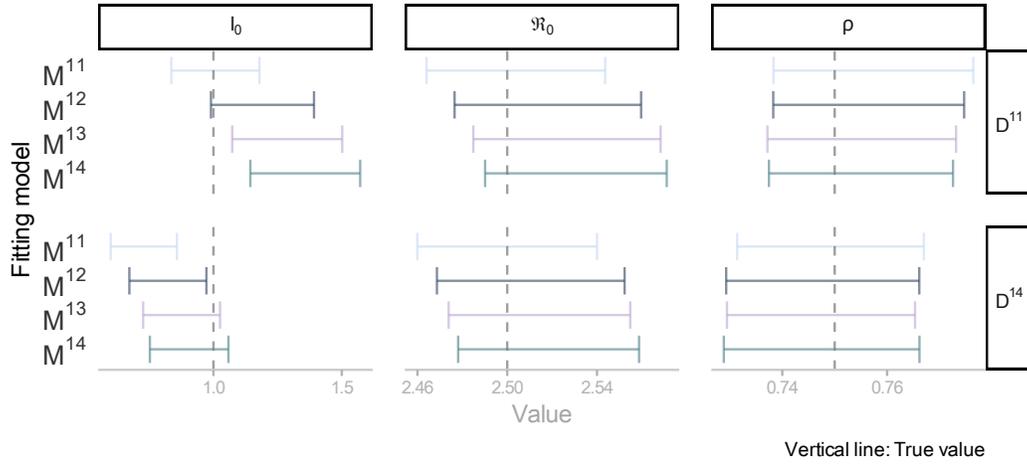


Figure B.16: Parameter estimates. Panel headings on the right-hand side indicate the data's origin.

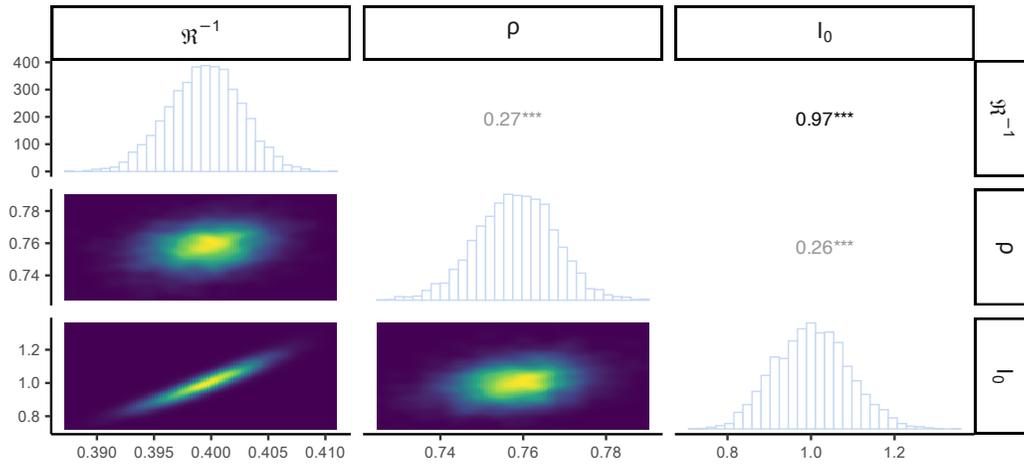


Figure B.17: Joint distribution

Table B.3: Alternative parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0
11	11	95%	100%	95%
11	12	95%	95%	45%
11	13	90%	95%	10%
11	14	85%	95%	5%
12	11	90%	100%	65%
12	12	90%	100%	95%
12	13	95%	100%	85%
12	14	95%	100%	75%
13	11	75%	100%	50%
13	12	95%	100%	95%
13	13	95%	100%	95%
13	14	100%	100%	90%
14	11	90%	100%	20%
14	12	90%	100%	60%
14	13	85%	100%	80%
14	14	85%	95%	95%

B.3 Fitting low-fidelity D^{1j}

This section describes the process of fitting various parameterisations of the SE^1I^jR model (M^{1j}) to **low-fidelity** D^{1j} incidence reports. The key difference between this section and the previous one lies in the presence of overdispersion in the data and fitting models. In particular, the latter are equipped with a Negative Binomial measurement component, which imposes the estimation of an additional parameter: (ϕ). Other than this difference, there is a large degree of redundancy between the two sections. Therefore, we assume that the reader is familiar with the content presented in the previous section.

B.3.1 Three-unknown parameterisation (traditional)

Candidates from this parameterisation have three unknowns (β , ρ , and I_0^1) in the system component and an additional one (ϕ^{-1}) in the measurement component. Fig B.18 shows the prior distribution for the latter.

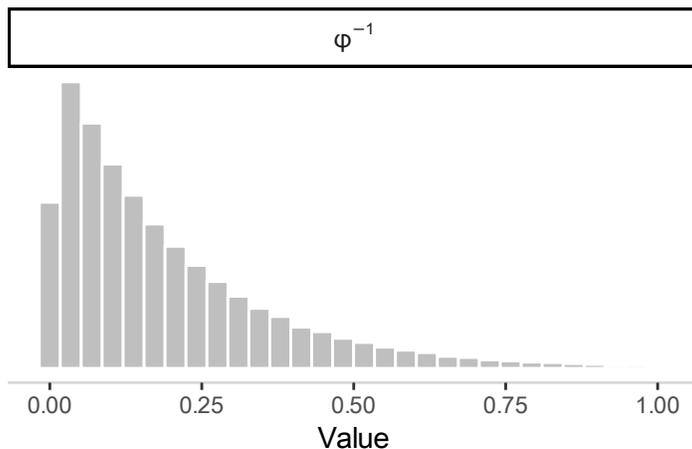


Figure B.18: Overdispersion's prior distribution

B.3.1.1 Inference

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) to each of the 80 low-fidelity D^{1j} incidence reports.

B.3.1.1.1 Incidence fit

Fig B.19 compares actual (points) and simulated (lines) latent incidence.

B.3. Fitting low-fidelity D^{1j}

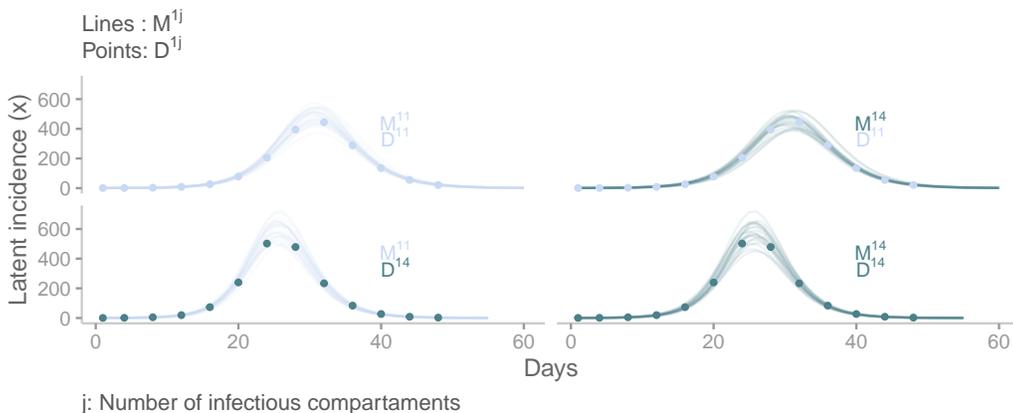


Figure B.19: Incidence fit

B.3.1.1.2 MASE

We employ the MASE to compare each of the 4000 simulated latent incidences against its true counterpart (x) and observed incidence (y). We summarise the results via histograms (Fig B.20). The left column (of panels) contains the comparison between simulated and actual latent incidences. The right column of panels displays the comparison between simulated latent incidence and the observed incidence. Overall, there is no substantial variation in the histograms as the fitting model changes (increasing j).

B.3.1.1.3 Parameter estimates

Fig B.21 shows parameter estimates (error bars) from four model candidates from this alternative parameterisation fitted to two incidence reports of different origin (D^{11} and D^{14}).

B.3.1.1.4 Joint distribution

Fig B.22 presents an example of a joint posterior distribution. Specifically, this posterior distribution was derived from fitting the M^{11} candidate to one D^{11} incidence report.

B.3.1.1.5 Coverage

Table B.4 summarises the traditional parameterisation's coverage.

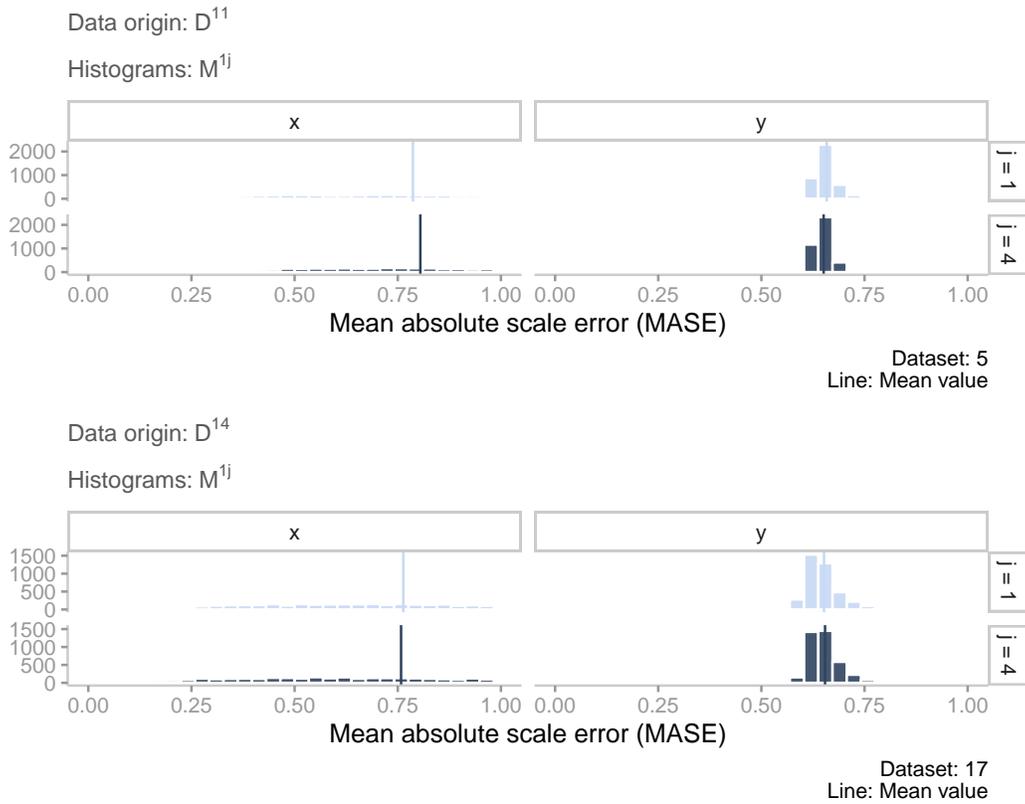


Figure B.20: Fit scores by candidate model and type of data. x denotes latent incidence, whereas y indicates observed incidence. Vertical line denotes the mean.

B.3. Fitting low-fidelity D^{1j}

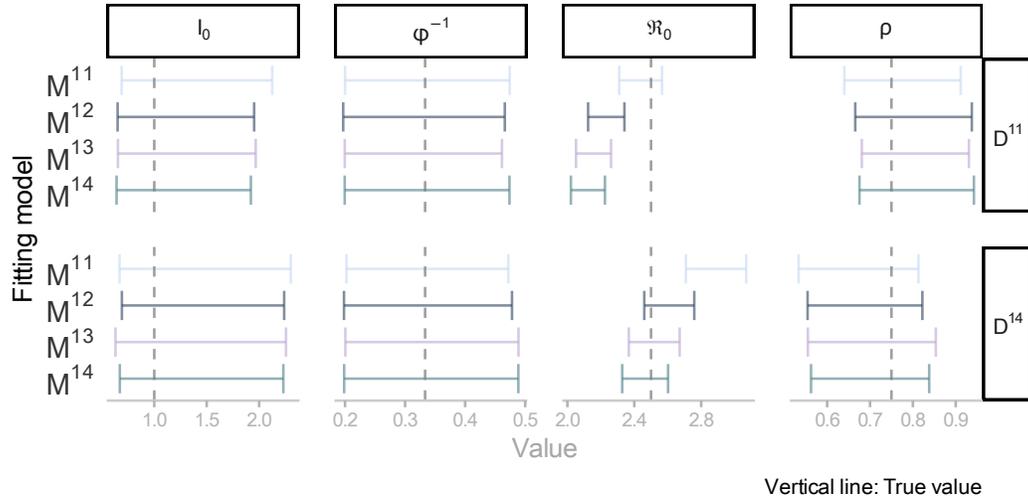


Figure B.21: Parameter estimates. Panel headings on the right-hand side indicate the data's origin.

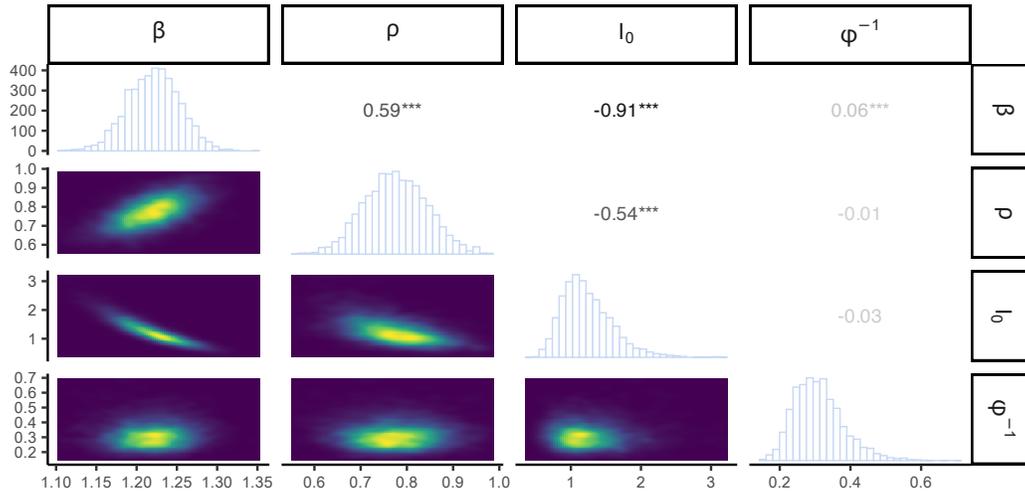


Figure B.22: Joint distribution

Table B.4: Traditional parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0	ϕ^{-1}
11	11	90%	90%	100%	100%
11	12	5%	85%	100%	100%
11	13	0%	80%	100%	100%
11	14	0%	80%	95%	100%
12	11	30%	100%	90%	100%
12	12	90%	100%	90%	100%
12	13	65%	100%	90%	100%
12	14	35%	100%	95%	100%
13	11	0%	95%	95%	95%
13	12	80%	95%	95%	95%
13	13	95%	90%	95%	95%
13	14	90%	90%	95%	95%
14	11	0%	95%	95%	95%
14	12	55%	90%	95%	95%
14	13	90%	90%	100%	95%
14	14	100%	90%	100%	95%

B.3.1.1.6 MLE criterion

Each bar (column) in in Fig B.23 represents the number of times that particular model candidate attained the largest likelihood score for a given set of incidence reports. Recall that each set of reports comprises 20 time-series. For instance, the first column (left to right) in the first panel (top-left) indicates that M^{11} outperforms its competitors 10 out of 20 times (50%) in fitting D^{11} incidence reports.

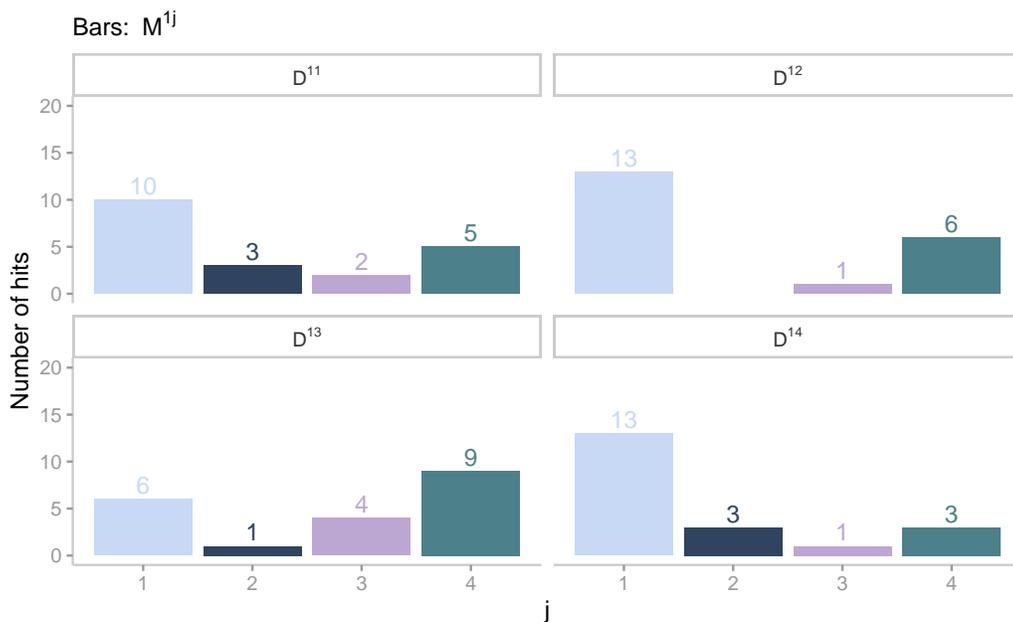


Figure B.23: Score summary

B.3.2 Four-unknown parameterisation

B.3.2.1 Inference

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from this parameterisation to each of the 80 low-fidelity D^{1j} incidence reports.

B.3.2.1.1 Incidence fit

Fig B.24 compares actual (points) and simulated (lines) latent incidence.

B.3.2.1.2 Parameter estimates

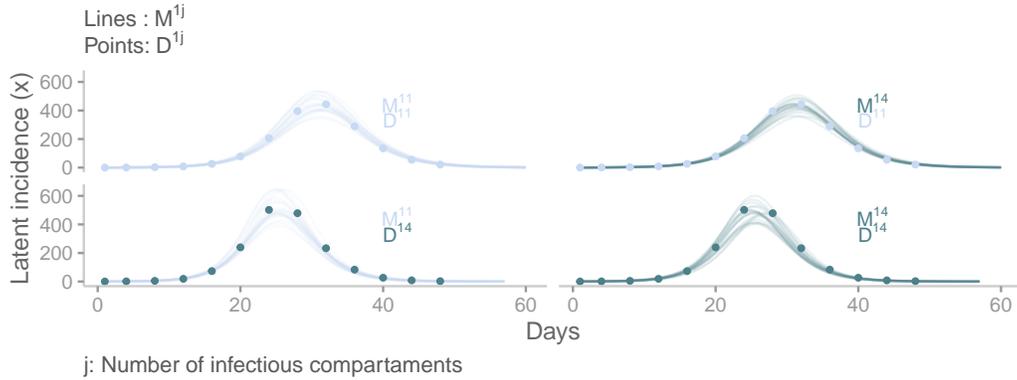


Figure B.24: Incidence fit

Fig B.25 shows parameter estimates (error bars) from four model candidates from this alternative parameterisation fitted to two incidence reports of different origin (D^{11} and D^{14}).

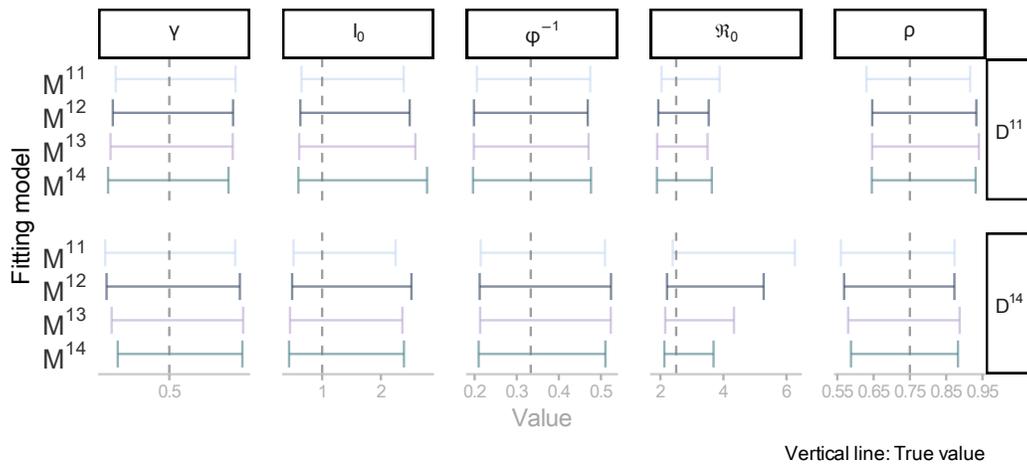


Figure B.25: Parameter estimates. Panel headings on the right-hand side indicate the data's origin.

B.3.2.1.3 Joint distribution

Fig B.26 presents an example of a joint posterior distribution. Specifically, this posterior distribution was derived from fitting the M^{12} candidate to one D^{12} incidence report.

B.3.2.1.4 Coverage

B.3. Fitting low-fidelity D^{1j}

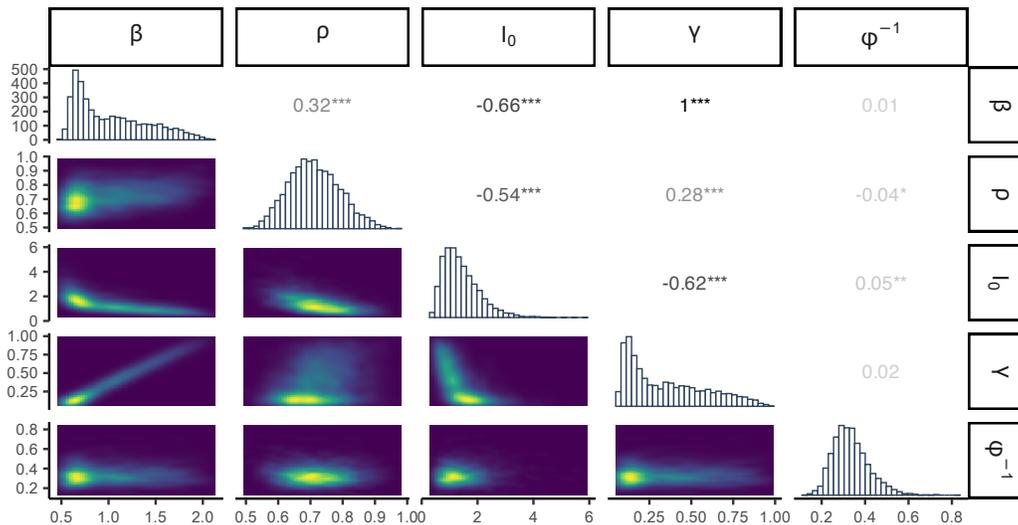


Figure B.26: Joint distribution

Table B.5 summarises the four-unknown parameterisation's coverage.

B.3.2.1.5 \mathfrak{R}_0 vs τ

Fig B.27 shows the linear relationship between the basic reproduction number and the mean generation time.

B.3.3 Three-unknown parameterisation (alternative)

B.3.3.1 Inference

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) to each of the 80 low-fidelity D^{1j} incidence reports.

B.3.3.1.1 Incidence fit

Fig B.28 compares actual (points) and simulated (lines) latent incidence.

B.3.3.1.2 Parameter estimates

Fig B.29 shows parameter estimates (error bars) from four model candidates from this alternative parameterisation fitted to two incidence reports of different origin (D^{11} and D^{14}).

Table B.5: Four-unknown parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	γ^{-1}	I_0	ϕ^{-1}
11	11	100%	100%	95%	100%	100%
11	12	100%	100%	90%	100%	100%
11	13	100%	100%	90%	100%	100%
11	14	95%	100%	90%	100%	100%
12	11	100%	100%	100%	95%	100%
12	12	100%	100%	100%	100%	100%
12	13	100%	100%	100%	100%	100%
12	14	100%	100%	100%	100%	100%
13	11	95%	100%	90%	95%	95%
13	12	100%	100%	95%	95%	95%
13	13	100%	100%	95%	100%	95%
13	14	100%	100%	95%	100%	95%
14	11	95%	100%	95%	100%	95%
14	12	100%	100%	95%	100%	90%
14	13	100%	100%	90%	100%	95%
14	14	100%	100%	90%	100%	95%

B.3. Fitting low-fidelity D^{1j}

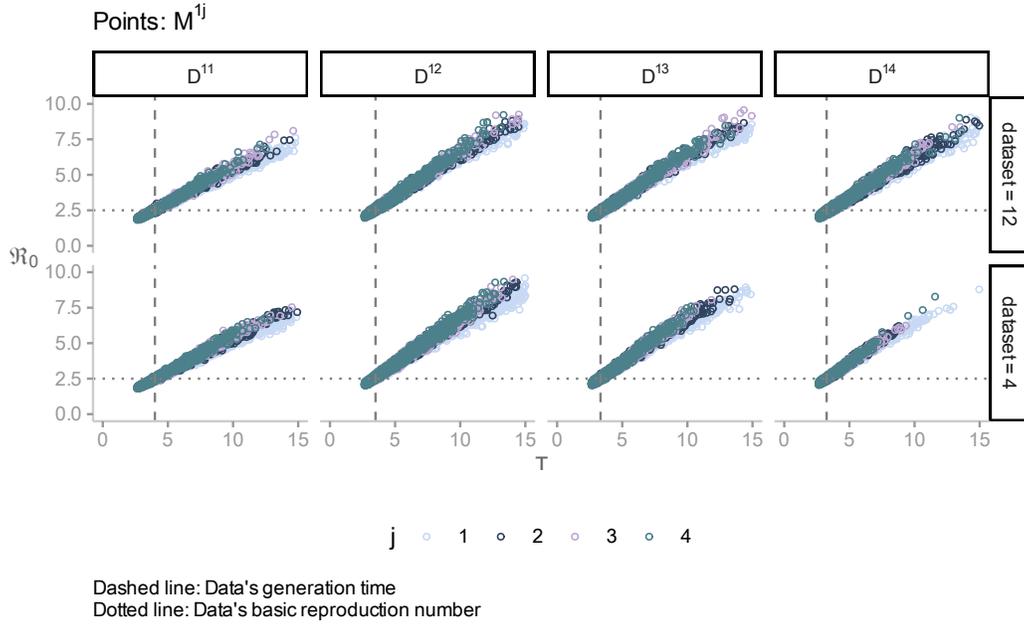


Figure B.27: Scatter plot

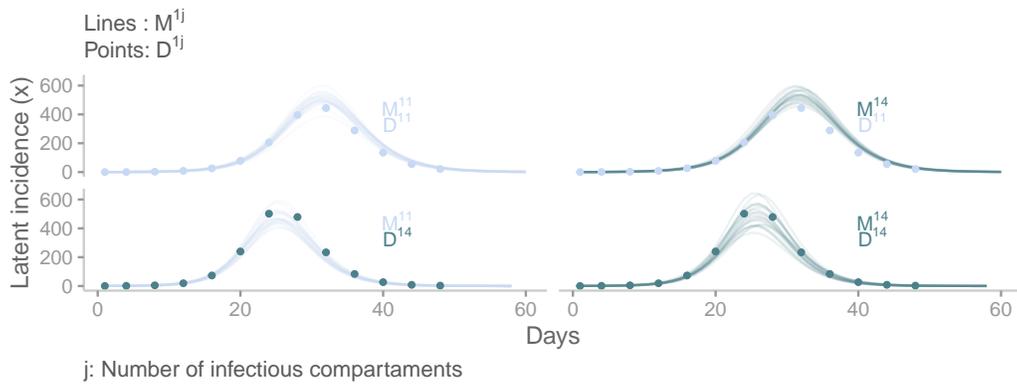


Figure B.28: Incidence fit

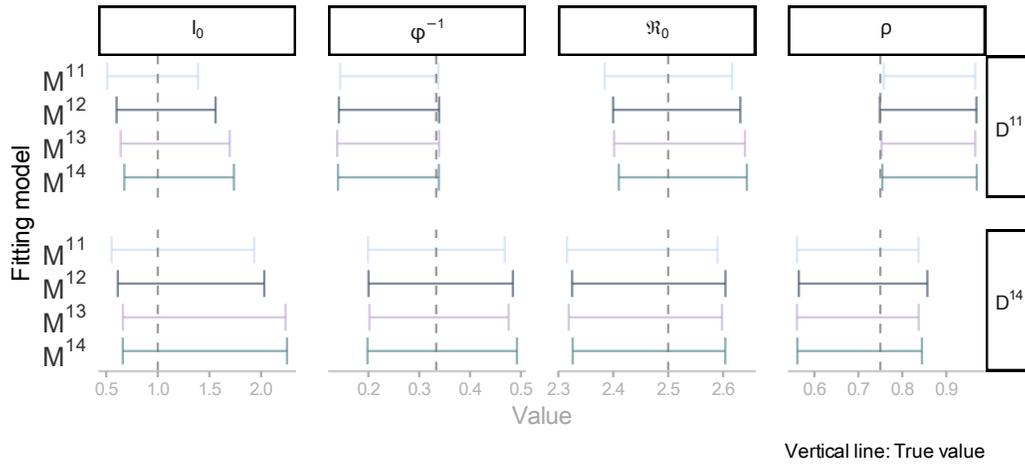


Figure B.29: Parameter estimates. Panel headings on the right-hand side indicate the data's origin.

B.3.3.1.3 Joint posterior distribution

Fig B.30 presents an example of a joint posterior distribution. Specifically, this posterior distribution was derived from fitting M^{14} to one D^{14} incidence report.

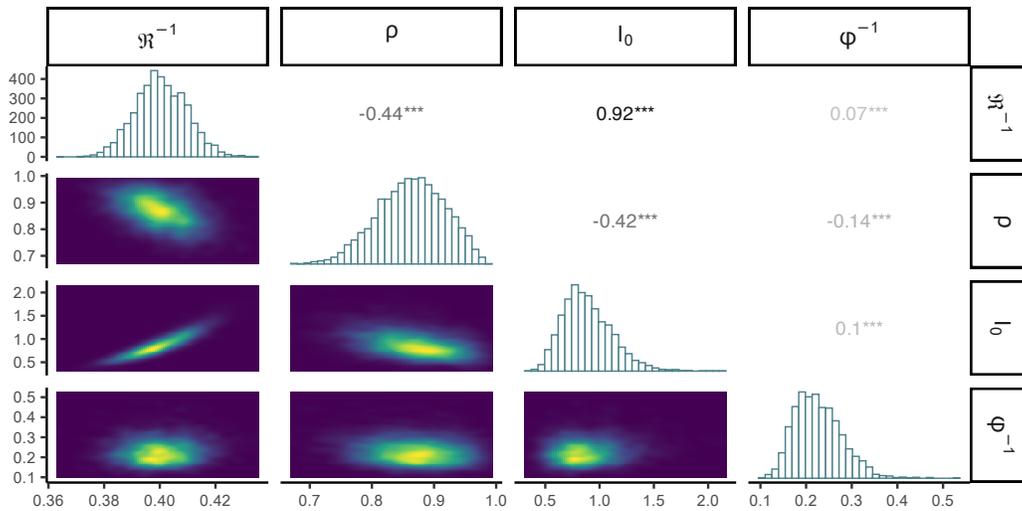


Figure B.30: Joint distribution

B.3.3.1.4 Coverage

Table B.6 summarises the alternative parameterisation's coverage.

Table B.6: Alternative parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0	ϕ^{-1}
11	11	95%	80%	100%	100%
11	12	95%	85%	85%	100%
11	13	95%	85%	80%	100%
11	14	95%	80%	75%	100%
12	11	90%	100%	100%	100%
12	12	90%	100%	90%	100%
12	13	90%	100%	90%	100%
12	14	90%	100%	85%	100%
13	11	90%	95%	95%	95%
13	12	95%	95%	95%	95%
13	13	95%	95%	95%	95%
13	14	95%	95%	95%	90%
14	11	90%	90%	90%	90%
14	12	95%	90%	100%	95%
14	13	95%	90%	100%	90%
14	14	100%	90%	100%	95%

B.4 Fitting low-fidelity D^{3j}

This section illustrates the process of fitting various parameterisations of the $SE^i I^j R$ model (M^{1j} & M^{3j}) to **low-fidelity** D^{3j} incidence reports. In particular, this section expands on the work presented in the preceding sections by exploring the consequences of misspecifying the latent period distribution. Thus, we assume the reader is familiar with the previous sections so as to avoid redundancy in the documentation.

B.4.1 Three-unknown parameterisation (traditional)

Candidates from this parameterisation have three unknowns (β , ρ , and I_0^1) in the system component and an additional one (ϕ^{-1}) in the measurement component.

B.4.1.1 Inference

We fit eight candidate models (M^{11} , M^{12} , M^{13} , M^{14} , M^{31} , M^{32} , M^{33} , and M^{34}) to each of the 80 low-fidelity D^{3j} incidence reports.

B.4.1.1.1 Incidence fit

Fig B.31 compares actual (points) and simulated (lines) latent incidence. Specifically, this comparison shows four candidate models from this parameterisation fitting one D^{33} report. Notice that misspecifying the latent period distribution leads to a less accurate fit.

B.4.1.1.2 MASE

Fig B.32 corroborates the previous observation. Notice that candidates with a gamma-distributed latent period (M^{31} and M^{33}) attain lower scores than their counterparts with an exponentially-distributed latent period. On the other hand, MASE scores are far less sensitive to changes in the infectious period distribution.

B.4.1.1.3 Parameter estimates

Fig B.33 shows parameter estimates (error bars) from eight model candidates from this traditional parameterisation fitted to one incidence report (D^{33}).

B.4. Fitting low-fidelity D^{3j}

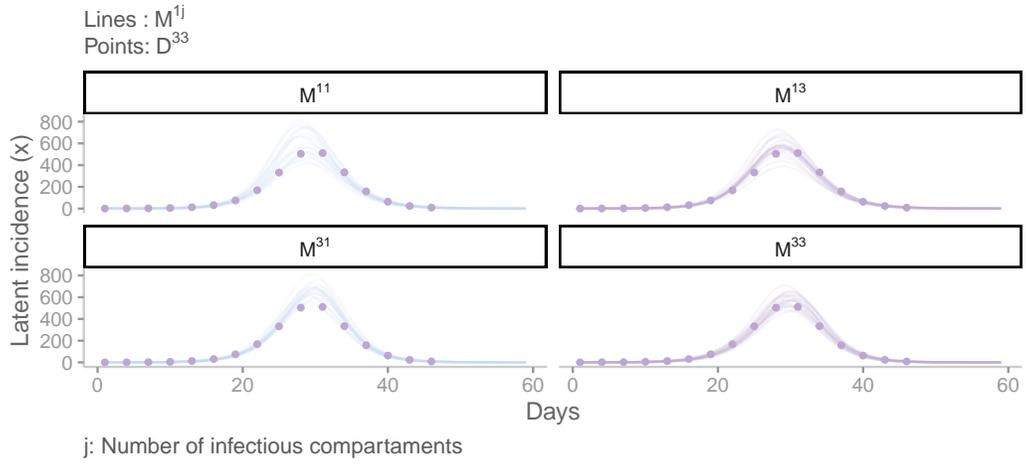


Figure B.31: Incidence fit

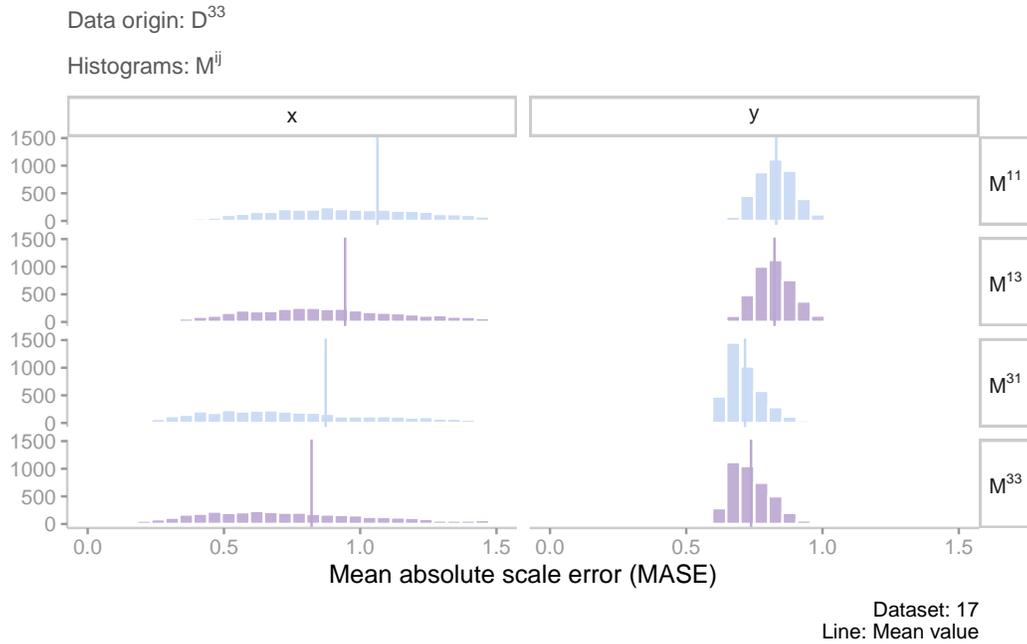


Figure B.32: MASE

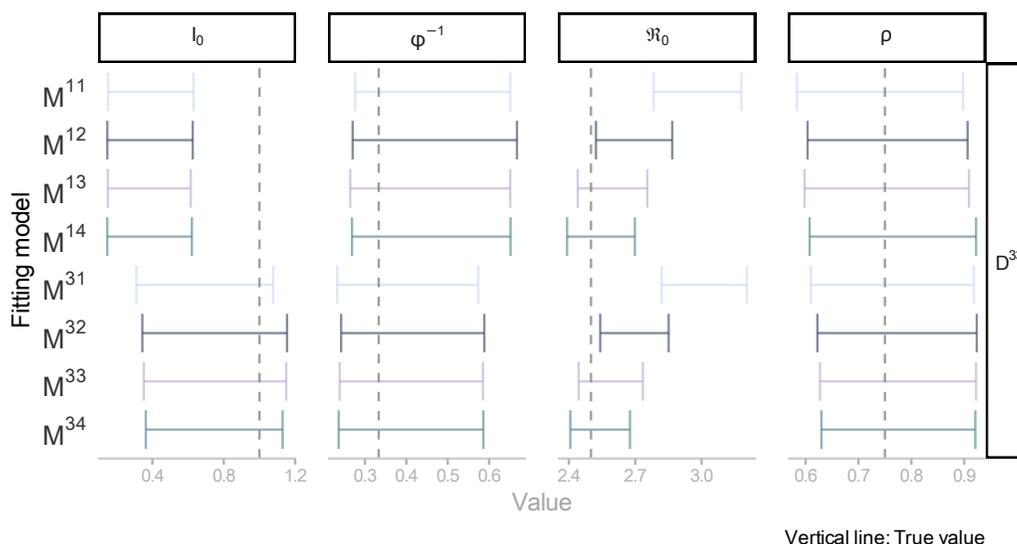


Figure B.33: Parameter estimates.

B.4.1.1.4 Joint posterior distribution

Fig B.34 presents an example of a joint posterior distribution. Specifically, this posterior distribution was derived from fitting M^{13} to one D^{13} incidence report.

B.4.1.1.5 Coverage

Table B.7 summarises this parameterisation's coverage.

B.4.2 Four-unknown parameterisation

B.4.2.1 Inference

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from this parameterisation to each of the 80 low-fidelity D^{3j} incidence reports. As mentioned in the main text, misspecifying the latent period distribution merely produces slight biases in the estimates of the initial number of infectious individuals (I_0) and incidence over time. Therefore, we restrict ourselves to reporting relationship between the basic reproduction number and the mean generation time and estimated coverage, given that the results are equivalent to those presented in the previous sections.

B.4. Fitting low-fidelity D^{3j}

Table B.7: Traditional parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0	ϕ^{-1}
31	31	95%	95%	95%	95%
31	32	0%	95%	95%	95%
31	33	0%	90%	95%	95%
31	34	0%	90%	95%	95%
31	11	90%	95%	65%	95%
31	12	0%	95%	60%	95%
31	13	0%	95%	55%	95%
31	14	0%	95%	55%	95%
32	31	5%	100%	95%	100%
32	32	100%	100%	100%	100%
32	33	80%	100%	100%	100%
32	34	45%	100%	100%	100%
32	11	10%	95%	45%	95%
32	12	100%	100%	40%	95%
32	13	85%	100%	30%	95%
32	14	60%	100%	30%	95%
33	31	0%	95%	80%	85%
33	32	80%	95%	90%	85%
33	33	85%	95%	90%	85%
33	34	75%	95%	90%	85%
33	11	0%	85%	70%	90%
33	12	80%	90%	60%	90%
33	13	90%	95%	60%	90%
33	14	85%	95%	60%	85%
34	31	0%	95%	90%	95%
34	32	55%	95%	90%	95%
34	33	90%	100%	85%	90%
34	34	90%	100%	85%	90%
34	11	0%	90%	60%	95%
34	12	55%	95%	45%	100%
34	13	90%	95%	45%	100%
34	14	95%	100%	45%	95%

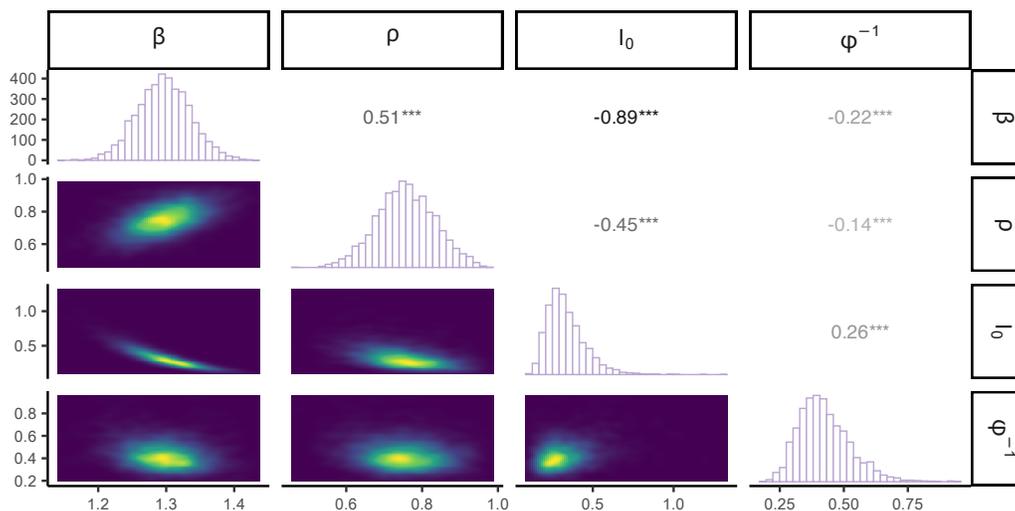


Figure B.34: Joint distribution

B.4.2.1.1 Coverage

Table B.8 summarises this parameterisation’s coverage.

B.4.2.1.2 \mathcal{R}_0 vs τ

Fig B.35 shows the linear relationship between the basic reproduction number and the mean generation time.

B.4.3 Three-unknown parameterisation (alternative)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each of the 80 low-fidelity D^{3j} incidence reports.

B.4.3.1 Coverage

We summarise the results in Table B.9.

B.4.4 Three-unknown (Alternative) - Poisson

In this section, we test the implications of amalgamating the alternative parameterisation with a Poisson measurement model (likelihood function) to fit overdispersed incidence data. We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) to each of the 80 low-fidelity D^{3j} incidence reports.

Table B.8: Four-unknown parameterisation's coverage table

D^{ij}	M^{ij}	\Re_0	ρ	γ^{-1}	I_0	ϕ^{-1}
31	11	95%	95%	95%	95%	95%
31	12	95%	100%	95%	100%	95%
31	13	100%	100%	90%	100%	95%
31	14	100%	100%	90%	100%	95%
32	11	100%	95%	100%	100%	100%
32	12	100%	100%	100%	100%	100%
32	13	100%	95%	100%	100%	100%
32	14	100%	100%	100%	100%	100%
33	11	65%	90%	90%	95%	80%
33	12	85%	90%	90%	95%	80%
33	13	85%	90%	90%	95%	80%
33	14	90%	95%	80%	95%	85%
34	11	65%	90%	95%	90%	95%
34	12	80%	90%	95%	90%	100%
34	13	80%	90%	80%	90%	95%
34	14	80%	90%	80%	90%	95%

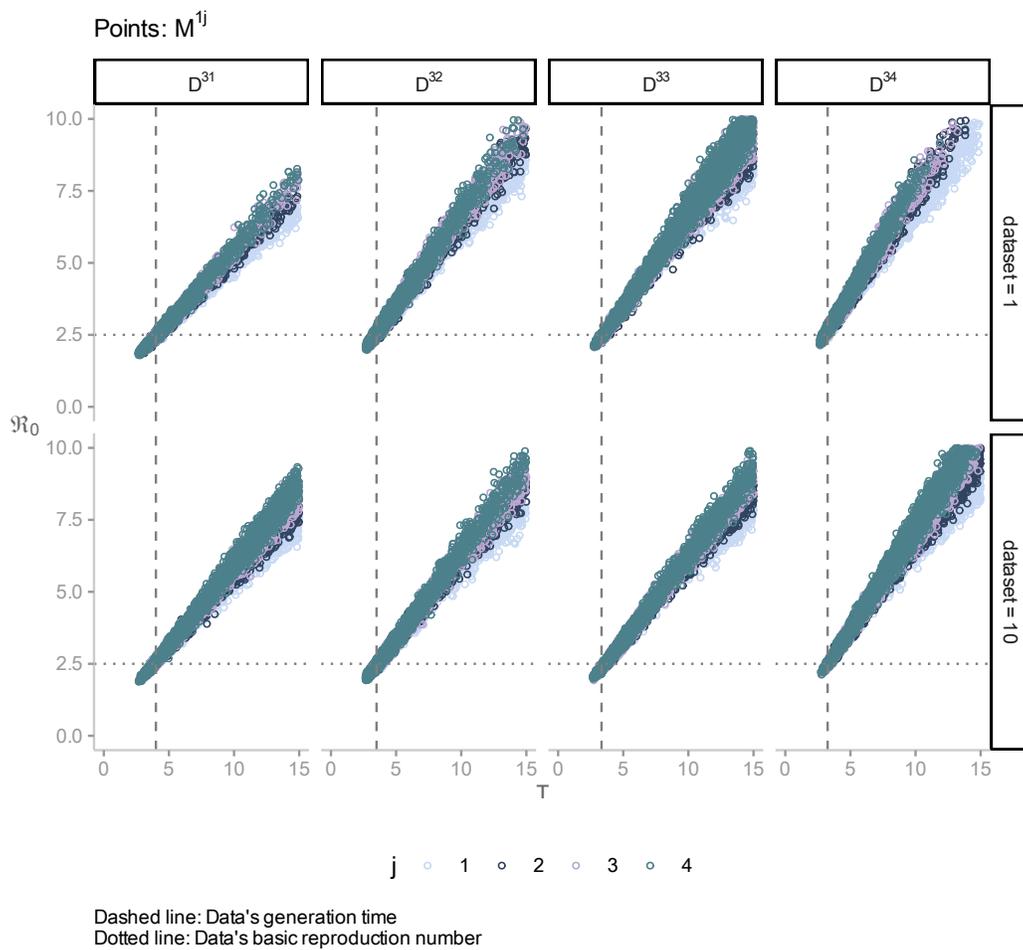


Figure B.35: Scatter plot

Table B.9: Alternative parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0	ϕ^{-1}
31	11	95%	95%	70%	95%
31	12	95%	95%	80%	95%
31	13	95%	95%	85%	95%
31	14	95%	95%	90%	95%
32	11	100%	100%	25%	95%
32	12	100%	95%	35%	95%
32	13	100%	95%	45%	95%
32	14	100%	100%	45%	95%
33	11	85%	95%	60%	90%
33	12	90%	95%	60%	90%
33	13	85%	95%	60%	85%
33	14	90%	95%	70%	90%
34	11	85%	95%	40%	85%
34	12	95%	95%	45%	95%
34	13	100%	95%	50%	100%
34	14	95%	95%	45%	95%

Table B.10: Alternative parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0
31	11	15%	30%	20%
31	12	15%	30%	15%
31	13	15%	30%	15%
31	14	15%	25%	10%
32	11	25%	20%	5%
32	12	30%	20%	0%
32	13	30%	20%	0%
32	14	30%	20%	0%
33	11	5%	10%	15%
33	12	5%	10%	20%
33	13	10%	10%	25%
33	14	10%	10%	25%
34	11	5%	0%	5%
34	12	10%	0%	5%
34	13	10%	0%	5%
34	14	10%	5%	5%

B.4.4.1 Coverage

We summarise the results in Table B.10.

B.5 Fitting high-fidelity D^{3j}

This section illustrates the process of fitting various parameterisations of the SE^iI^jR model (M^{1j}) to **high-fidelity** D^{3j} incidence reports. These results verify the insights obtained in the previous section. That is, misspecifying the latent period distribution does not have a detrimental effect on the estimation of \mathfrak{R}_0 . Consequently, we restrict the presentation of results to coverage tables and the linear relationship between the basic reproduction number and the mean generation time obtained from the four-unknown parameterisation.

Table B.11: Traditional parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0
31	11	95%	90%	0%
31	12	0%	15%	0%
31	13	0%	5%	0%
31	14	0%	5%	0%
32	11	0%	35%	0%
32	12	100%	95%	0%
32	13	0%	90%	0%
32	14	0%	65%	0%
33	11	0%	10%	0%
33	12	0%	85%	0%
33	13	90%	95%	0%
33	14	55%	85%	0%
34	11	0%	5%	0%
34	12	0%	75%	0%
34	13	30%	90%	0%
34	14	85%	95%	0%

B.5.1 Three-unknown parameterisation (traditional)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the traditional parameterisation to each of the 80 high-fidelity D^{3j} incidence reports.

B.5.1.1 Coverage

We summarise the results in Table B.11.

B.5.2 Four-unknown parameterisation

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the four-unknown parameterisation to each of the 80 high-fidelity D^{3j} incidence reports.

B.5.2.1 Coverage

We summarise the results in Table B.12.

Table B.12: Four-unknown parameterisation's coverage table

D^{ij}	M^{ij}	\Re_0	ρ	γ^{-1}	I_0
31	11	15%	10%	15%	40%
31	12	25%	25%	5%	80%
31	13	30%	25%	0%	95%
31	14	40%	25%	0%	95%
32	11	0%	0%	0%	30%
32	12	0%	0%	0%	30%
32	13	0%	0%	0%	35%
32	14	0%	0%	0%	40%
33	11	0%	0%	0%	15%
33	12	0%	0%	0%	15%
33	13	0%	0%	0%	40%
33	14	0%	0%	0%	40%
34	11	0%	0%	0%	15%
34	12	0%	0%	0%	10%
34	13	0%	0%	0%	15%
34	14	0%	0%	0%	25%

B.5.2.2 \mathcal{R}_0 vs τ

Fig B.36 shows the linear relationship between the basic reproduction number and the mean generation time.

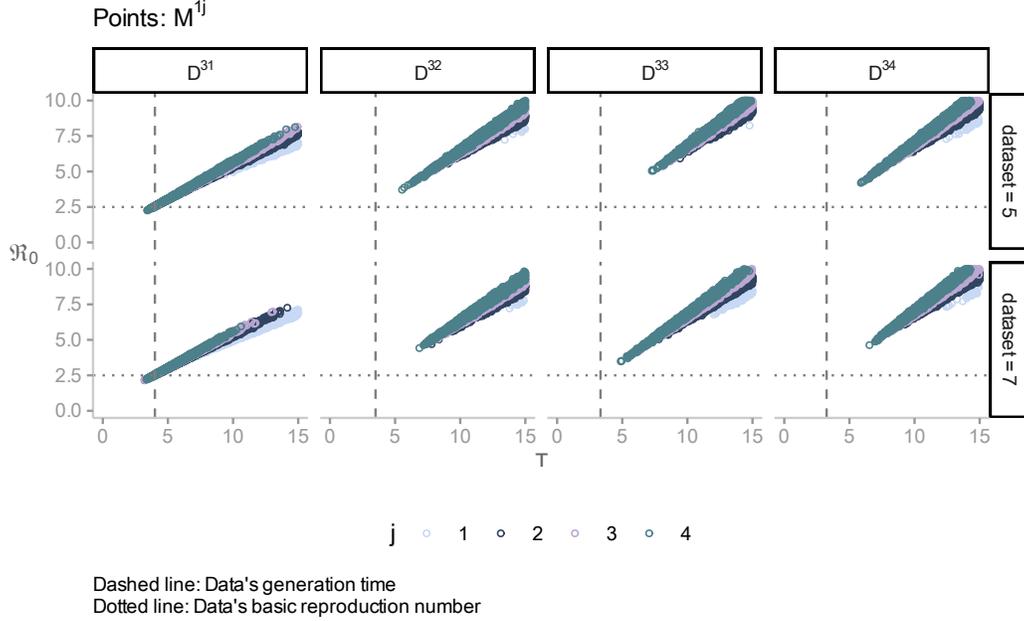


Figure B.36: Scatter plot

B.5.3 Three-unknown parameterisation (alternative)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each of the 80 high-fidelity D^{3j} incidence reports.

B.5.3.1 Coverage

We summarise the results in Table B.13.

B.6 Sensitivity analysis

This section aims to illustrate the performance of the alternative parameterisation under various transmissibility levels (\mathcal{R}_0) and mean generation times. For simplicity, we restrict the $SE^i I^j R$ to four instances ($i = 1$, and $j = \{1, 2, 3, 4\}$) for data generation and fitting. In particular, we test the alternative parameterisation under four additional scenarios (Table B.14). The

Table B.13: Alternative parameterisation's coverage table

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0
31	11	95%	90%	0%
31	12	90%	90%	0%
31	13	70%	90%	0%
31	14	55%	95%	0%
32	11	85%	90%	0%
32	12	100%	90%	0%
32	13	85%	95%	0%
32	14	85%	95%	0%
33	11	95%	90%	0%
33	12	90%	95%	0%
33	13	90%	95%	0%
33	14	85%	95%	0%
34	11	90%	90%	0%
34	12	95%	95%	0%
34	13	90%	95%	0%
34	14	85%	95%	0%

Table B.14: Sensitivity analysis scenarios

Scenario	\mathfrak{R}_0	τ_e
1	2.5	4
2	2.5	8
3	2.5	13
4	9.0	4
5	17.0	4

first scenario corresponds to Appendix B.3. The term *scenario* refers to a particular configuration of \mathfrak{R}_0 and the mean generation time (τ) in the SE^iI^jR framework to produce the synthetic data. Since τ varies with the infectious period distribution (j), we identify each scenario by the mean generation time obtained from an exponentially-distributed infectious period (denoted by τ_e). For each instance and data-fidelity level, we generate 20 synthetic incidence reports. Recall that fidelity levels correspond to two configurations of the Negative Binomial measurement component: high ($\phi^{-1} = 0$) and low fidelity ($\phi^{-1} = \frac{1}{3}$). As a result, we produce 160 synthetic incidence reports per scenario. Then, we fit four candidates per incidence report, assuming appropriate measurement components.

B.6.1 Scenario 2

In comparison to Scenario 1, we increase the reference mean generation time (τ_e) from 4 to 8 to produce the synthetic data (Fig B.37).

B.6.1.1 Inference (high-fidelity data)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each incidence report.

B.6.1.1.1 Coverage

We summarise the results in Table B.15.

B.6.1.1.2 \mathfrak{R}_0 estimates

Fig B.38 displays parameter estimates from fitting four candidates (two from the traditional and two from the alternative parameterisation) to one

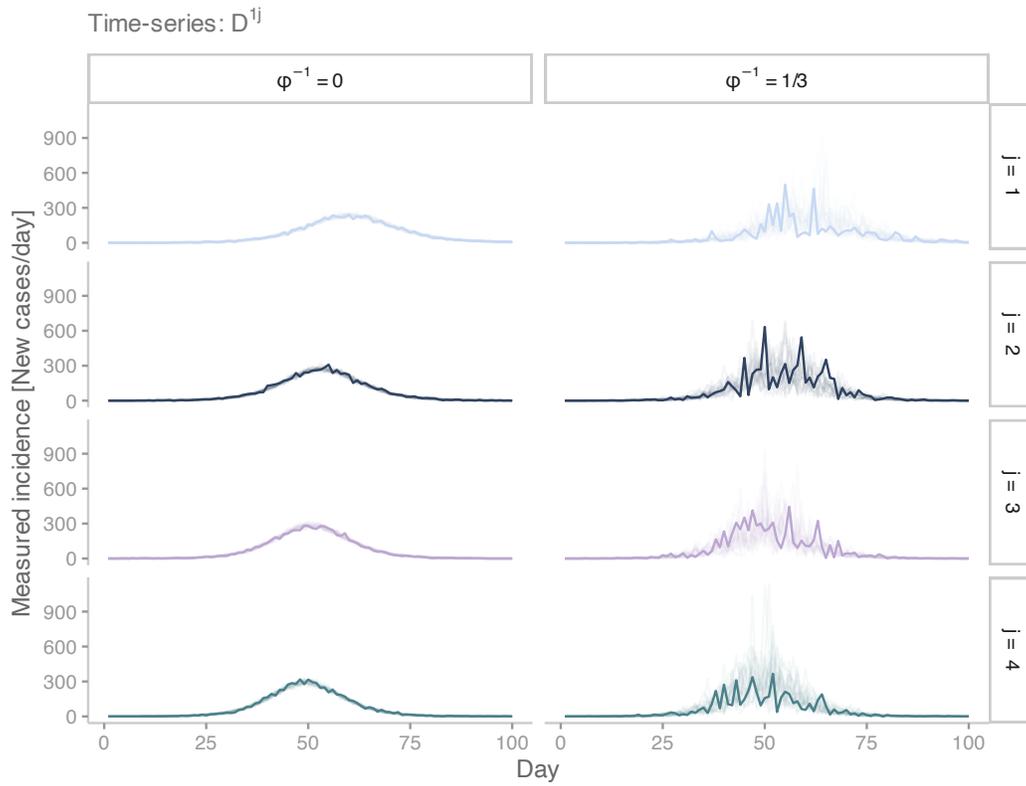


Figure B.37: Scenario 2's simulated incidence reports. Measurement noise from the Poisson (no overdispersion) and Negative Binomial distributions was added to the smooth trajectories obtained from SEIR instances with an exponential-distributed latent period. There are 20 incidence reports overlapped in each panel.

Table B.15: Scenario 2's coverage table. Poisson noise

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0
11	11	95%	90%	95%
11	12	80%	95%	35%
11	13	35%	95%	10%
11	14	30%	90%	0%
12	11	80%	80%	45%
12	12	95%	80%	95%
12	13	90%	80%	80%
12	14	75%	80%	65%
13	11	60%	85%	25%
13	12	95%	85%	85%
13	13	95%	90%	95%
13	14	100%	95%	85%
14	11	60%	100%	25%
14	12	75%	100%	70%
14	13	75%	95%	85%
14	14	90%	95%	85%

D^{11} incidence report. Although both M^{14} candidates miss the actual value, the estimate from the alternative parameterisation is much closer than that one of the traditional parameterisation. The value in the middle of error indicate the percentage difference between the actual value and posterior distribution's mean.

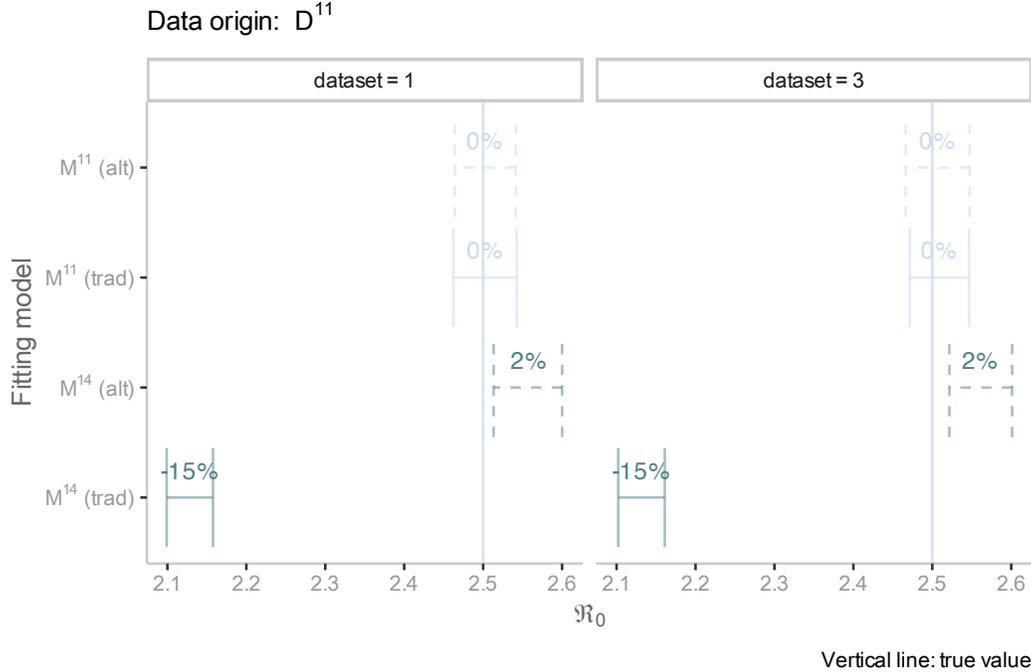


Figure B.38: Estimates of the basic reproduction number by model candidate.

B.6.1.2 Inference (low-fidelity data)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each incidence report.

B.6.1.2.1 Coverage

We summarise the results in Table B.16.

B.6.2 Scenario 3

In comparison to Scenario 2, we increase the reference mean generation time (τ_e) from 8 to 13 to produce the synthetic data (Fig B.39).

Table B.16: Scenario 2's coverage table. Overdispersion

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0	ϕ^{-1}
11	11	95%	95%	95%	95%
11	12	90%	90%	80%	95%
11	13	85%	90%	75%	95%
11	14	85%	95%	75%	95%
12	11	100%	90%	95%	85%
12	12	100%	90%	95%	90%
12	13	100%	90%	100%	90%
12	14	100%	90%	100%	90%
13	11	85%	95%	95%	95%
13	12	90%	100%	95%	95%
13	13	90%	95%	90%	95%
13	14	90%	100%	85%	95%
14	11	90%	95%	80%	95%
14	12	90%	95%	90%	95%
14	13	95%	95%	95%	95%
14	14	95%	95%	95%	95%

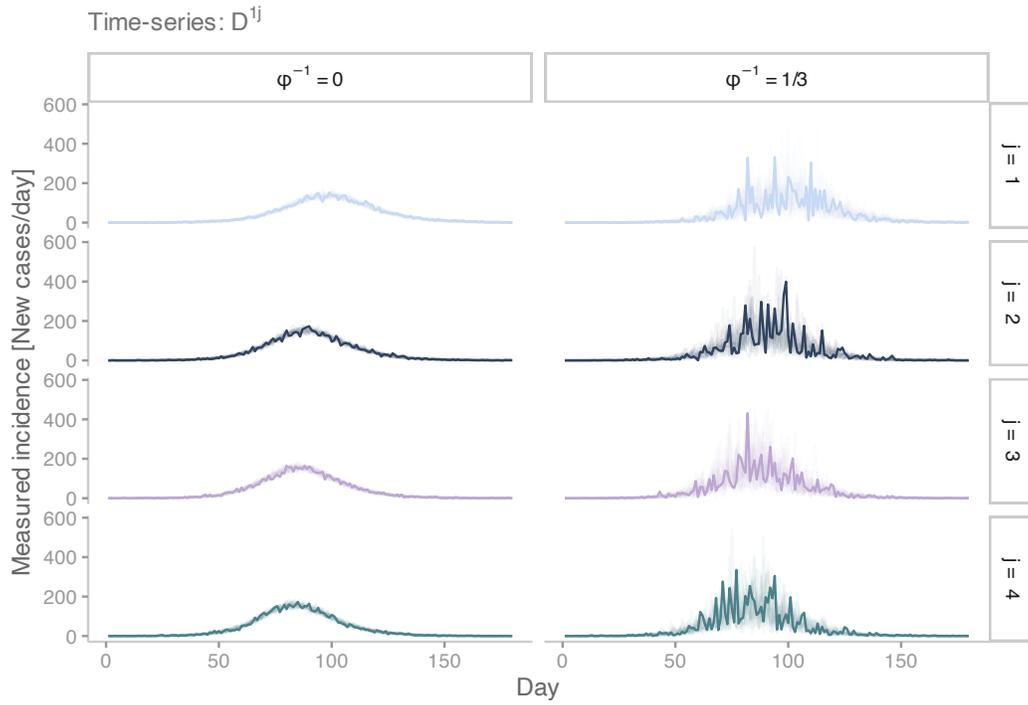


Figure B.39: Scenario 3's simulated incidence reports. Measurement noise from the Poisson (no overdispersion) and Negative Binomial distributions was added to the smooth trajectories obtained from SEIR instances with an exponential-distributed latent period.

Table B.17: Scenario 3's coverage table. Poisson noise

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0
11	11	100%	100%	100%
11	12	95%	100%	75%
11	13	90%	100%	55%
11	14	90%	100%	45%
12	11	95%	90%	85%
12	12	90%	90%	100%
12	13	90%	95%	90%
12	14	90%	95%	85%
13	11	95%	100%	65%
13	12	100%	95%	95%
13	13	100%	90%	95%
13	14	95%	90%	100%
14	11	90%	100%	75%
14	12	95%	100%	90%
14	13	90%	100%	90%
14	14	90%	100%	90%

B.6.2.1 Inference (high-fidelity data)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each incidence report.

B.6.2.1.1 Coverage

We summarise the results in Table [B.17](#).

B.6.2.2 Inference (low-fidelity data)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each incidence report.

B.6.2.2.1 Coverage

We summarise the results in Table [B.18](#).

Table B.18: Scenario 3's coverage table. Overdispersion

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0	ϕ^{-1}
11	11	95%	100%	95%	90%
11	12	100%	95%	95%	95%
11	13	90%	100%	95%	95%
11	14	90%	100%	90%	95%
12	11	90%	100%	100%	85%
12	12	95%	100%	100%	90%
12	13	95%	100%	90%	80%
12	14	95%	100%	90%	80%
13	11	100%	95%	95%	90%
13	12	100%	95%	100%	90%
13	13	100%	95%	100%	90%
13	14	100%	95%	100%	85%
14	11	85%	95%	90%	85%
14	12	90%	95%	90%	85%
14	13	90%	90%	90%	85%
14	14	90%	95%	90%	85%

B.6.3 Scenario 4

In comparison to Scenario 1, we increase \mathfrak{R}_0 from **2.5** to **9** to produce the synthetic data (Fig B.40).

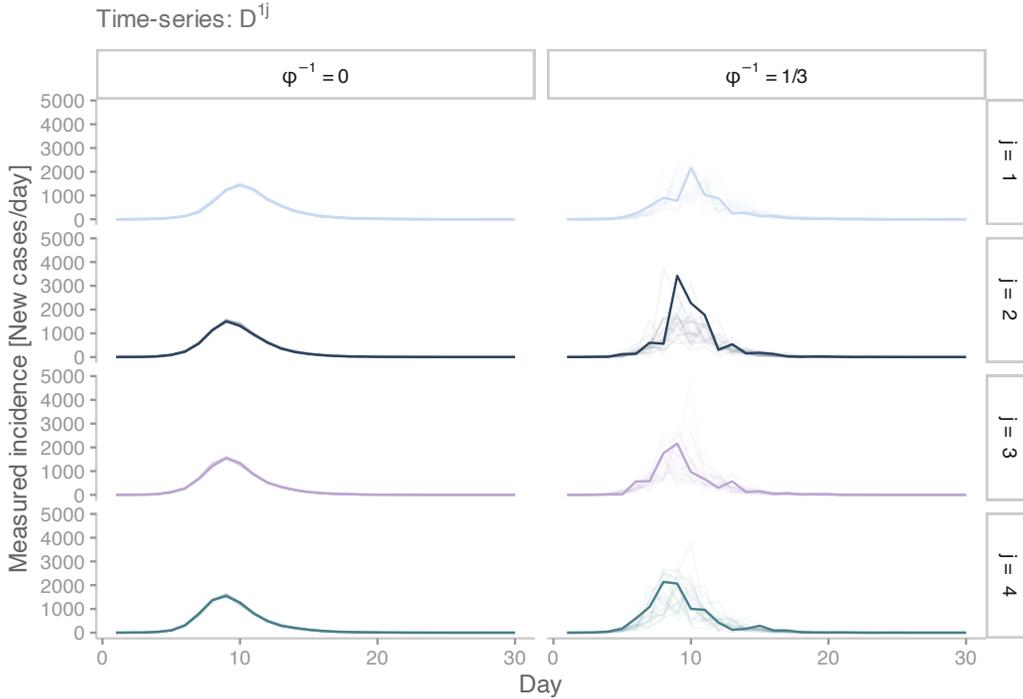


Figure B.40: Scenario 4’s simulated incidence reports. Measurement noise from the Poisson (no overdispersion) and Negative Binomial distributions was added to the smooth trajectories obtained from SEIR instances with an exponential-distributed latent period.

B.6.3.1 Inference (high-fidelity data)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each incidence report.

B.6.3.1.1 Coverage

We summarise the results in Table B.19.

B.6.3.2 Inference (low-fidelity data)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each incidence report.

Table B.19: Scenario 4's coverage table. Poisson noise

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0
11	11	95%	80%	95%
11	12	10%	80%	90%
11	13	0%	80%	80%
11	14	0%	80%	65%
12	11	45%	95%	70%
12	12	90%	95%	85%
12	13	55%	95%	90%
12	14	35%	95%	85%
13	11	10%	100%	75%
13	12	80%	100%	90%
13	13	95%	100%	90%
13	14	90%	100%	95%
14	11	0%	85%	50%
14	12	60%	90%	85%
14	13	100%	90%	95%
14	14	100%	80%	100%

Table B.20: Scenario 4's coverage table. Overdispersion

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0	ϕ^{-1}
11	11	95%	100%	95%	80%
11	12	95%	100%	95%	90%
11	13	90%	100%	95%	80%
11	14	80%	100%	95%	90%
12	11	85%	85%	100%	90%
12	12	90%	95%	95%	90%
12	13	95%	90%	95%	90%
12	14	95%	90%	95%	90%
13	11	75%	100%	100%	95%
13	12	90%	100%	100%	95%
13	13	95%	100%	100%	95%
13	14	100%	100%	100%	95%
14	11	85%	100%	100%	100%
14	12	100%	100%	100%	100%
14	13	100%	100%	100%	100%
14	14	100%	100%	100%	95%

B.6.3.2.1 Coverage

We summarise the results in Table B.20.

B.6.4 Scenario 5

In comparison to Scenario 4, we increase \mathfrak{R}_0 from **9** to **17** to produce the synthetic data (Fig B.41).

B.6.4.1 Inference (high-fidelity data)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each incidence report.

B.6.4.1.1 Coverage

We summarise the results in Table B.21.

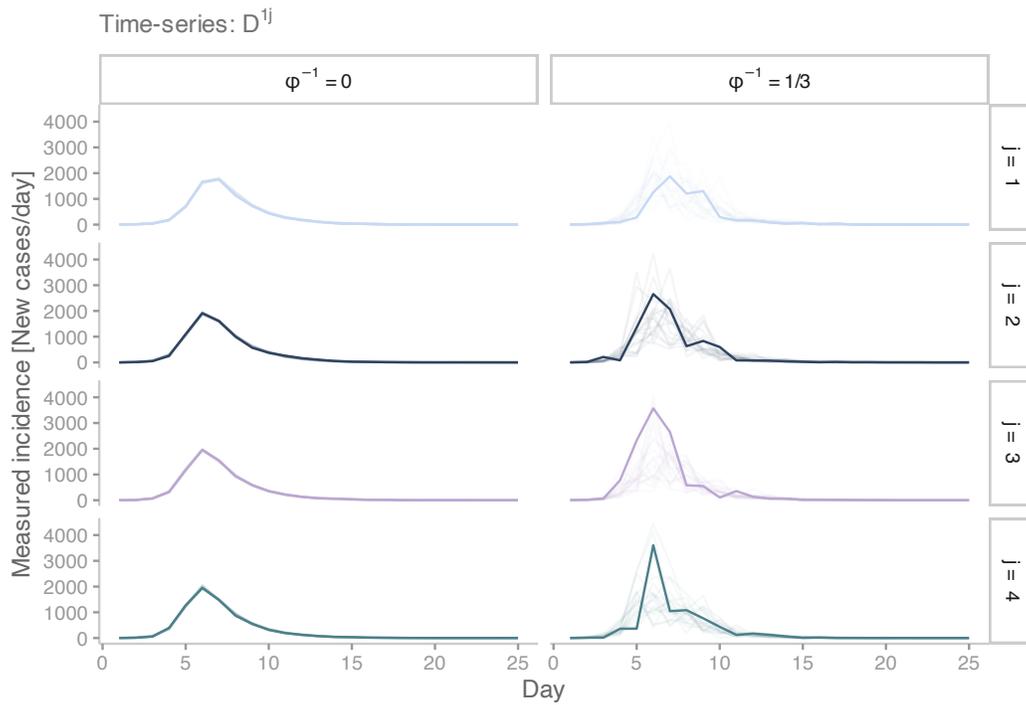


Figure B.41: Scenario 5's simulated incidence reports. Measurement noise from the Poisson (no overdispersion) and Negative Binomial distributions was added to the smooth trajectories obtained from SEIR instances with an exponential-distributed latent period.

Table B.21: Scenario 5's coverage table. Poisson noise

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0
11	11	100%	95%	100%
11	12	0%	95%	100%
11	13	0%	95%	100%
11	14	0%	95%	100%
12	11	25%	95%	90%
12	12	95%	95%	95%
12	13	60%	95%	100%
12	14	30%	95%	95%
13	11	0%	100%	95%
13	12	60%	100%	100%
13	13	95%	100%	100%
13	14	95%	100%	90%
14	11	0%	100%	100%
14	12	25%	100%	100%
14	13	85%	100%	95%
14	14	95%	100%	95%

Table B.22: Scenario 5’s coverage table. Overdispersion

D^{ij}	M^{ij}	\mathfrak{R}_0	ρ	I_0	ϕ^{-1}
11	11	100%	100%	100%	100%
11	12	100%	100%	100%	100%
11	13	95%	100%	100%	100%
11	14	80%	100%	100%	100%
12	11	80%	100%	95%	95%
12	12	100%	100%	95%	100%
12	13	100%	100%	90%	95%
12	14	95%	100%	90%	95%
13	11	70%	100%	100%	90%
13	12	95%	95%	100%	90%
13	13	95%	100%	100%	90%
13	14	100%	95%	100%	90%
14	11	65%	100%	95%	100%
14	12	90%	100%	95%	100%
14	13	90%	100%	95%	100%
14	14	95%	100%	95%	100%

B.6.4.2 Inference (low-fidelity data)

We fit four candidate models (M^{11} , M^{12} , M^{13} , and M^{14}) from the alternative parameterisation to each incidence report.

B.6.4.2.1 Coverage

We summarise the results in Table B.22.

B.7 Fitting Cumberland’s data

In this section, we illustrate the procedure to infer the basic reproduction number (\mathfrak{R}_0) of influenza during the second wave of the **1918 pandemic** in Cumberland (Maryland). Specifically, we fit four model candidates from the **alternative parameterisation** to incidence data. The inference process yields almost identical \mathfrak{R}_0 estimates, regardless of the infectious period

distribution. Furthermore, we compare the results of the alternative parameterisation to those of the *traditional* one.

B.7.1 Incidence report

After enduring a wave of influenza infections during the spring of 1918, the U.S. Public Health Service organised special surveys in several localities to determine as accurately as possible the proportion of the population infected during the second wave of infections in the autumn of 1918. Fig B.42 shows the report of new cases detected in the city of Cumberland (Maryland) over that period.

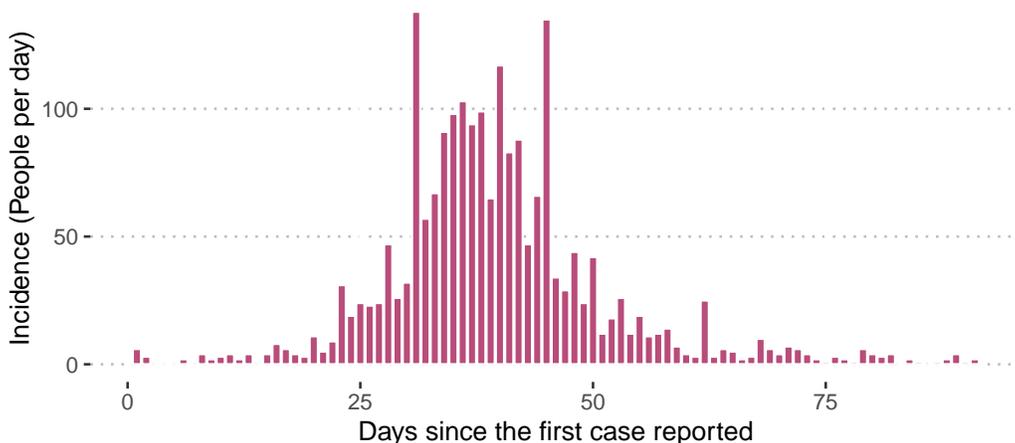


Figure B.42: Cumberland’s incidence data

B.7.2 Inference

We employ four candidates per parameterisation (traditional and alternative). On the one hand, the traditional parameterisation refers to the approach of fixing the mean of the epidemiological delays (latent and infectious periods) to values obtained from the literature, irrespective of their particular distribution. On the other hand, the proposed alternative parameterisation refers to the special emphasis placed on mean generation time of the **SEIR**, while flexibilising the mean and distribution of the epidemiological delays. Namely, the epidemiological delays can take any mean or shape provided that as a whole conform to the observed mean generation time. Furthermore, we assume an exponentially-distributed latent period (SE^1I^jR), where $j = \{1, 2, 3, 4\}$.

B.7.2.1 Incidence fit

Fig B.43 compares observed data (points) and against simulations from the full observational model. We stratify results by candidate and parameterisation. The solid line indicates the predicted mean and ribbons the 95% credible interval.

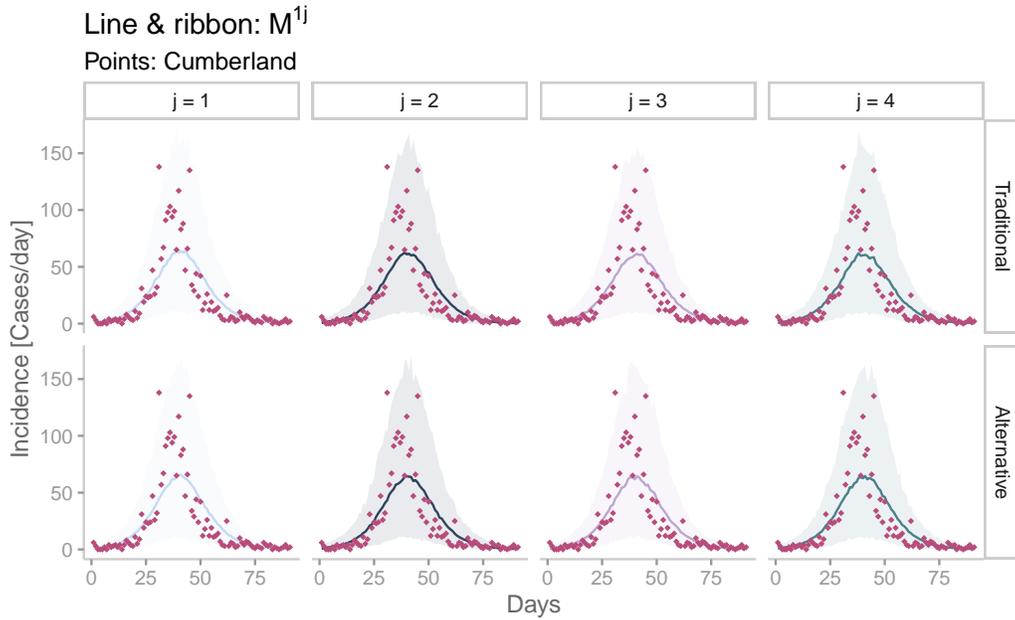


Figure B.43: Posterior predictive checks

B.7.2.2 Basic reproduction number (\mathcal{R}_0)

Fig B.44 compares estimates of basic reproduction number by model candidate and parameterisation

B.7. Fitting Cumberland's data

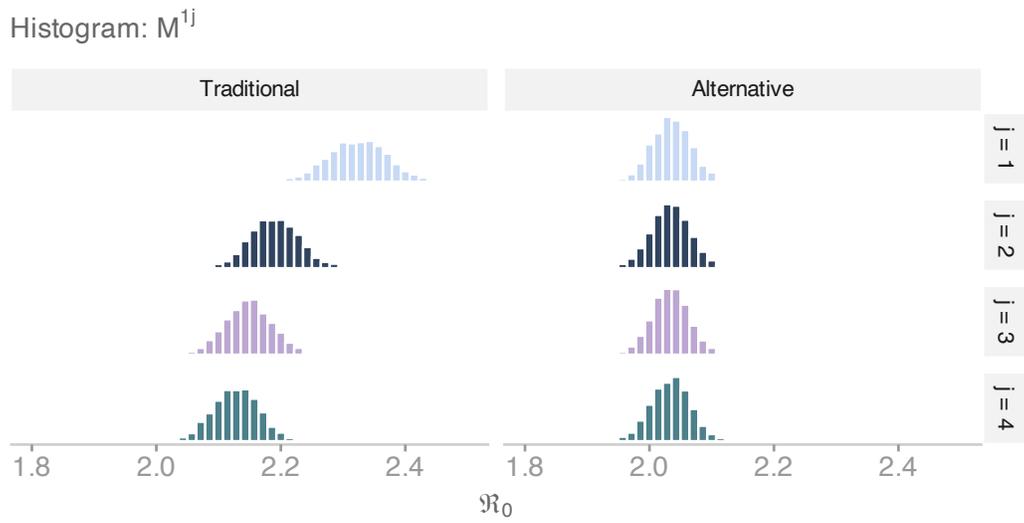


Figure B.44: \mathfrak{R}_0 estimates

Appendix C

Supplementary Information (Ch. 4)

C.1 Assumed parameters

Table C.1 shows the parameter values assumed for the within-host profile.

C.2 Formulation of semi-deterministic candidates

Model candidates comprise the amalgamation of process and measurement components. These candidates are semi-deterministic whose stochastic component corresponds to the formulation of the effective contact rate. This rate can have one of two formulations: Geometric Brownian Motion (GBM) or Cox-Ingersoll-Ross (CIR). Moreover, each process component can also be coupled with one of six measurement components, yielding 12 semi-deterministic structures (Table C.2).

C.2.1 Measurement components

$$\begin{aligned} \frac{dC}{dt} &= \eta P_t - C_t \delta(t \bmod 1) \\ y_d^1 &\sim Pois(C_t) \end{aligned} \tag{C.1}$$

C.2. Formulation of semi-deterministic candidates

Table C.1: Parameter values

Name	Symbol	Value	Units	Source
Incubation period	σ^{-1}	3	Days	Davies (2020)
Duration of preclinical infectiousness	η^{-1}	2.1	Days	Davies (2020)
Duration of clinical infectiousness	γ^{-1}	2.9	Days	Davies (2020)
Clinical fraction	ω	0.7	Unitless	HPSC (2020)
Asymptomatic infectious period	κ^{-1}	5	Days	Davies (2020)
Population	N_0	4937796	People	United Nations (2019)
Relative infectiousness	μ	0.5	Unitless	Davies (2020)
Initial susceptible	S_0	$N_0 - P_0$	People	Assumption
Initial exposed	E_0	0	People	Assumption
Initial clinical infectious	I_0	0	People	Assumption
Initial recovered	R_0	0	People	Assumption
Initial subclinical infectious	A_0	0	People	Assumption
Initial reported cases	C_0	0	People	By definition
Initial mobility effect	Z_0	1	Unitless	By definition

Table C.2: Semi-deterministic model candidates

Candidate	Process	Measurement
1	GBM	Eq C.1
2	GBM	Eq C.2
3	GBM	Eq C.3
4	GBM	Eq C.4
5	GBM	Eq C.5
6	GBM	Eq C.6
7	CIR	Eq C.1
8	CIR	Eq C.2
9	CIR	Eq C.3
10	CIR	Eq C.4
11	CIR	Eq C.5
12	CIR	Eq C.6

$$\begin{aligned}\frac{dC}{dt} &= \eta P_t - C_t \delta(t \bmod 1) \\ y_d^1 &\sim Nbin(C_t, \phi^{-1})\end{aligned}\tag{C.2}$$

$$\begin{aligned}\frac{dC}{dt} &= \eta P_t - C_t \delta(t \bmod 1) \\ y_d^1 &\sim Pois(C_t) \\ y_d^2 &\sim Normal(Z_t, \tau)\end{aligned}\tag{C.3}$$

$$\begin{aligned}\frac{dC}{dt} &= \eta P_t - C_t \delta(t \bmod 1) \\ y_d^1 &\sim Nbin(C_t, \phi^{-1}) \\ y_d^2 &\sim Normal(Z_t, \tau)\end{aligned}\tag{C.4}$$

$$\begin{aligned}\frac{dC}{dt} &= \eta P_t - C_t \delta(t \bmod 7) \\ y_w^1 &\sim Pois(C_t)\end{aligned}\tag{C.5}$$

$$\begin{aligned}\frac{dC}{dt} &= \eta P_t - C_t \delta(t \bmod 7) \\ y_w^1 &\sim Pois(C_t) \\ y_w^2 &\sim Normal(Z_t, \tau)\end{aligned}\tag{C.6}$$

C.3 Convergence tests

A critical requirement for performing inference on State-Space Models concerns the robustness of likelihood estimates. Namely, different runs must yield similar log-likelihood values from a single set of parameters (model configuration). To validate this requirement, we use five point estimates per group of candidates (Table C.3) as probes to assess the reliability of likelihood estimates. Depending on the specific measurement component of a candidate model, certain values of a probe are ignored. For instance, Candidate 1 ignores ϕ and τ values of all probes given that its measurement component does not require such parameters. Also, these point estimates correspond to *plausible* values obtained in an earlier iteration of this analysis. For the final version, we assume such values as **given**, and conduct the analysis from the start. Considering that these tests are not intended to be

C.3. Convergence tests

Table C.3: Probes

Candidates	Probe ID	ζ	P_0	τ	α	ϕ	ν	v
1 to 6	1	1.179	2.392	0.165	0.181	0.168	NA	NA
1 to 6	2	1.109	2.774	0.156	0.183	0.069	NA	NA
1 to 6	3	1.179	2.521	0.154	0.201	0.468	NA	NA
1 to 6	4	1.128	2.392	0.154	0.198	0.021	NA	NA
1 to 6	5	1.203	2.392	0.163	0.200	0.039	NA	NA
7 to 12	1	1.158	2.410	0.156	0.054	0.511	0.061	0.245
7 to 12	2	1.146	2.182	0.145	0.057	0.214	0.084	0.235
7 to 12	3	1.079	3.396	0.166	0.059	0.083	0.052	0.287
7 to 12	4	1.151	3.876	0.136	0.047	0.036	0.076	0.184
7 to 12	5	1.278	2.587	0.164	0.043	0.037	0.041	0.146

exhaustive but rather of exploratory nature, any arbitrary point is valid as a probe. For each probe and model candidate, we run the particle filter 16 times. We repeat the process for various particle and integration step sizes. For each batch of 16 runs, we estimate the log-likelihood mean and standard error.

In a nutshell, the results indicate that when one models daily incidence with the Poisson distribution, the particle filter does not converge as the number of particles increases. This finding suggests model misspecification under such an assumption. Furthermore, the accuracy of log-likelihood estimates is not compromised by relative large integration steps (e.g., $\frac{1}{8}$), which diminishes the burden on computational resources. On the contrary, and as expected in Monte Carlo simulation, the accuracy relies on the number of samples (particles). In the following subsections, we disaggregate the results by model candidate.

C.3.1 Candidate 1

This convergence test (Fig C.1) shows that the likelihood standard error **does NOT** tend to zero as the number of particles increases. We ascribe this result to model misspecification and exclude this candidate from further analyses.

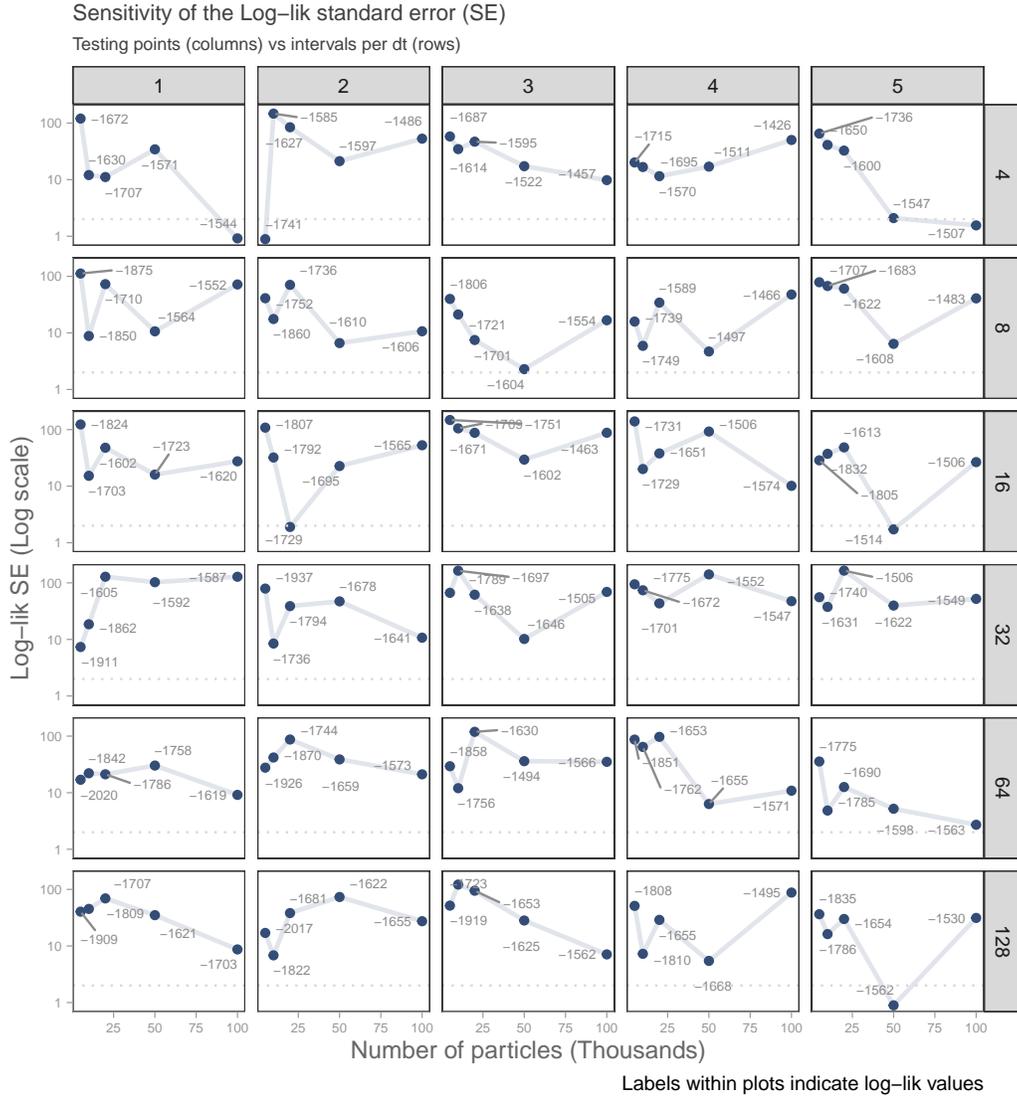


Figure C.1: Candidate 1's convergence test

C.3. Convergence tests

C.3.2 Candidate 2

Fig C.2 suggests that the likelihood standard error approximates zero as the number of particles increases.

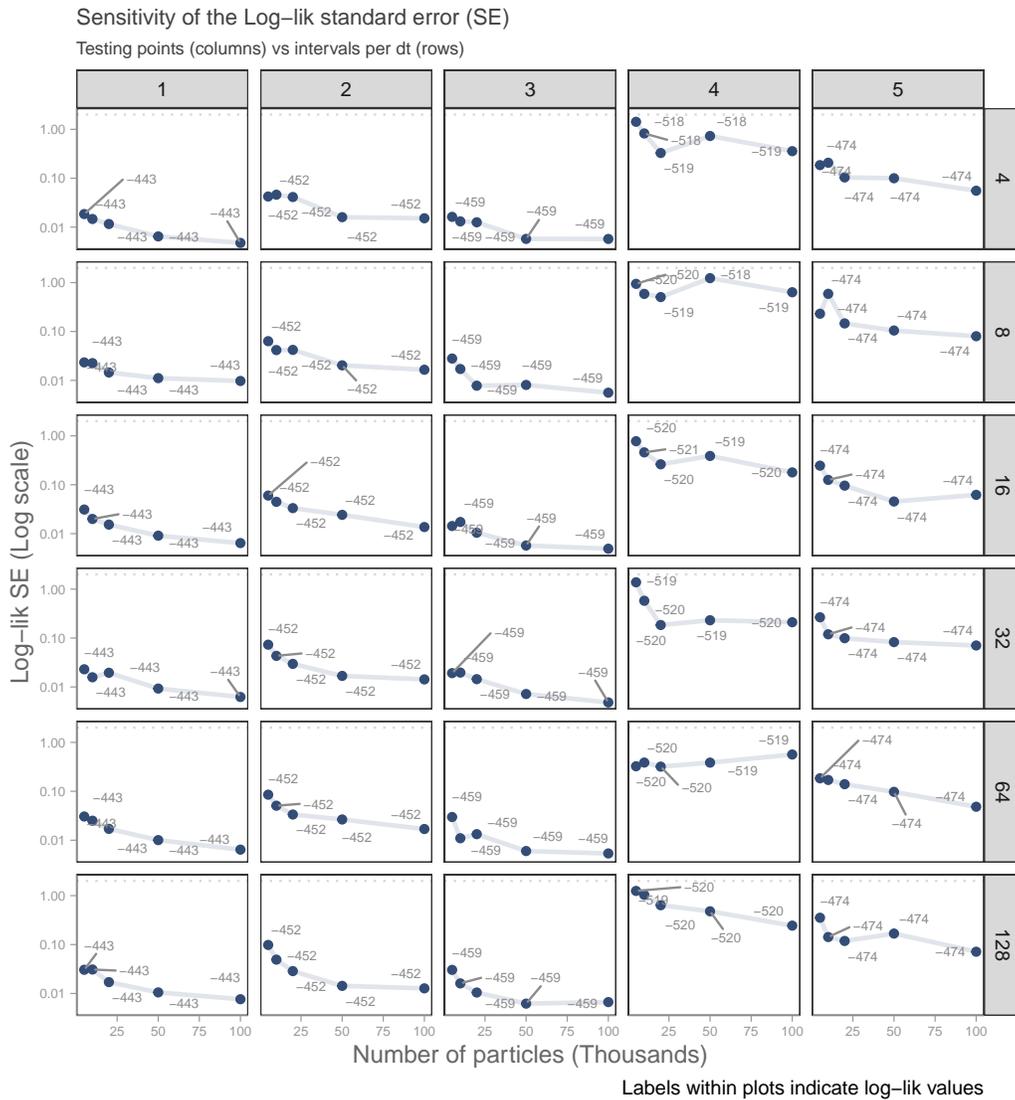


Figure C.2: Candidate 2's convergence test

C.3.3 Candidate 3

Fig C.3 suggests that mobility data does not redress convergence issues in the assumption that daily incidence measurements are distributed according to the Poisson distribution. Therefore, we exclude this candidate from further analyses.

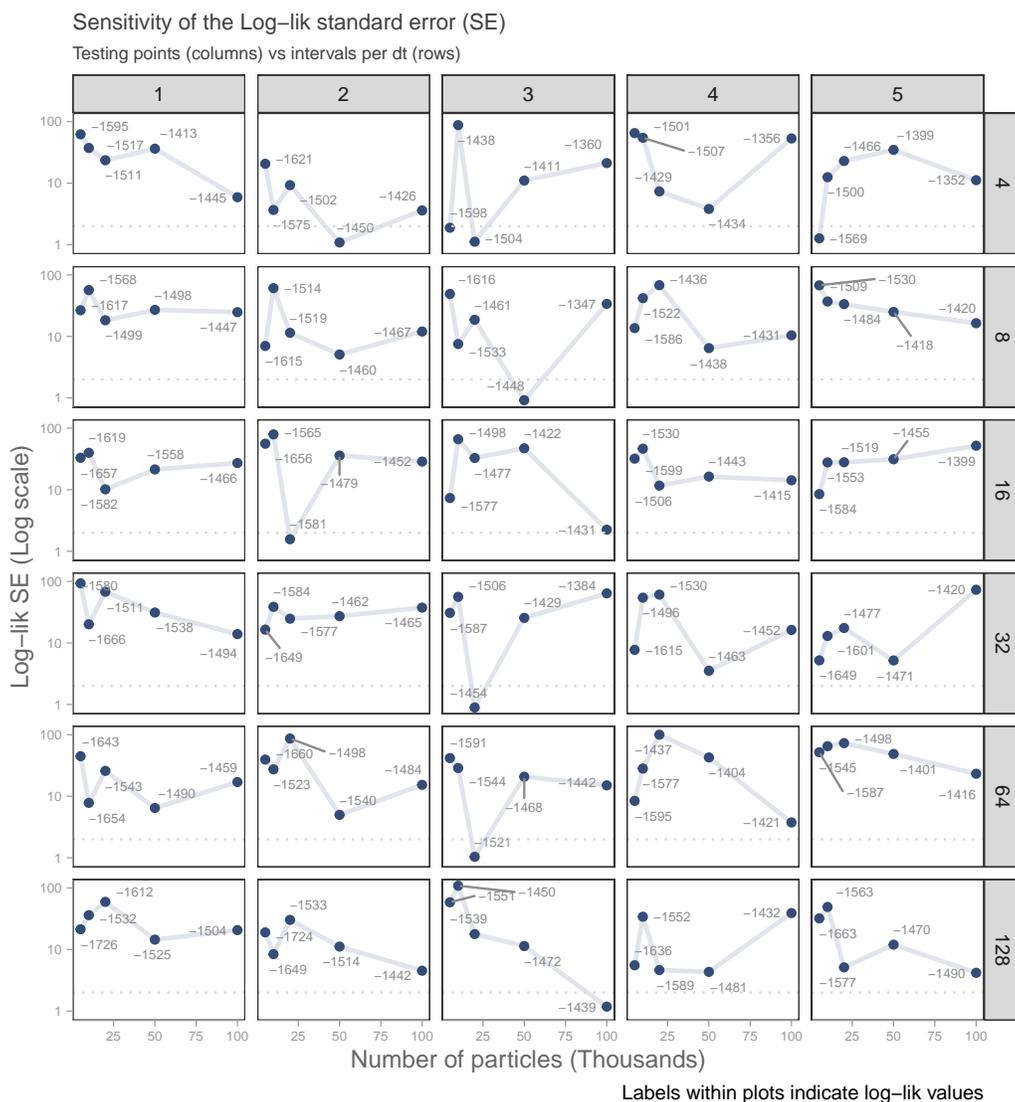


Figure C.3: Candidate 3's convergence test

C.3. Convergence tests

C.3.4 Candidate 4

Fig C.4 suggests that the likelihood standard error approximates zero as the number of particles increases.

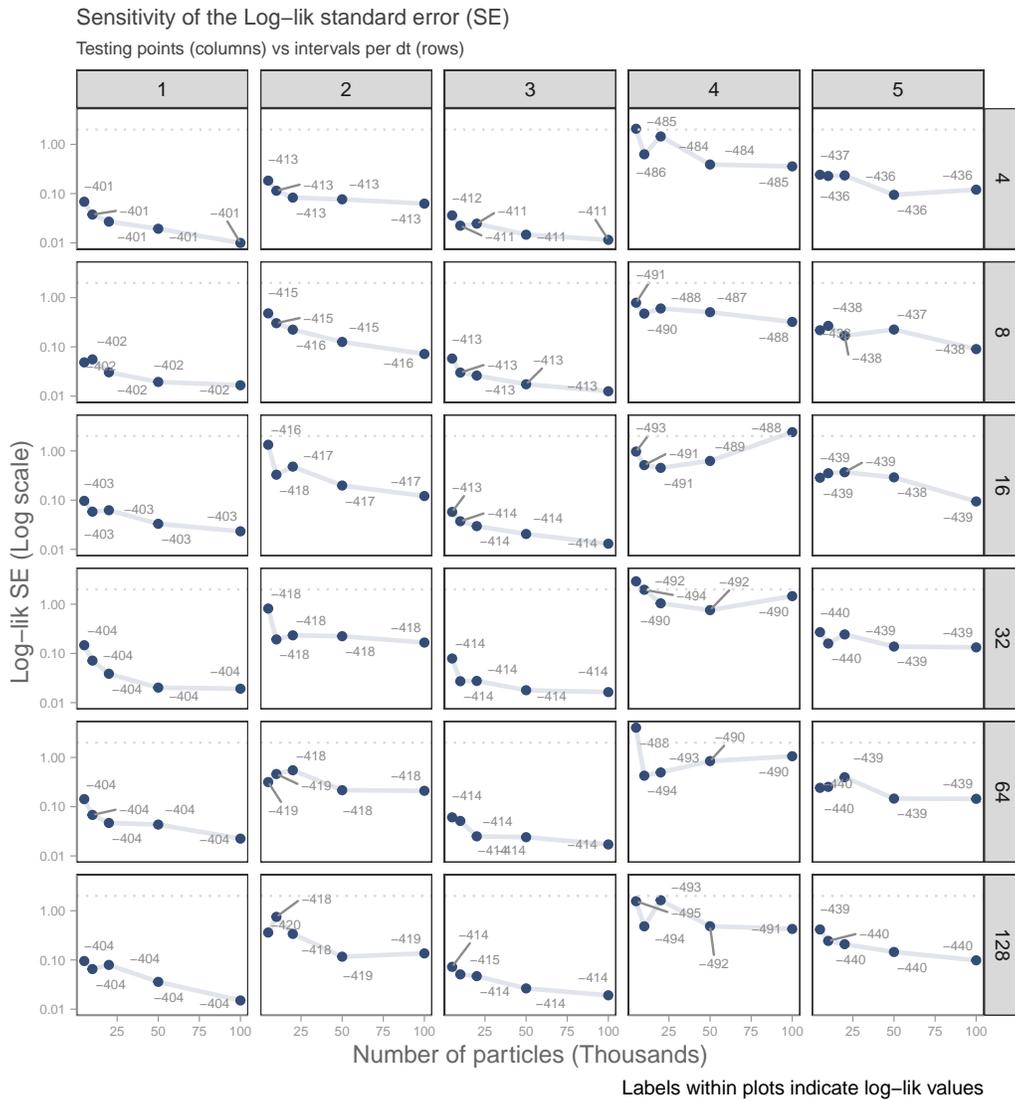


Figure C.4: Candidate 4's convergence test

C.3.5 Candidate 5

Fig C.5 suggests that the likelihood standard error approximates zero as the number of particles increases.

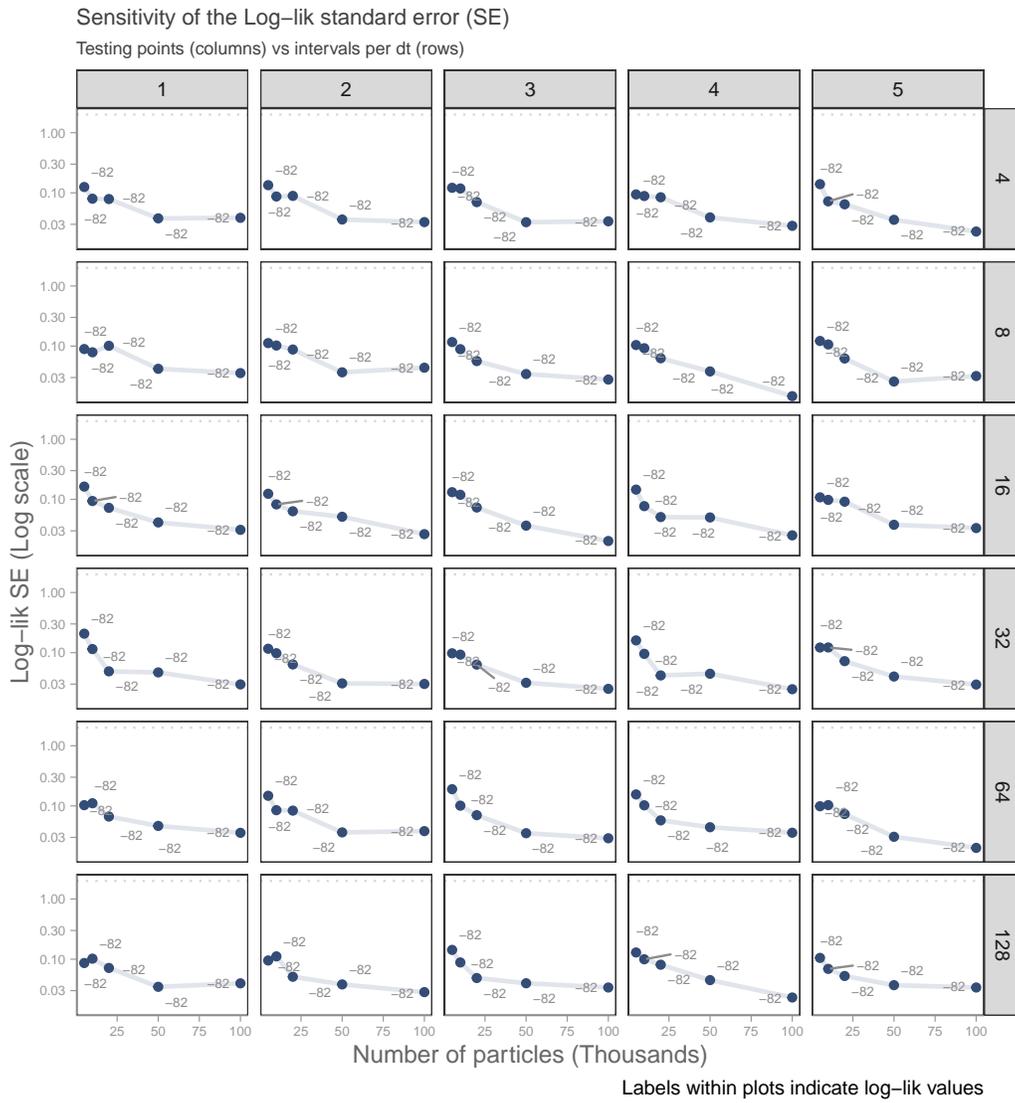


Figure C.5: Candidate 5's convergence test

C.3.6 Candidate 6

Fig C.6 suggests that the likelihood standard error approximates zero as the number of particles increases.

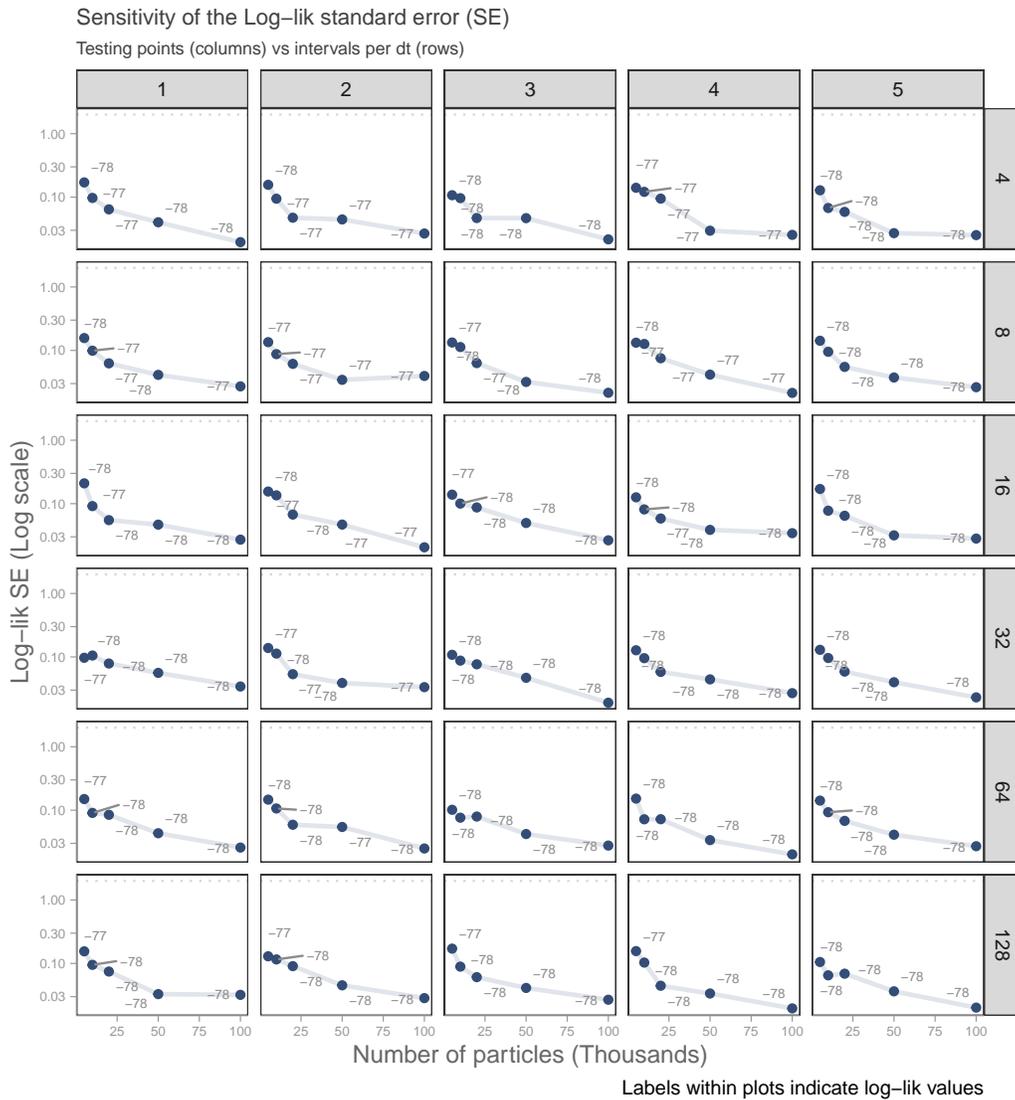


Figure C.6: Candidate 6's convergence test

C.3.7 Candidates 7-12

Candidates 7-12 exhibit the same pattern as that observed in Candidates 1-6. For brevity reasons, we only show one probe and one integration step per candidate in Fig C.7.

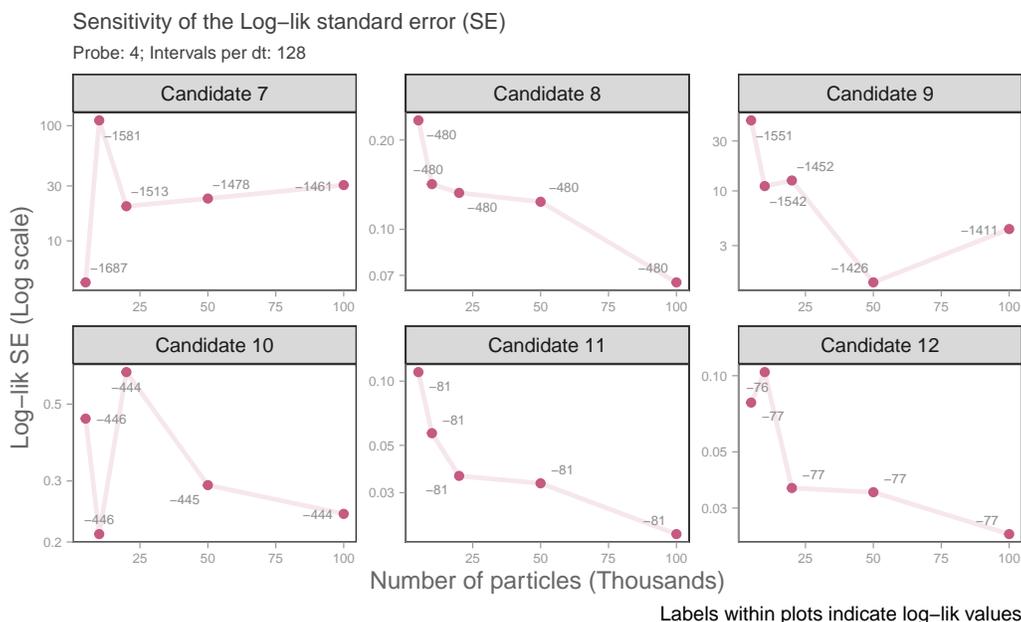


Figure C.7: Candidates 7-12 convergence test

C.4 Inference on DGP1's Candidate 2

This section illustrates the inference process performed on Candidate 2. This candidate's process model consists of an SEIR-type formulation whose relative effective contact rate is described by Geometric Brownian Motion. Moreover, the measurement model assumes that daily incidence counts are distributed according to the **Negative binomial distribution**. In particular, we apply *Iterated Filtering* and the *Particle Filter* to obtain estimates (via samples) for the effective reproductive number and other parameters that explain Ireland's first wave of COVID-19 in 2020.

C.4.1 Parameter inference

Table C.4: Candidate 2's estimated parameters

Name	Symbol	Units
Initial effective contact rate	ζ	People / day
Volatility of effective contact rate	α	Unitless
Variance of the measured transmission rate	τ	Unitless
Dispersion parameter	ϕ	Unitless
Initial preclinical infectious	P_0	People

C.4.1.1 Unknown parameters

Table C.4 indicates the parameters that will be subject to inference.

C.4.1.2 Local search

We start the inference process with a preliminary test. Specifically, we verify that Iterated Filtering algorithm converges to regions of high likelihood. Therefore, we search for the Maximum Likelihood Estimate (MLE) using Iterated Filtering from a single point in the parameter space. We repeat this process twenty times (Fig C.8).

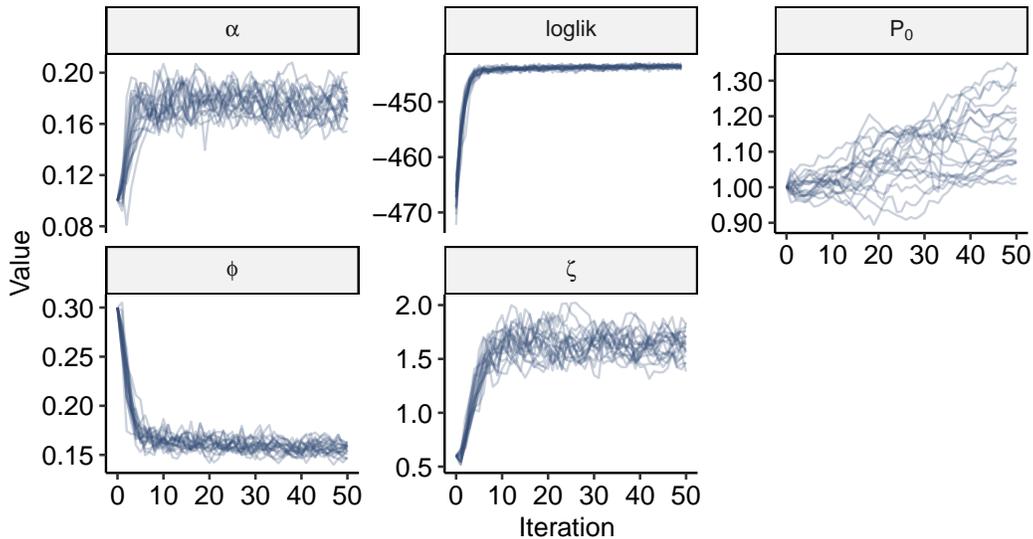


Figure C.8: Local search

C.4.1.3 Global search

In this step, we increase the number of starting points (300) and filtering iterations. Also, there is only one run for each starting point. We refer to this step as a *global search* (Fig C.9), which aims to construct a likelihood surface that allows us to identify regions of high plausibility.

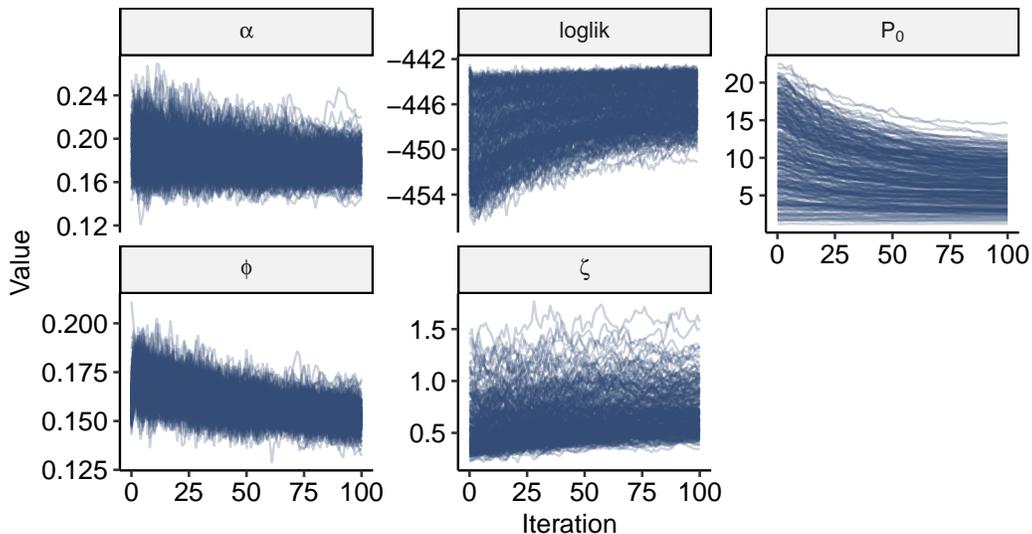


Figure C.9: Global search

C.4.1.4 Likelihood estimates

The likelihood estimates obtained from the Iterated Filtering algorithm are merely an approximation to the actual values at those points. This difference occurs for two reasons¹: the Iterated Filtering algorithm is run with fewer particles than are needed for a good likelihood evaluation; 2) the stochastic perturbations applied to the inferred parameters at each iteration. Consequently, it is necessary to run the Particle Filter to obtain reliable likelihood estimates. Specifically, we use the values from each run's final filtering iteration as inputs to the Particle Filter. In Fig C.10, grey dots denote starting points, whereas the other dots are the point estimates obtained from the Iterated Filtering algorithm. We notice that the estimates tend to converge to certain regions of the parameter space.

¹<https://kingaa.github.io/sbied/mif/slides.pdf>

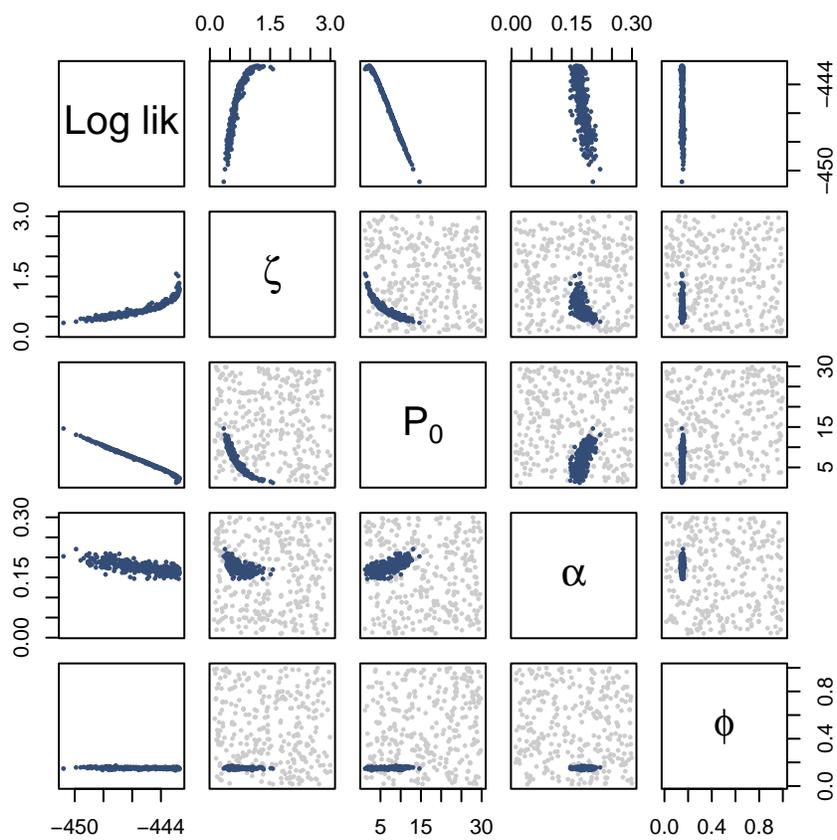


Figure C.10: Likelihood estimates

C.4.1.5 Profile likelihood

We employ the Profile Likelihood method to estimate confidence intervals. To illustrate the mechanics of this method, we illustrate each step followed to calculate the initial effective contact rate's (ζ) uncertainty bounds. We repeat this process for the remaining parameters. To begin with, we plot the initial effective contact rate's likelihood surface (Fig C.11) using the information from the global search to gain insight into the curvature of ζ . We subsequently define a region near the MLE (hypercube) from which we draw several hundreds of samples. The specific number of samples varies according to the complexity of exploring each parameter's space. For ζ , we draw 1,000 samples (Fig C.12).

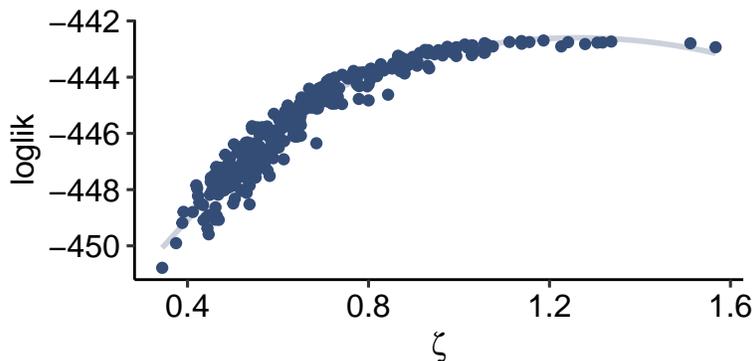


Figure C.11: Raw likelihood

For each sample, we run the Iterated Filtering algorithm, holding ζ constant, while the remaining parameters are perturbed (Fig C.13). As with local and global searches, we estimate, via the Particle Filter, the likelihood for each point estimate obtained from the Iterated Filtering algorithm. From this calculation, we construct the confidence intervals using the Profile Likelihood method and its refined version, the Monte Carlo-adjusted profile. The red dashed line indicates the cut-off at the 95% confidence level (Fig C.14).

Furthermore, after repeating the above process for each unknown parameter, we collate all likelihood estimates from the previous steps into a single database. The resulting likelihood surfaces exhibit quadratic shapes (Fig C.15). We, therefore, assume that these surfaces are approximations of the likelihood profiles. Following this assumption, we estimate each parameter's 95% confidence intervals. Overall, we employed three methods to calculate these quantities (Fig C.16).

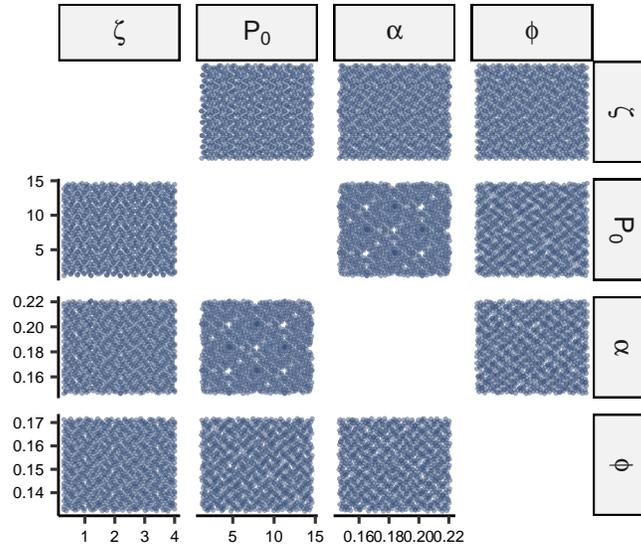


Figure C.12: Exploration space

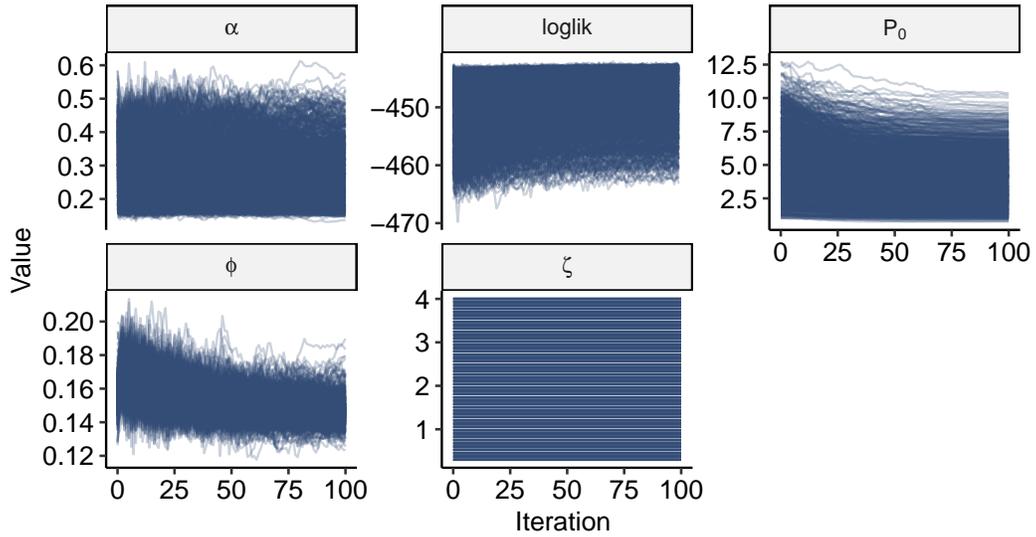


Figure C.13: MLE estimation while holding ζ constant

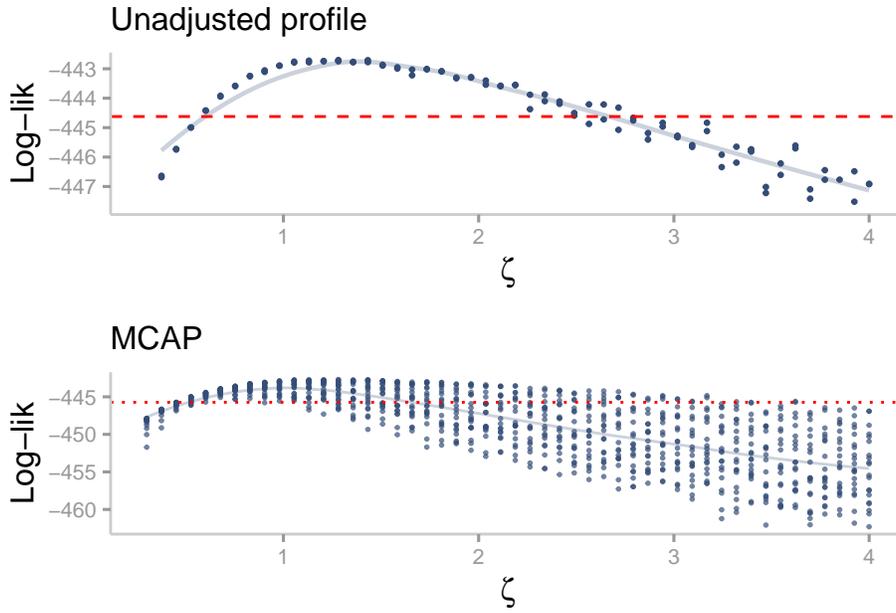


Figure C.14: ζ confidence intervals

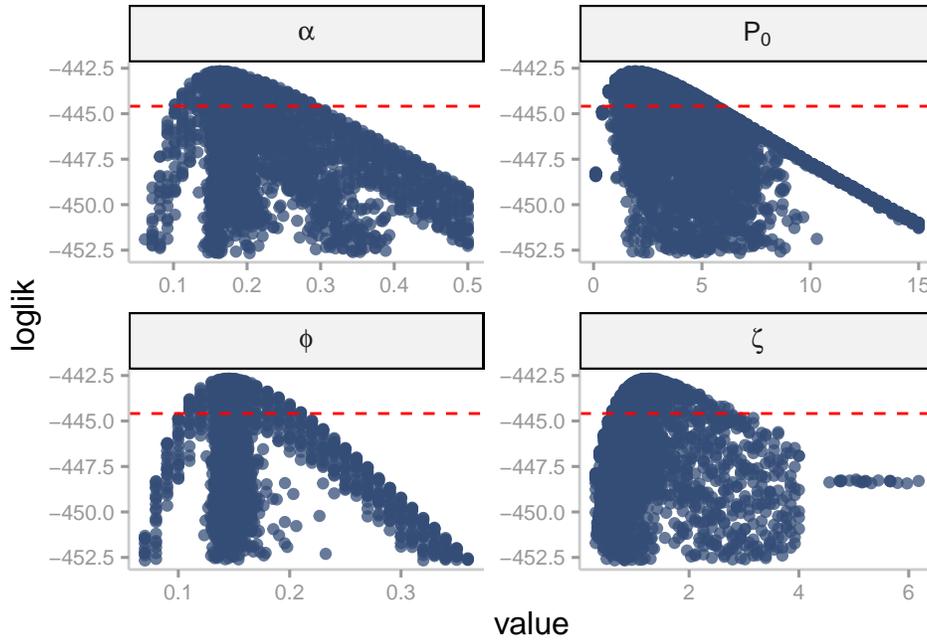


Figure C.15: Likelihood surface

C.5. Inference on DGP1's Candidate 4

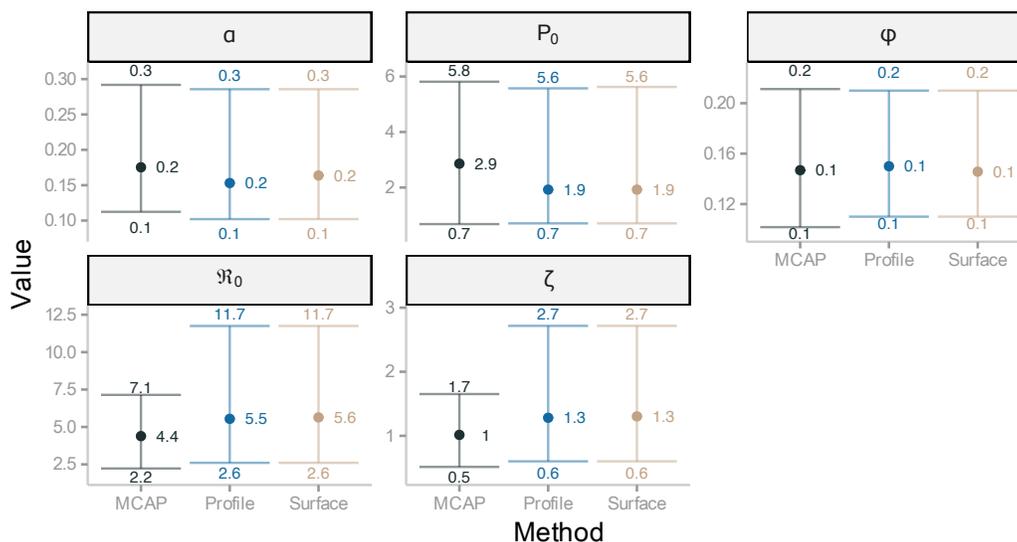


Figure C.16: 95% confidence intervals by method

C.4.2 Hidden states

We employ the filtering distribution to obtain predictions for the latent states. Since this distribution is intractable, we approximate it through simulation. To do so, first, we define a hypercube near the MLE (neighbourhood), from which we draw 200 samples to feed the Particle Filter. This method returns a set of draws weighted by its corresponding likelihood to approximate the filtering distribution at each time step (Fig C.17).

C.5 Inference on DGP1's Candidate 4

In this section, we illustrate the inference process performed on DGP1's Candidate 4. This candidate's process model consists of an SEIR-type formulation with a relative effective contact rate described by Geometric Brownian Motion. Furthermore, this candidate's measurement model assumes that daily incidence counts follow a Negative Binomial distribution and that mobility data serves as a proxy for the relative effective contact rate. As we replicate the inference process described in the previous section, we present the results succinctly.

C.5.1 Parameter inference

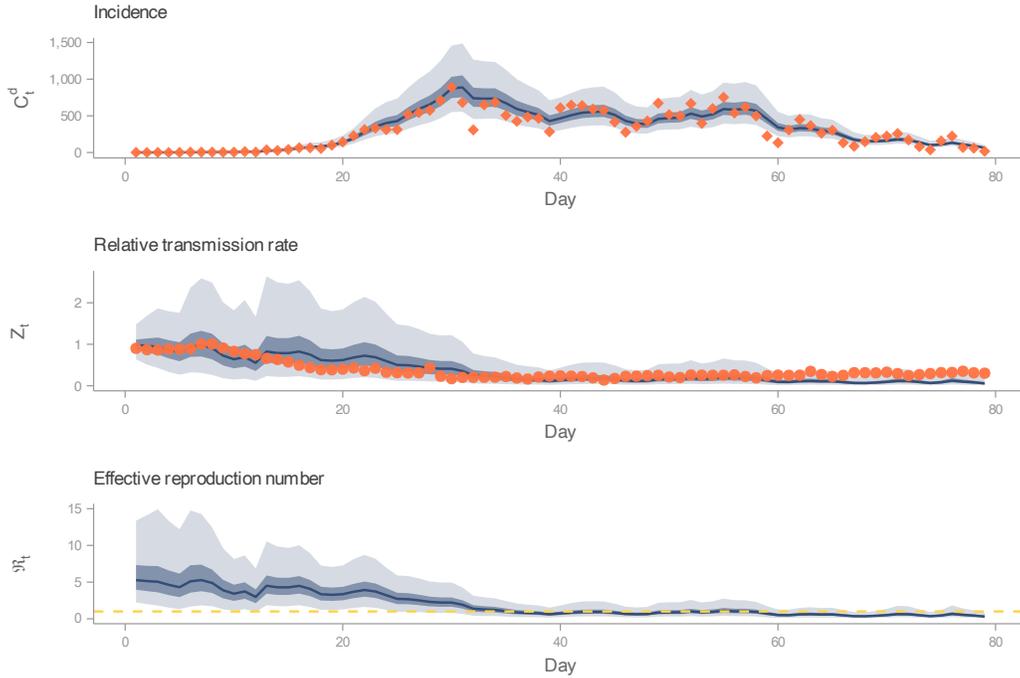


Figure C.17: Candidate 2's hidden states

C.5.1.1 Unknown parameters

Table C.5 shows the parameters we will infer.

C.5.1.2 Profile likelihood

Fig C.18 shows Candidate 4's parameter estimates.

Table C.5: Candidate 4's estimated parameters

Name	Symbol	Units
Initial effective contact rate	ζ	People / day
Volatility of effective contact rate	α	Unitless
Variance of the measured transmission rate	τ	Unitless
Dispersion parameter	ϕ	Unitless
Initial preclinical infectious	P_0	People

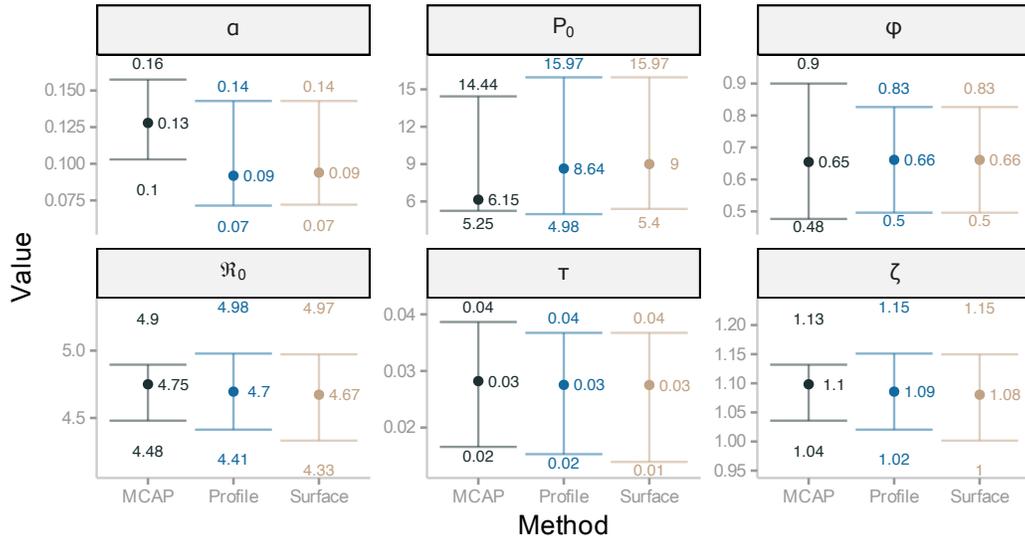


Figure C.18: 95% confidence intervals by method

C.5.2 Hidden states

Fig C.19 displays Candidate 4's filtering distribution.

C.6 Inference on DGP1's Candidate 5

This section illustrates the inference process performed on Candidate 5. This candidate's process model consists of an SEIR-type formulation whose relative effective contact rate is described by Geometric Brownian Motion. Moreover, this candidate's measurement model assumes that weekly incidence counts are distributed according to the **Poisson distribution**.

C.6.1 Parameter inference

C.6.1.1 Unknown parameters

Table C.6 shows the parameters we will infer.

C.6.1.2 Profile likelihood

Fig C.20 shows Candidate 5's parameter estimates.

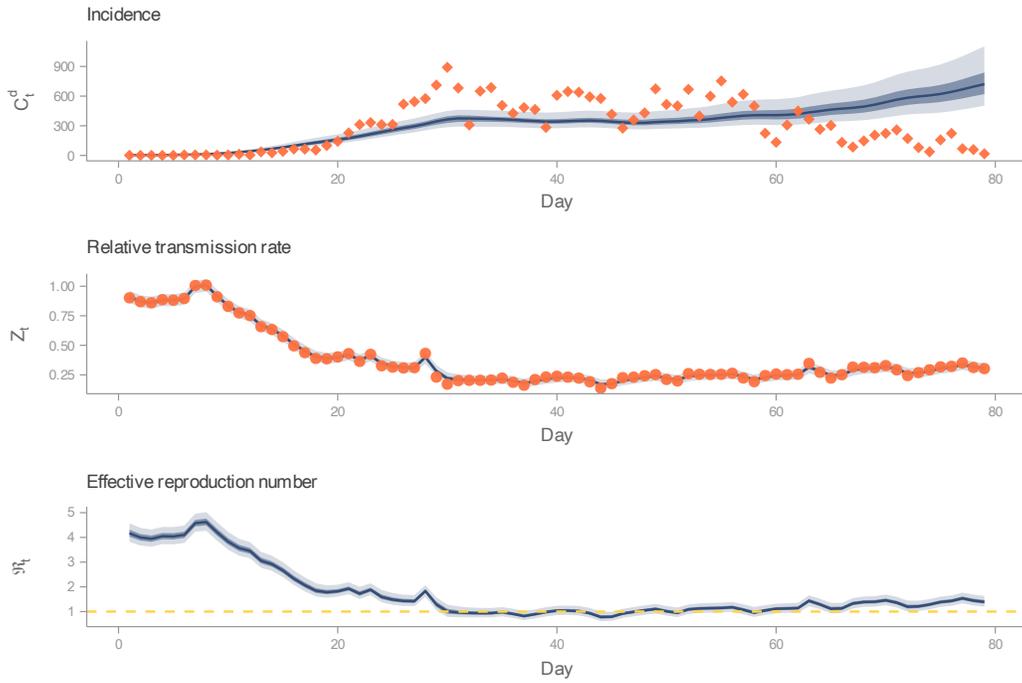


Figure C.19: Candidate 4's hidden states

Table C.6: Candidate 5's estimated parameters

Name	Symbol	Units
Initial effective contact rate	ζ	People / day
Volatility of effective contact rate	α	Unitless
Initial preclinical infectious	P_0	People

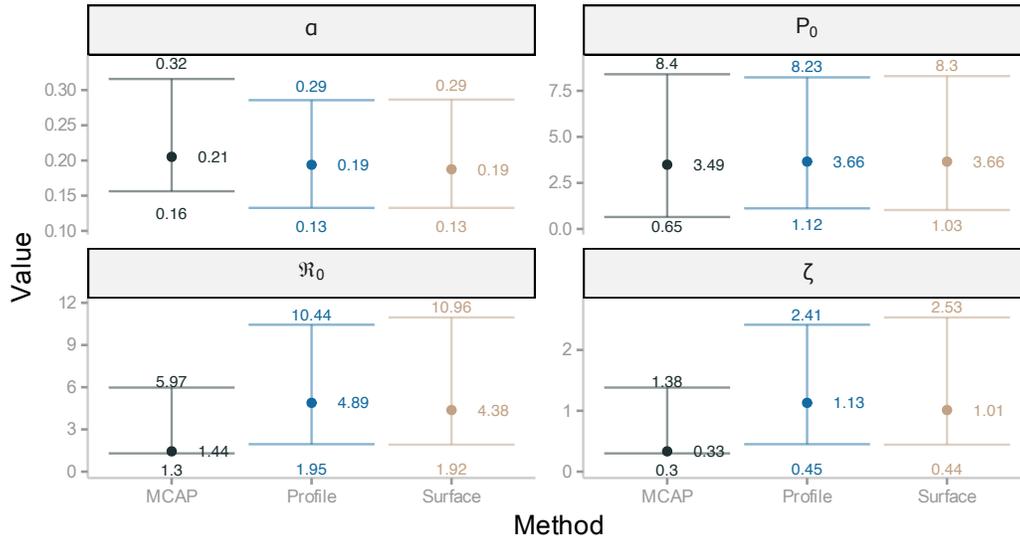


Figure C.20: 95% confidence intervals by method

C.6.2 Hidden states

Fig C.21 displays Candidate 6's filtering distribution.

C.7 Inference on DGP1's Candidate 6

This sections illustrates the inference process performed on Candidate 6. This candidate's process model consists of an SEIR-type formulation whose relative effective contact rate is described by Geometric Brownian Motion. Moreover, this candidate's measurement model assumes that weekly incidence counts are distributed according to the **Poisson distribution**, and that **mobility data is a proxy** measurement for the relative effective contact rate.

C.7.1 Parameter inference

C.7.1.1 Unknown parameters

Table C.7 shows the parameters we will infer.

C.7.1.2 Profile likelihood

Fig C.22 shows Candidate 6's parameter estimates.

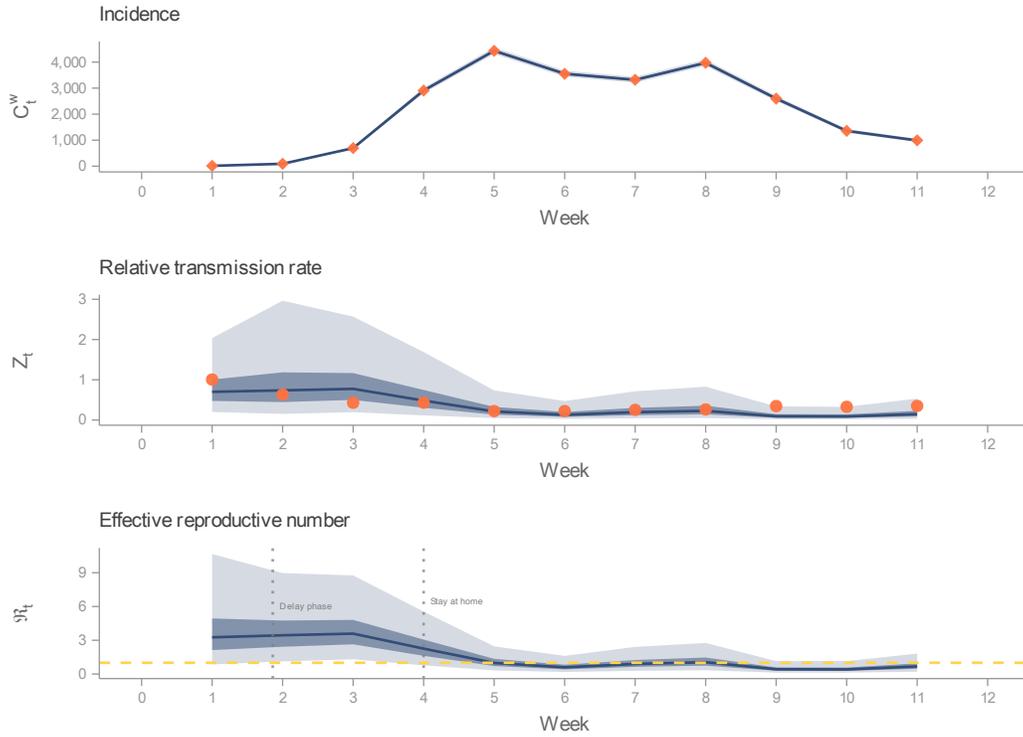


Figure C.21: Candidate 5's hidden states

Table C.7: Candidate 6's estimated parameters

Name	Symbol	Units
Initial effective contact rate	ζ	People / day
Volatility of effective contact rate	α	Unitless
Variance of the measured transmission rate	τ	Unitless
Initial preclinical infectious	P_0	People

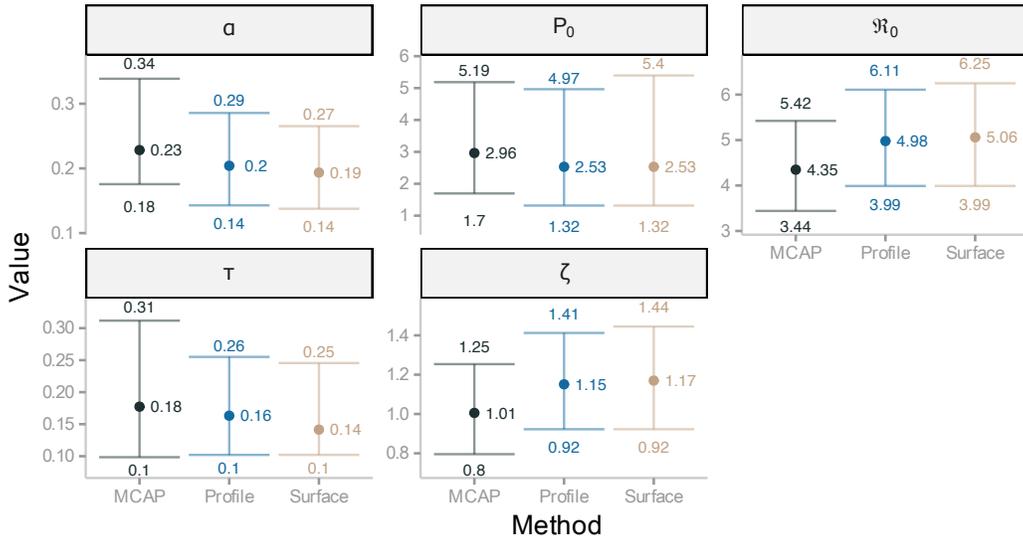


Figure C.22: 95% confidence intervals by method

C.7.2 Hidden states

Fig C.23 displays Candidate 5's filtering distribution.

C.8 Inference on DGP2

This section illustrates the inference process performed on DGP2. Its process model consists of an SEIR-type formulation whose relative effective contact rate is described by a Cox-Ingersoll-Ross structure. Moreover, DGP2's measurement model assumes that weekly incidence counts are distributed according to the **Poisson distribution**, and that **mobility data is a proxy** measurement for the relative effective contact rate.

C.8.1 Parameter inference

C.8.1.1 Unknown parameters

Table C.8 shows the parameters we will infer.

C.8.1.2 Profile likelihood

Fig C.24 shows Candidate 6's parameter estimates.

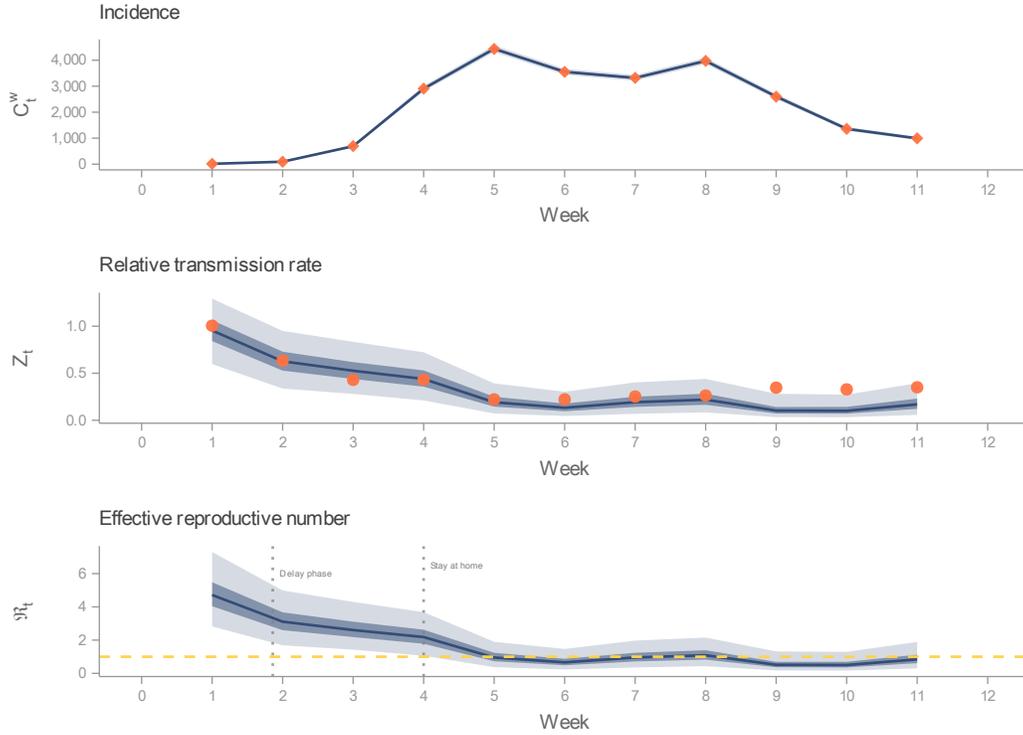


Figure C.23: Candidate 6's hidden states

Table C.8: DGP2's estimated parameters

Name	Symbol	Units
Initial effective contact rate	ζ	People / day
Volatility of effective contact rate	α	Unitless
Variance of the measured transmission rate	τ	Unitless
Initial preclinical infectious	P_0	People
Long-term goal	v	Unitless
Adjustment speed	ν	day^{-1}

C.8. Inference on DGP2

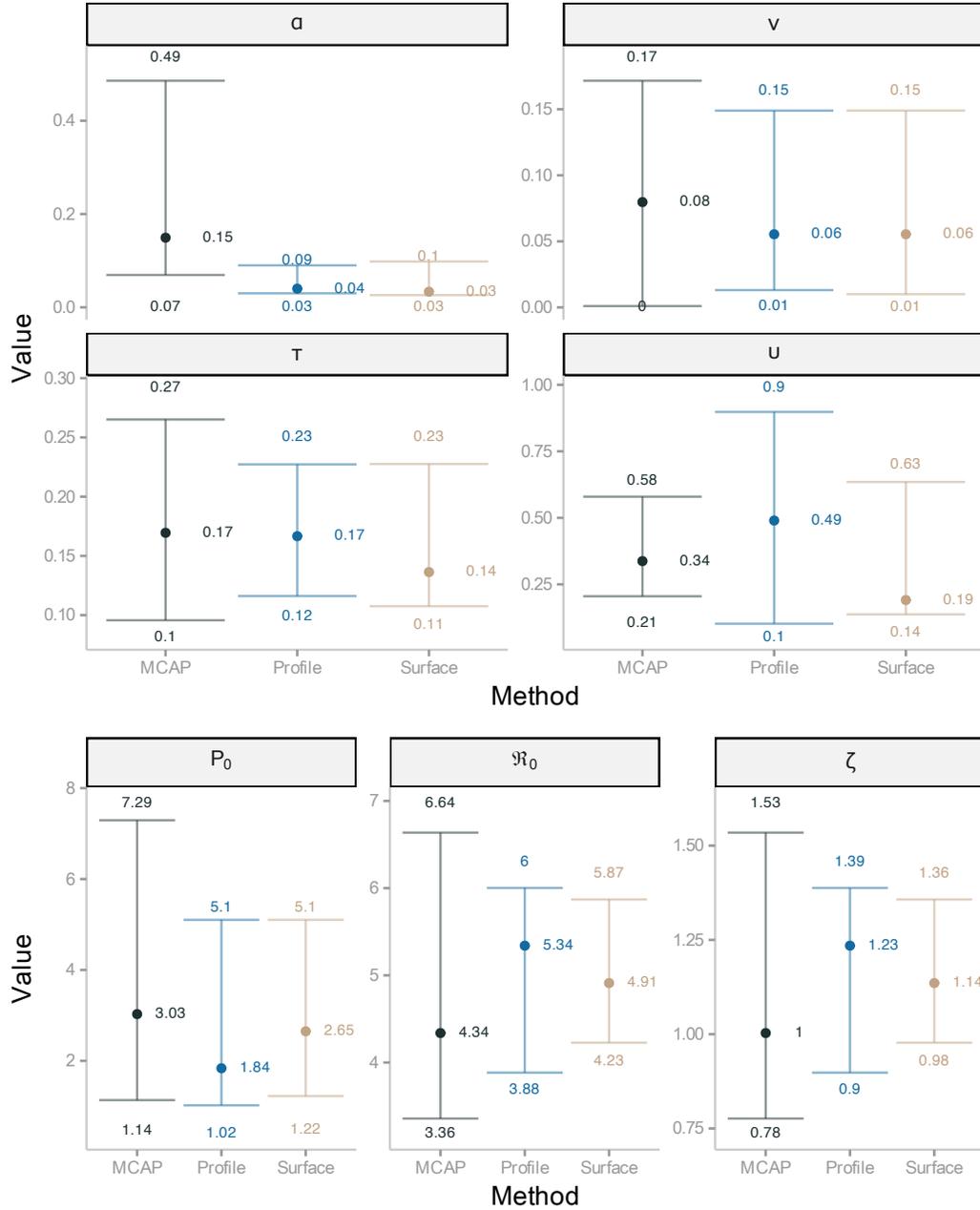


Figure C.24: 95% confidence intervals by method

C.8.2 Hidden states

We model the MLE’s neighbourhood as a copula from which 200 samples are drawn. Then, we run the particle filter and estimate their likelihood. Even though the copula prevents, to some extent, the exploration of undesired regions of the parameter space that may bias the results, some runs yield abnormal likelihood values. Fig C.25 shows likelihood values in a violin plot at different cut-offs. Here, we notice that the likelihood concentrates on a common region, but several outliers skew the results. This outcome suggests that using all the samples may bias the estimates towards low probability regions, or even worse, produce computational overflows. Thus, we opt for a cut-off of 20 log-likelihood units (or 485×10^6 likelihood units) given that it contains 67 % of all the starting points and appears to include only a few outliers. Such *outliers* are concentrated at the edges of the likelihood surface (high values of v and ν). These results highlight the complex surface created by this particular high-dimensional DGP. Fig C.26 below compares included (light colour) and excluded (dark colour) parameter values. After excluding outliers, we estimate the filtering distribution (Fig C.27).

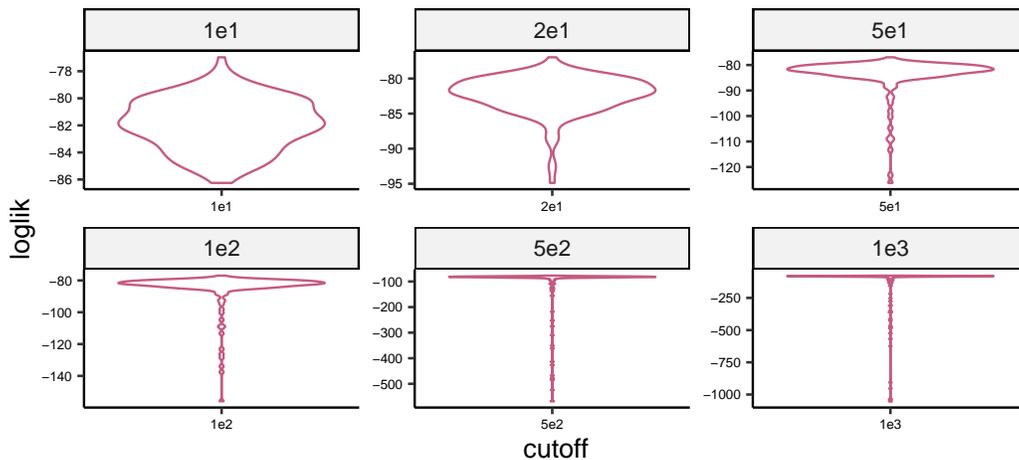


Figure C.25: Likelihood estimates

C.9 Inference on DGP3

This section aims to illustrate the inference process performed on DGP3. The formulation of this structure involves nine candidate deterministic process components (PM3) and the assumption that the observation of daily cases

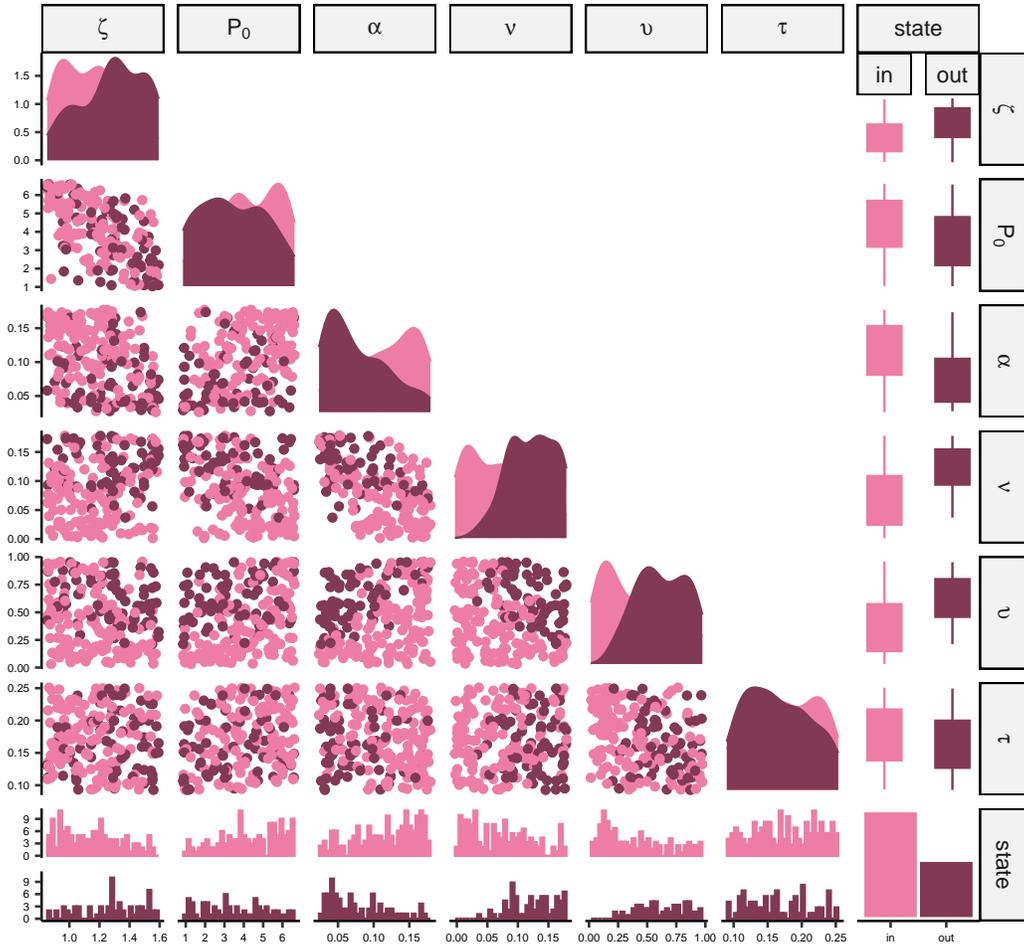


Figure C.26: Analysis of outliers

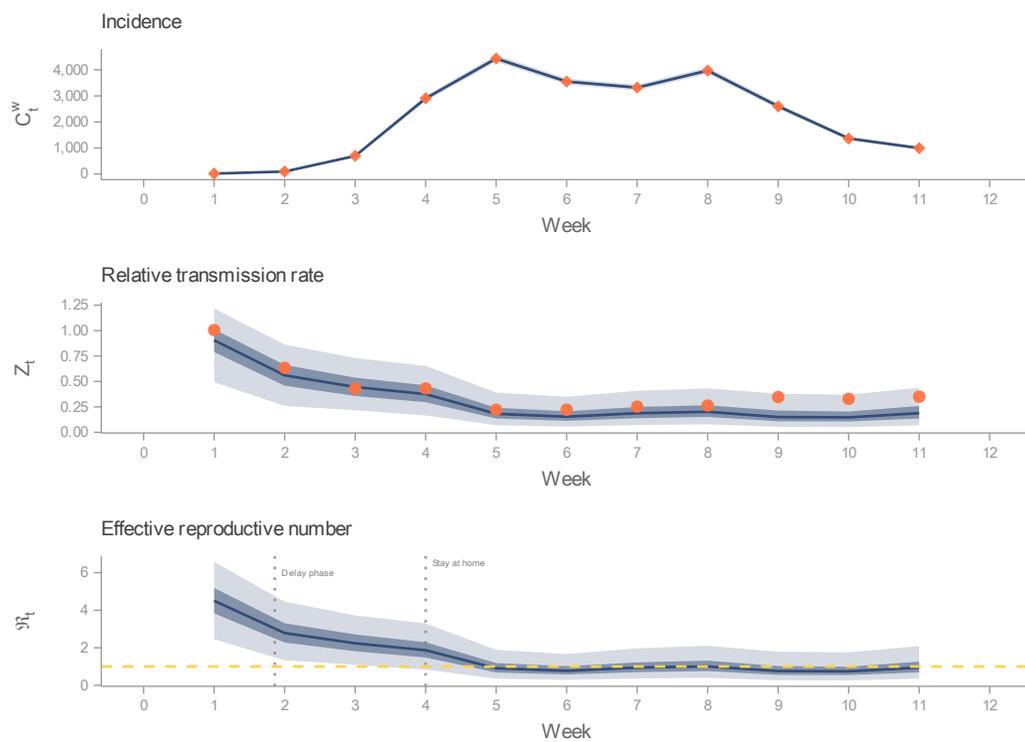


Figure C.27: DGP's hidden states

follows the Poisson distribution. We approach the inference process from a Bayesian perspective, which entails the estimation of posterior distributions. We estimate these distributions using Hamiltonian Monte Carlo (HMC).

C.9.1 Prior distributions

For the nine candidate models, we formulate the priors presented in Fig C.28.

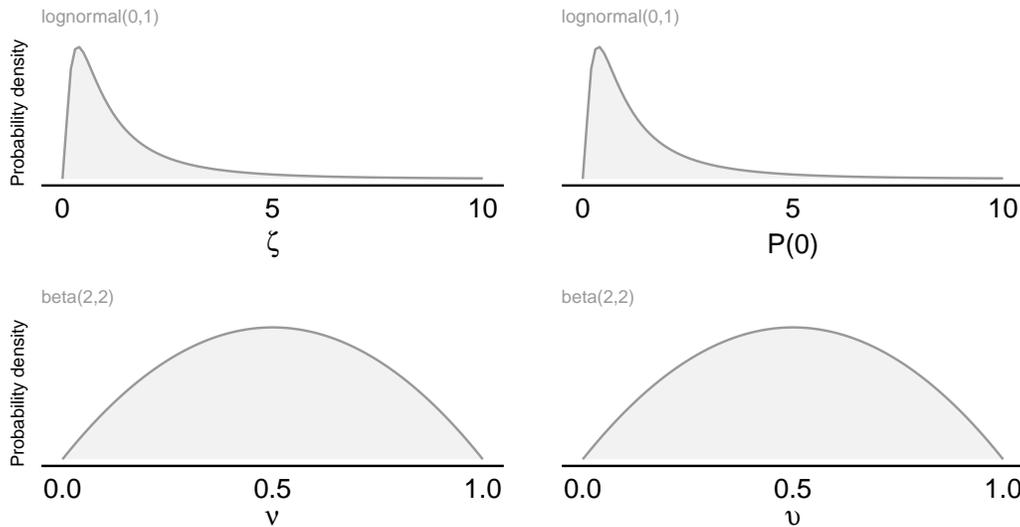


Figure C.28: Prior distributions

C.9.2 Posterior distributions

For each candidate fitted to the incidence data, Stan indicates that no pathological behaviour was present during the sampling process. Additionally, the sampled draws yield adequate Effective Sample Sizes and potential scale reduction factors. These outcomes suggest that the Markov chains converged.

C.9.2.1 Likelihood by delay order

Fig C.29 shows the 95% credible intervals of likelihood estimates by model candidate.

C.9.2.2 Fit accuracy

In the main text, we present a plot that compares predicted incidences and relative contact rates against incidence and mobility data, respectively. Here,

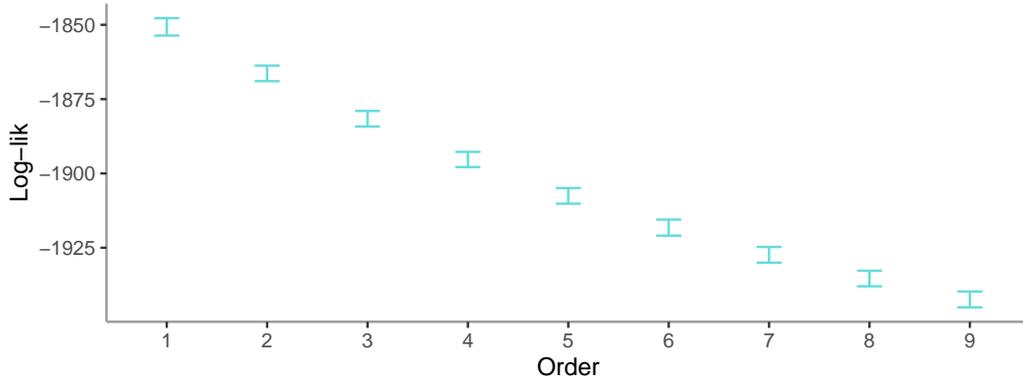


Figure C.29: Log-likelihood estimates by model candidate

we provide a quantitative assessment (Fig C.30) of how well the predicted values capture the data using the mean absolute scale error (MASE). Values below one (dotted line) indicate good performance.

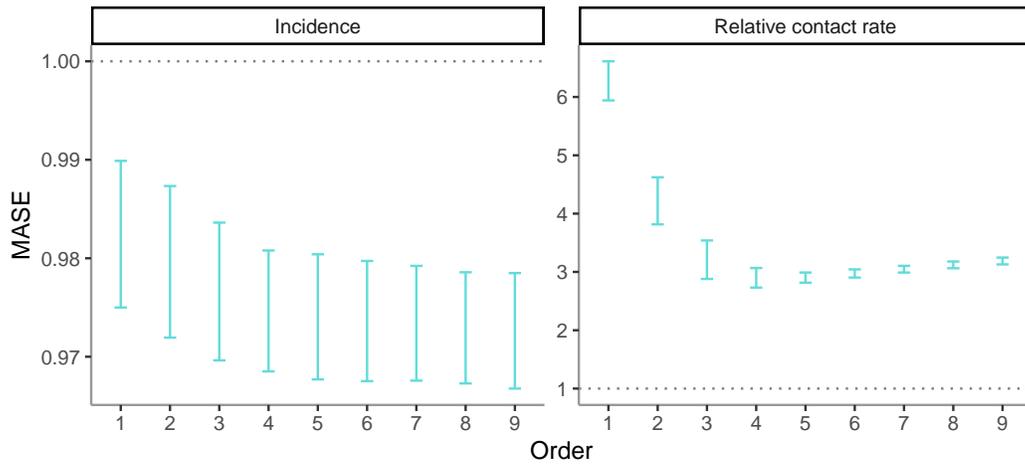


Figure C.30: MASE by indicator. Error bars indicate 95% credible intervals.

C.9.2.3 Parameter estimates

Fig C.31 shows parameter estimates by delay order. Here, we notice that the standard deviations of ν and v are noticeably small. In other words, the probability mass is located in a low-volume and high-density region of the parameter space.

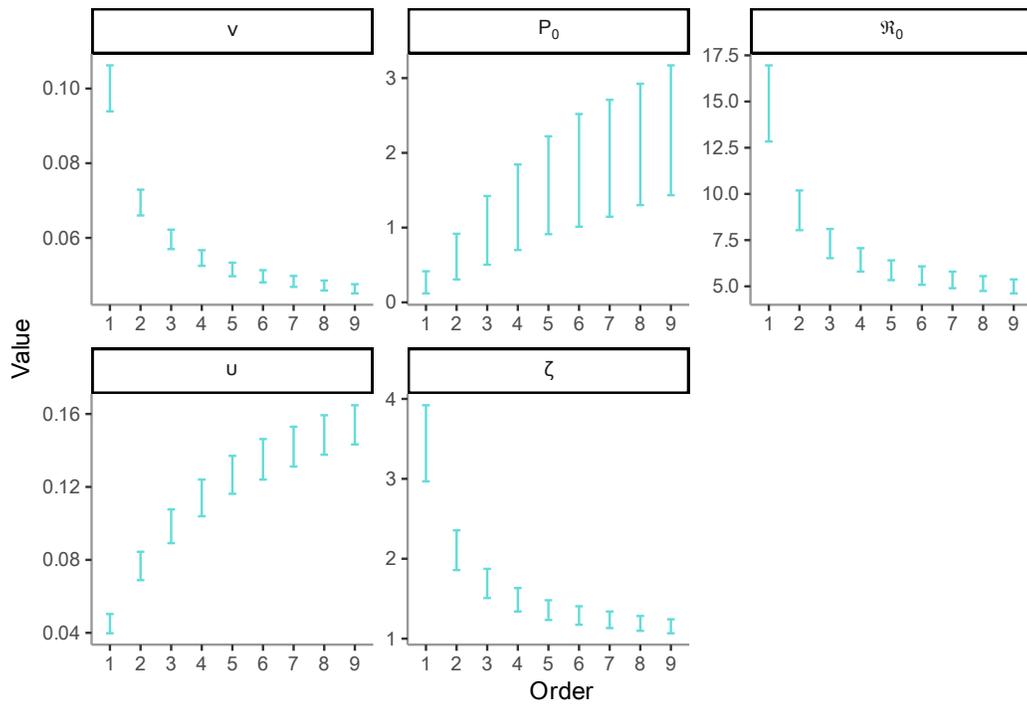


Figure C.31: Parameter estimates. Error bars indicate 95% credible intervals.

C.9.3 Alternative measurement component

$$y_a^1 \sim NBin(C_t, \phi) \tag{C.7}$$

Misspecification in the measurement component, such as unaccounted overdispersion and unmodelled variability, can lead to overly confident conclusions (Bretó 2018) or biased estimates. Above, we employed a stringent measurement model (Poisson), which ties the observation mean and variance. In this section, we replace the Poisson model with the Negative Binomial distribution (Eq (C.7)), denoted by Nbin, a structure that allows the DGP to handle overdispersion (if present) in the observations. Although the Nbin is more flexible, it also increases the DGP’s complexity by adding a new parameter: ϕ . The reader should recall that as $\phi \rightarrow \infty$, NBin converges to the Poisson distribution.

In order to understand these new parameter spaces, we fit the daily incidence data to the nine process model candidates, which are coupled with the NBin observational model. Here, we assume ζ , ν , v , ϕ and P_0 as unknown parameters. For each model, we run eight Markov chains from different starting points. The results indicate that this alternative measurement component yields a complex bimodal posterior distribution (Fig C.32). That is, chains reach either of two equilibrium regions.

To elaborate on this bimodality feature, we focus on candidate 4’s trace plot (Fig C.33). Here, we notice a distinctive pattern. Chains settle either on high-density (log-likelihood)/low-volume (narrow-band chains) or low-density/high-volume regions. Furthermore, Stan diagnostics (see Github repository) confirm such pathological behaviour in this parameter space by signalling the occurrence of divergent transitions and abnormal *energies*. Interestingly, Stan only detects divergences and abnormal energy values in the low-density/high-volume region. We thus refer to chains in the high-density region as *well-behaved* chains.

Further, we consider incidence fit as a more immediate appraisal. That is, we compare the predicted incidence against the actual data, discriminating by chain type (Fig C.34). Here, it can be seen that only the well-behaved chains fit the data. On the other hand, parameter estimates from these chains (Table C.9) provide similar insights to those obtained from the Poisson distribution. For instance, both distributions yield notably narrow estimates for ν & v , which determine the dynamics of the relative effective contact rate. Admittedly, there is a slight bias in estimates from the Poisson distribution,

C.9. Inference on DGP3

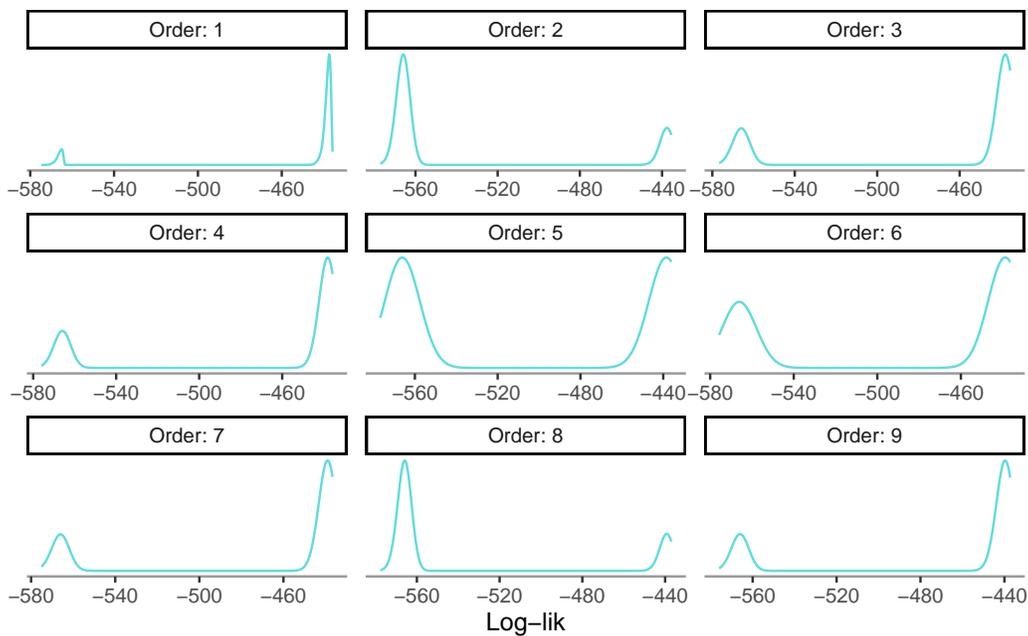


Figure C.32: Log-lik density by candidate

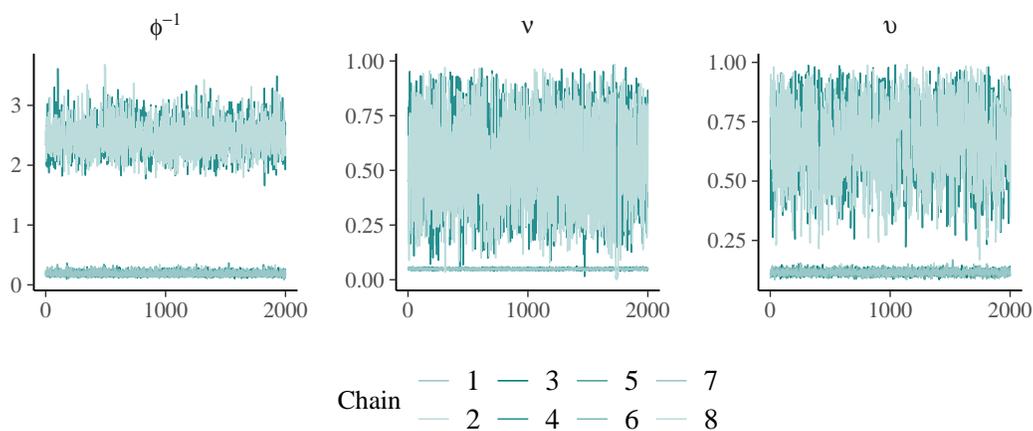


Figure C.33: Candidate 4's trace plot. This candidate has been amalgamated with the Nbin distribution.

Table C.9: Parameter estimates from the well-behaved chains

Parameter	Mean	SD	2.5%	97.5%
P_0	1.712	0.678	0.714	3.346
\mathfrak{R}_0	5.686	0.559	4.716	6.912
ν	0.049	0.003	0.043	0.056
v	0.115	0.010	0.097	0.135
ζ	1.315	0.129	1.091	1.598

which overestimates \mathfrak{R}_0 . However, it should be remarked that ignoring pathological chains is not a sound approach. Namely, we cannot assume that the parameter space is well-behaved when the evidence tells otherwise. Thus, we employ such estimated values for comparison and exploration purposes rather than for an inference one.

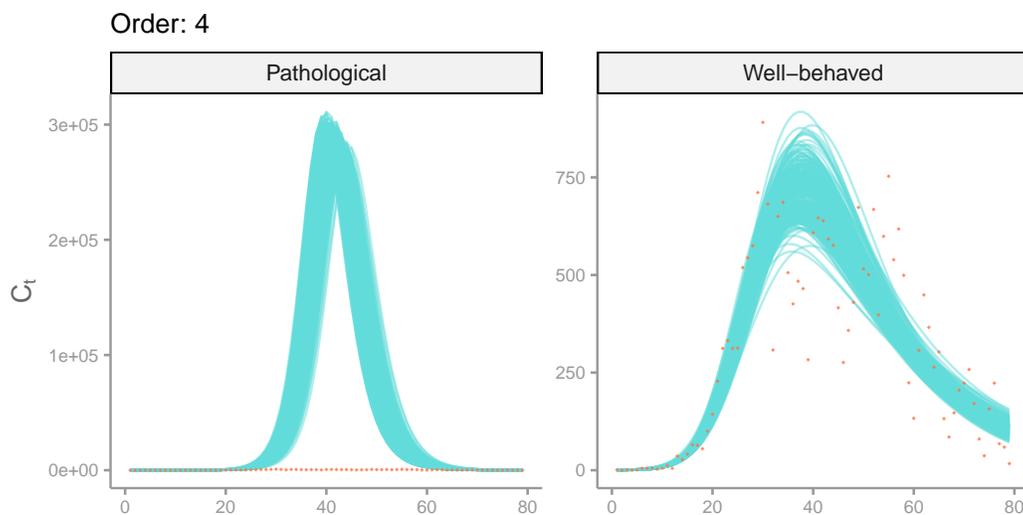


Figure C.34: Incidence fit region

To provide further evidence of the complexity generated by the Nbin distribution, we identify that bimodality persists even with only one unknown parameter in the 1st-order delay model (Fig C.35). To illustrate this finding, we assume as unmodelled predictors or known values, the mean values of the well-behaved chains. We do so for parameters P_0 , v , ν , and ϕ , leaving ζ as the only unknown.

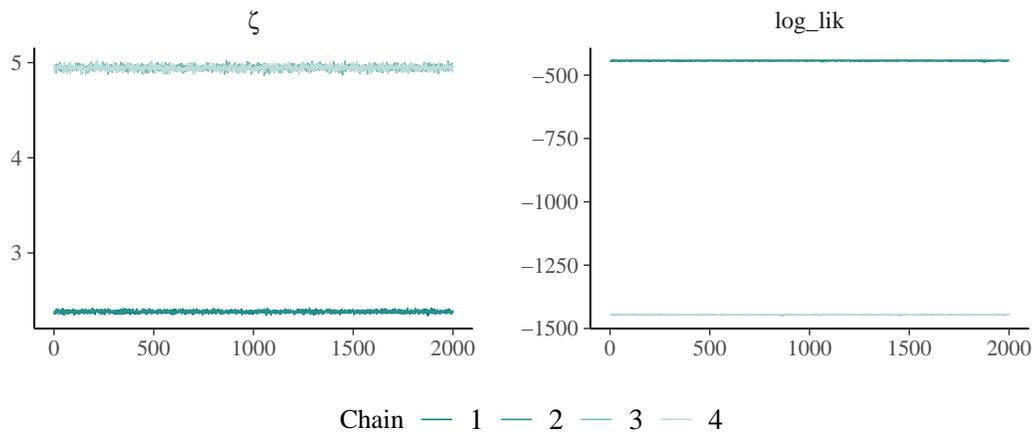


Figure C.35: Bimodality with one unknown parameter.

References

- Abbott, S, J Hellewell, RN Thompson, K Sherratt, HP Gibbs, NI Bosse, JD Munday, et al. 2020. “Estimating the Time-Varying Reproduction Number of SARS-CoV-2 Using National and Subnational Case Counts [Version 2; Peer Review: 1 Approved, 1 Approved with Reservations].” *Wellcome Open Research* 5 (112). <https://doi.org/10.12688/wellcomeopenres.16006.2>.
- Allen, Linda J. S. 2017. “A Primer on Stochastic Epidemic Models: Formulation, Numerical Simulation, and Analysis.” *Infectious Disease Modelling* 2 (2): 128–42. <https://doi.org/10.1016/j.idm.2017.03.001>.
- Anderson, Dorothy, and Ray Watson. 1980. “On the Spread of a Disease with Gamma Distributed Latent and Infectious Periods.” *Biometrika* 67: 191–98. <https://doi.org/10.1093/biomet/67.1.191>.
- Anderson, Roy. 1991. “Discussion: The Kermack-McKendrick Epidemic Threshold Theorem.” *Bulletin of Mathematical Biology* 53 (1): 3–32. [https://doi.org/10.1016/S0092-8240\(05\)80039-4](https://doi.org/10.1016/S0092-8240(05)80039-4).
- Anderson, Roy, and Robert May. 1992. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.
- Andrade, Jair. 2021. *readsdr: Translate Models from System Dynamics Software into 'R'*. <https://CRAN.R-project.org/package=readsdr>.
- Andrade, Jair, and Jim Duggan. 2020. “An Evaluation of Hamiltonian Monte Carlo Performance to Calibrate Age-Structured Compartmental SEIR Models to Incidence Data.” *Epidemics* 33: 100415. <https://doi.org/10.1016/j.epidem.2020.100415>.
- . 2021. “A Bayesian approach to calibrate system dynamics models using Hamiltonian Monte Carlo.” *System Dynamics Review* 37 (4): 283–309. <https://doi.org/10.1002/sdr.1693>.
- . 2022. “Inferring the Effective Reproductive Number from Deterministic and Semi-Deterministic Compartmental Models Using Incidence and Mobility Data.” *PLOS Computational Biology* 18 (6): 1–25.

- <https://doi.org/10.1371/journal.pcbi.1010206>.
- Ansah, John P., Victoria Koh, Chi-Tsun Chiu, Choy-Lye Chei, Yi Zeng, Zhao-Xue Yin, Xiao-Ming Shi, and David B. Matchar. 2017. “Projecting the Number of Elderly with Cognitive Impairment in China Using a Multi-State Dynamic Population Model.” *System Dynamics Review* 33 (2): 89–111. <https://doi.org/10.1002/sdr.1581>.
- Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp. 2002. “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking.” *IEEE Transactions on Signal Processing* 50 (2): 174–88. <https://doi.org/10.1109/78.978374>.
- Bailey, Norman. 1954. “A Statistical Method of Estimating the Periods of Incubation and Infection of an Infectious Disease.” *Nature* 174 (4420): 139–40. <https://doi.org/10.1038/174139a0>.
- Banks, H. Thomas, Marie Davidian, John R. Samuels, and Karyn L. Sutton. 2009. “An Inverse Problem Statistical Methodology Summary.” In *Mathematical and Statistical Estimation Approaches in Epidemiology*, edited by Gerardo Chowell, James M. Hyman, Luis M. A. Bettencourt, and Carlos Castillo-Chavez, 249–302. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-2313-1_11.
- Barlas, Yaman. 1996. “Formal aspects of model validity and validation in system dynamics.” *System Dynamics Review* 12 (3): 183–210. [https://doi.org/10.1002/\(SICI\)1099-1727\(199623\)12:3%3C183::AID-SDR103%3E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-1727(199623)12:3%3C183::AID-SDR103%3E3.0.CO;2-4).
- Barlas, Yaman, and Hakan Yasarcan. 2008. “A Comprehensive Model of Goal Dynamics in Organizations: Setting, Evaluation and Revision.” In *Complex Decision Making: Theory and Practice*, edited by H. Quadrat-Ullah, J. M. Spector, and P. I. Davidsen, 295–320. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73665-3_15.
- Beraha, Mario, Daniele Falco, and Alessandra Guglielmi. 2021. “JAGS, NIMBLE, Stan: A Detailed Comparison Among Bayesian MCMC Software.” <https://arxiv.org/abs/2107.09357>.
- Betancourt, Michael. 2015. “Adiabatic Monte Carlo.” <https://arxiv.org/abs/1405.3489>.
- . 2017. “The Convergence of Markov Chain Monte Carlo Methods: From the Metropolis Method to Hamiltonian Monte Carlo.” <https://arxiv.org/abs/1706.01520>.
- . 2018. “A Conceptual Introduction to Hamiltonian Monte Carlo.” <https://arxiv.org/abs/1701.02434>.

- Betancourt, Michael, Simon Byrne, Samuel Livingstone, and Mark Girolami. 2014. “The Geometric Foundations of Hamiltonian Monte Carlo.” <https://arxiv.org/abs/1410.5110>.
- Bjørnstad, Ottar. 2023. “Stochastics.” In *Epidemics: Models and Data Using R*, edited by Ottar N. Bjørnstad, 165–84. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-12056-5%7B/_%7D9.
- Blitzstein, J. K., and J. Hwang. 2019. *Introduction to Probability, Second Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press.
- Bolker, Benjamin. 2008. *Ecological Models and Data in R*. Princeton University Press.
- Bolker, Benjamin, and Bryan Thomas Grenfell. 1995. “Space, Persistence and Dynamics of Measles Epidemics.” *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 348 (1325): 309–20. <https://doi.org/10.1098/rstb.1995.0070>.
- Box, G. E. P., J. S. Hunter, and W. G. Hunter. 2005. *Statistics for Experimenters: Design, Innovation, and Discovery*. Wiley Series in Probability and Statistics. Wiley.
- Brauer, Fred, and Carlos Castillo-Chavez. 2012. “Models for Endemic Diseases.” In *Mathematical Models in Population Biology and Epidemiology*, 411–64. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-1686-9_10.
- Bretó, Carles. 2018. “Modeling and Inference for Infectious Disease Dynamics: A Likelihood-Based Approach.” *Statistical Science* 33 (1): 57–69. <https://doi.org/10.1214/17-STS636>.
- Bretó, Carles, Daihai He, Edward L. Ionides, and Aaron A. King. 2009. “Time series analysis via mechanistic models.” *The Annals of Applied Statistics* 3 (1): 319–48. <https://doi.org/10.1214/08-AOAS201>.
- Brett, Tobias, Marco Ajelli, Quan-Hui Liu, Mary G. Krauland, John J. Grefenstette, Willem G. van Panhuis, Alessandro Vespignani, John M. Drake, and Pejman Rohani. 2020. “Detecting Critical Slowing down in High-Dimensional Epidemiological Systems.” *PLOS Computational Biology* 16 (3): 1–19. <https://doi.org/10.1371/journal.pcbi.1007679>.
- Brouwer, Andrew F. 2022. “Why the Spectral Radius? An Intuition-Building Introduction to the Basic Reproduction Number.” *Bulletin of Mathematical Biology* 84 (September): 96. <https://doi.org/10.1007/s11538-022-01057-9>.
- Bubar, Kate M., Kyle Reinholt, Stephen M. Kissler, Marc Lipsitch, Sarah Cobey, Yonatan H. Grad, and Daniel B. Larremore. 2021. “Model-Informed COVID-19 Vaccine Prioritization Strategies by Age and Serosta-

- tus.” *Science* 371 (6532): 916–21. <https://doi.org/10.1126/science.abe6959>.
- Camacho, A., A. J. Kucharski, S. Funk, J. Breman, P. Piot, and W. J. Edmunds. 2014. “Potential for Large Outbreaks of Ebola Virus Disease.” *Epidemics* 9: 70–78. <https://doi.org/10.1016/j.epidem.2014.09.003>.
- Camacho, A., Adam Kucharski, Yvonne Aki-Sawyers, Mark A White, Stefan Flasche, Marc Baguelin, Timothy Pollington, et al. 2015. “Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: a Real-time Modelling Study.” *PLoS Currents* 7 (February): ecurrents.outbreaks.406ae55e83ec0b5193e30856b9235ed2. <https://doi.org/10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2>.
- Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software* 76 (1).
- Carpenter, Bob, Matthew D. Hoffman, Marcus Brubaker, Daniel Lee, Peter Li, and Michael Betancourt. 2015. “The Stan Math Library: Reverse-Mode Automatic Differentiation in c++.” <https://arxiv.org/abs/1509.07164>.
- Chatzilena, Anastasia, Edwin van Leeuwen, Oliver Ratmann, Marc Baguelin, and Nikolaos Demiris. 2019. “Contemporary Statistical Inference for Infectious Disease Models Using Stan.” *Epidemics* 29: 100367. <https://doi.org/10.1016/j.epidem.2019.100367>.
- Chopin, N., and O. Papaspiliopoulos. 2020. *An Introduction to Sequential Monte Carlo*. Springer Series in Statistics. Springer International Publishing.
- Chowell, Gerardo, C. E. Ammon, N. W. Hengartner, and J. M. Hyman. 2006. “Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions.” *Journal of Theoretical Biology* 241 (July): 193–204. <https://doi.org/10.1016/j.jtbi.2005.11.026>.
- Chowell, Gerardo, Hiroshi Nishiura, and Luis M. A Bettencourt. 2007. “Comparative Estimation of the Reproduction Number for Pandemic Influenza from Daily Case Notification Data.” *Journal of The Royal Society Interface* 4 (12): 155–66. <https://doi.org/10.1098/rsif.2006.0161>.
- Cintrón-Arias, Ariel, Carlos Castillo-Chávez, Luis M. A. Bettencourt, Alun L. Lloyd, and H. T. Banks. 2020. “The Estimation of the Effective Reproductive Number from Disease Outbreak Data.” <https://arxiv.org/abs/2004.06827>.

- Cori, Anne, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. 2013. “A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics.” *American Journal of Epidemiology* 178 (9): 1505–12. <https://doi.org/10.1093/aje/kwt133>.
- Coulson, Tim, Pejman Rohani, and Mercedes Pascual. 2004. “Skeletons, Noise and Population Growth: The End of an Old Debate?” *Trends in Ecology & Evolution* 19 (7): 359–64. <https://doi.org/10.1016/j.tree.2004.05.008>.
- Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross. 1985. “A Theory of the Term Structure of Interest Rates.” *Econometrica* 53 (2): 385–407. <https://doi.org/10.2307/1911242>.
- Daems, Rudi, Giuseppe Del Giudice, and Rino Rappuoli. 2005. “Anticipating Crisis: Towards a Pandemic Flu Vaccination Strategy Through Alignment of Public Health and Industrial Policy.” *Vaccine* 23 (50): 5732–42. <https://doi.org/10.1016/j.vaccine.2005.10.011>.
- Dangerfield, Brian, and Jim Duggan. 2020. “Optimization of System Dynamics Models.” In *System Dynamics: Theory and Applications*, edited by Brian Dangerfield, 139–52. New York, NY: Springer US. https://doi.org/10.1007/978-1-4939-8790-0_542.
- Davies, Nicholas G., Sam Abbott, Rosanna C. Barnard, Christopher I. Jarvis, Adam J. Kucharski, James D. Munday, Carl A. B. Pearson, et al. 2021. “Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England.” *Science* 372 (6538). <https://doi.org/10.1126/science.abg3055>.
- Davies, Nicholas G., Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, Carl A. B. Pearson, Billy J. Quilty, et al. 2020. “Age-Dependent Effects in the Transmission and Control of COVID-19 Epidemics.” *Nature Medicine* 26 (8): 1205–11. <https://doi.org/10.1038/s41591-020-0962-9>.
- Dehning, Jonas, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. 2020. “Inferring Change Points in the Spread of COVID-19 Reveals the Effectiveness of Interventions.” *Science* 369 (6500). <https://doi.org/10.1126/science.abb9789>.
- Delamater, Paul, Erica Street, Timothy Leslie, Y. Tony Yang, and Kathryn Jacobsen. 2019. “Complexity of the Basic Reproduction Number (R_0).” *Emerging Infectious Diseases* 25 (1): 1. <https://doi.org/10.3201/eid2501.171901>.
- Diaz, Paul, Paul Constantine, Kelsey Kalmbach, Eric Jones, and Stephen Pankavich. 2018. “A Modified SEIR Model for the Spread of Ebola in Western Africa and Metrics for Resource Allocation.” *Applied Mathematics*

- and Computation* 324: 141–55. <https://doi.org/10.1016/j.amc.2017.11.039>.
- Diekmann, O., H. Heesterbeek, and T. Britton. 2013. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press.
- Diekmann, O., J. A. P. Heesterbeek, and J. A. J. Metz. 1990. “On the Definition and the Computation of the Basic Reproduction Ratio R_0 in Models for Infectious Diseases in Heterogeneous Populations.” *Journal of Mathematical Biology* 28 (4): 365–82. <https://doi.org/10.1007/BF00178324>.
- Dietz, K. 1975. “Transmission and Control of Arbovirus Diseases.” In *Epidemiology*, 104–21. Society for Industrial and Applied Mathematics.
- . 1993. “The Estimation of the Basic Reproduction Number for Infectious Diseases.” *Statistical Methods in Medical Research* 2 (1): 23–41. <https://doi.org/10.1177/096228029300200103>.
- Dogan, Gokhan. 2007. “Bootstrapping for Confidence Interval Estimation and Hypothesis Testing for Parameters of System Dynamics Models.” *System Dynamics Review* 23 (4): 415–36. <https://doi.org/10.1002/sdr.362>.
- Douglas, Margaret, Srinivasa Vittal Katikireddi, Martin Taulbut, Martin McKee, and Gerry McCartney. 2020. “Mitigating the Wider Health Effects of Covid-19 Pandemic Response.” *BMJ* 369. <https://doi.org/10.1136/bmj.m1557>.
- Duane, Simon, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth. 1987. “Hybrid Monte Carlo.” *Physics Letters B* 195 (2): 216–22. [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X).
- Dublin, Louis I., and Alfred J. Lotka. 1925. “On the True Rate of Natural Increase.” *Journal of the American Statistical Association* 20 (September): 305. <https://doi.org/10.2307/2965517>.
- Duggan, Jim. 2016. *System Dynamics Modeling with R*. Lecture Notes in Social Networks. Springer International Publishing.
- Duggan, Jim, Jair Andrade, Thomas Brendan Murphy, James P. Gleeson, Cathal Walsh, and Philip Nolan. 2024. “An Age-Cohort Simulation Model for Generating COVID-19 Scenarios: A Study from Ireland’s Pandemic Response.” *European Journal of Operational Research* 313 (1): 343–58. <https://doi.org/10.1016/j.ejor.2023.08.011>.
- Dureau, Joseph, Konstantinos Kalogeropoulos, and Marc Baguelin. 2013. “Capturing the time-varying drivers of an epidemic using stochastic dynamical systems.” *Biostatistics* 14 (3): 541–55. <https://doi.org/10.1093>

- [biostatistics/kxs052](#).
- Efron, B., and R. Tibshirani. 1986. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy." *Statistical Science* 1 (1): 54–75. <https://doi.org/10.1214/ss/1177013815>.
- Endo, Akira, Edwin van Leeuwen, and Marc Baguelin. 2019. "Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers." *Epidemics* 29: 100363. <https://doi.org/10.1016/j.epidem.2019.100363>.
- Epstein, Joshua M., Erez Hatna, and Jennifer Crodelle. 2021. "Triple Contagion: A Two-Fears Epidemic Model." *Journal of The Royal Society Interface* 18 (181): 20210186. <https://doi.org/10.1098/rsif.2021.0186>.
- Farrington, C. P., M. N. Kanaan, and N. J. Gay. 2001. "Estimation of the Basic Reproduction Number for Infectious Diseases from Age-Stratified Serological Survey Data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 50 (3): 251–92. <https://doi.org/10.1111/1467-9876.00233>.
- Fine, Paul E. M. 1993. "Herd Immunity: History, Theory, Practice." *Epidemiologic Reviews* 15 (2): 265–302. <https://doi.org/10.1093/oxfordjournals.epirev.a036121>.
- Fisher, I. 2006. *The Purchasing Power of Money: Its' Determination And Relation to Credit Interest And Crises*. Cosimo Classics Economics. Cosimo Classics.
- Flaxman, Seth, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, et al. 2020. "Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe." *Nature* 584 (7820): 257–61. <https://doi.org/10.1038/s41586-020-2405-7>.
- Forrester, Jay W. 1993. "System Dynamics and the Lessons of 35 Years." In *A Systems-Based Approach to Policymaking*, edited by Kenyon B. De Greene, 199–240. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-3226-2_7.
- Freedman, David, Robert Pisani, and Roger Purves. 2007. *Statistics*. International student ed., 4. ed. New York: Norton.
- Frost, W. H., and Edgar Sydenstricker. 1919. "Influenza in Maryland: Preliminary Statistics of Certain Localities." *Public Health Reports (1896-1970)* 34 (11): 491–504. <http://www.jstor.org/stable/4575056>.
- Funk, Sebastian, Anton Camacho, Adam J. Kucharski, Rosalind M. Eggo, and W. John Edmunds. 2018. "Real-Time Forecasting of Infectious Disease Dynamics with a Stochastic Semi-Mechanistic Model." *Epidemics* 22: 56–61. <https://doi.org/10.1016/j.epidem.2016.11.003>.
- Funk, Sebastian, and Aaron A. King. 2020. "Choices and Trade-Offs

- in Inference with Infectious Disease Models.” *Epidemics* 30: 100383. <https://doi.org/10.1016/j.epidem.2019.100383>.
- Gabry, Jonah, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. 2019. “Visualization in Bayesian Workflow.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182 (2): 389–402. <https://doi.org/10.1111/rssa.12378>.
- Gamado, Kokouvi M., George Strettaris, and Stan Zachary. 2014. “Modelling Under-Reporting in Epidemics.” *Journal of Mathematical Biology* 69 (3): 737–65. <https://doi.org/10.1007/s00285-013-0717-z>.
- Gelfand, Alan E., and Adrian F. M. Smith. 1990. “Sampling-Based Approaches to Calculating Marginal Densities.” *Journal of the American Statistical Association* 85 (410): 398–409. <http://www.jstor.org/stable/289776>.
- Gelman, Andrew. 2023. “What Is a Standard Error?” *Journal of Econometrics* 237 (1): 105516. <https://doi.org/10.1016/j.jeconom.2023.105516>.
- Gelman, Andrew, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- Gelman, Andrew, and J. Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.
- Gelman, Andrew, and Donald B. Rubin. 1992. “Inference from Iterative Simulation Using Multiple Sequences.” *Statistical Science* 7 (4): 457–72. <http://www.jstor.org/stable/2246093>.
- Gelman, Andrew, Daniel Simpson, and Michael Betancourt. 2017. “The Prior Can Often Only Be Understood in the Context of the Likelihood.” *Entropy* 19 (10). <https://doi.org/10.3390/e19100555>.
- Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. 2020. “Bayesian Workflow.” <https://arxiv.org/abs/2011.01808>.
- Geman, S., and D. Geman. 1984. “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6 (6): 721–41. <https://doi.org/10.1109/TPAMI.1984.4767596>.
- Geyer, Charles J. 2011. “Introduction to Markov Chain Monte Carlo.” In *Handbook of Markov Chain Monte Carlo*, 3–48. CRC Press.
- Ghaffarzadegan, Navid, and Hazhir Rahmandad. 2020. “Simulation-Based Estimation of the Early Spread of COVID-19 in Iran: Actual Versus

- Confirmed Cases.” *System Dynamics Review* 36 (1): 101–29. <https://doi.org/10.1002/sdr.1655>.
- Girolami, Mark, and Ben Calderhead. 2011. “Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (2): 123–214. <https://doi.org/https://doi.org/10.1111/j.1467-9868.2010.00765.x>.
- Gleeson, James P., Thomas Brendan Murphy, Joseph D. O’Brien, Nial Friel, Norma Bargary, and David J. P. O’Sullivan. 2022. “Calibrating COVID-19 Susceptible-Exposed-Infected-Removed Models with Time-Varying Effective Contact Rates.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 380 (2214): 20210120. <https://doi.org/10.1098/rsta.2021.0120>.
- Gordon, N. J., A. F. M. Smith, and D. J. Salmond. 1993. “Novel approach to nonlinear/non-Gaussian Bayesian state estimation.” *IEE Proceedings F (Radar and Signal Processing)* 140: 107–113(6). <https://doi.org/10.1049/ip-f-2.1993.0015>.
- Gostic, Katelyn M., Lauren McGough, Edward B. Baskerville, Sam Abbott, Keya Joshi, Christine Tedijanto, Rebecca Kahn, et al. 2020. “Practical Considerations for Measuring the Effective Reproductive Number, R_t .” *PLOS Computational Biology* 16 (12): 1–21. <https://doi.org/10.1371/journal.pcbi.1008409>.
- Graham, AK. 1980. “Parameter Estimation in System Dynamics Modeling.” In *Elements of the System Dynamics Method*, 143–61. Productivity Press.
- Greenhalgh, Scott, and Carly Rozins. 2021. “A Generalized Differential Equation Compartmental Model of Infectious Disease Transmission.” *Infectious Disease Modelling* 6: 1073–91. <https://doi.org/10.1016/j.idm.2021.08.007>.
- Grinsztajn, Léo, Elizaveta Semenova, Charles C. Margossian, and Julien Riou. 2021. “Bayesian Workflow for Disease Transmission Modeling in Stan.” *Statistics in Medicine* 40 (27): 6209–34. <https://doi.org/10.1002/sim.9164>.
- Hastie, Trevor, and Robert Tibshirani. 1993. “Varying-Coefficient Models.” *Journal of the Royal Statistical Society. Series B (Methodological)* 55 (4): 757–96. <http://www.jstor.org/stable/2345993>.
- Hastings, W. K. 1970. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications.” *Biometrika* 57 (1): 97–109. <http://www.jstor.org/stable/2334940>.
- He, Daihai, Edward L. Ionides, and Aaron A. King. 2010. “Plug-and-play inference for disease dynamics: measles in large and small populations as a case study.” *Journal of The Royal Society Interface* 7 (43): 271–83.

- <https://doi.org/10.1098/rsif.2009.0151>.
- Heesterbeek, J. A. P. 2002. “A Brief History of R_0 and a Recipe for Its Calculation.” *Acta Biotheoretica* 50 (3): 189–204. <https://doi.org/10.1023/A:1016599411804>.
- Heesterbeek, J. A. P., and K. Dietz. 1996. “The Concept of R_0 in Epidemic Theory.” *Statistica Neerlandica* 50 (1): 89–110. <https://doi.org/10.1111/j.1467-9574.1996.tb01482.x>.
- Heffernan, J. M., R. J. Smith, and L. M. Wahl. 2005. “Perspectives on the Basic Reproductive Ratio.” *Journal of The Royal Society Interface* 2 (September): 281–93. <https://doi.org/10.1098/rsif.2005.0042>.
- Hirotsu, Nobuo, Hideyuki Ikematsu, Norio Iwaki, Naoki Kawai, Takeshi Shigematsu, Osamu Kunishima, and Seizaburo Kashiwagi. 2004. “Effects of Antiviral Drugs on Viral Detection in Influenza Patients and on the Sequential Infection to Their Family Members—Serial Examination by Rapid Diagnosis (Capilia) and Virus Culture.” *International Congress Series* 1263: 105–8. <https://doi.org/10.1016/j.ics.2004.02.020>.
- Hoffman, Matthew D., and Andrew Gelman. 2011. “The No-u-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo.” <https://arxiv.org/abs/1111.4246>.
- Hollingsworth, T. Déirdre, Don Klinkenberg, Hans Heesterbeek, and Roy Anderson. 2011. “Mitigation Strategies for Pandemic Influenza a: Balancing Conflicting Policy Objectives.” *PLOS Computational Biology* 7 (2): 1–11. <https://doi.org/10.1371/journal.pcbi.1001076>.
- HPSC. 2020. “Preliminary report of the results of the Study to Investigate COVID-19 Infection in People Living in Ireland (SCOPI): A national seroprevalence study, June-July 2020.” 2020. <https://www.hpsc.ie/a-z/respiratory/coronavirus/novelcoronavirus/scopi/>.
- Huang, Chaolin, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, et al. 2020. “Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China.” *The Lancet* 395 (10223): 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- Hurtado, Paul J., and Adam S. Kiro Singh. 2019. “Generalizations of the ‘Linear Chain Trick’: Incorporating More Flexible Dwell Time Distributions into Mean Field ODE Models.” *Journal of Mathematical Biology* 79 (5): 1831–83. <https://doi.org/10.1007/s00285-019-01412-w>.
- Hyndman, Rob J., and Anne B. Koehler. 2006. “Another Look at Measures of Forecast Accuracy.” *International Journal of Forecasting* 22 (4): 679–88. <https://doi.org/10.1016/j.ijforecast.2006.03.001>.
- Ionides, Edward L., Anindya Bhadra, Yves Atchadé, and Aaron King. 2011.

- “Iterated filtering.” *The Annals of Statistics* 39 (3): 1776–1802. <https://doi.org/10.1214/11-AOS886>.
- Ionides, Edward L., C. Breto, J. Park, R. A. Smith, and A. A. King. 2017. “Monte Carlo profile confidence intervals for dynamic systems.” *Journal of The Royal Society Interface* 14 (132): 20170126. <https://doi.org/10.1098/rsif.2017.0126>.
- Ionides, Edward L., C. Bretó, and A. A. King. 2006. “Inference for Nonlinear Dynamical Systems.” *Proceedings of the National Academy of Sciences* 103 (49): 18438–43. <https://doi.org/10.1073/pnas.0603181103>.
- Jacquez, J. A. 1972. *Compartmental Analysis in Biology and Medicine: Kinetics of Distribution of Tracer-Labeled Materials*. Elsevier Publishing Company.
- James, G., D. Witten, T. Hastie, and R. Tibshirani. 2021. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer US.
- Katul, Gabriel G., Assaad Mrad, Sara Bonetti, Gabriele Manoli, and Anthony J. Parolari. 2020. “Global Convergence of COVID-19 Basic Reproduction Number and Estimation from Early-Time SIR Dynamics.” *PLOS ONE* 15 (9): 1–22. <https://doi.org/10.1371/journal.pone.0239800>.
- Keeling, Matt J., and Bryan T. Grenfell. 2002. “Understanding the Persistence of Measles: Reconciling Theory, Simulation and Observation.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269 (1489): 335–43. <https://doi.org/10.1098/rspb.2001.1898>.
- Keeling, Matt J., and Pejman Rohani. 2002. “Estimating Spatial Coupling in Epidemiological Systems: A Mechanistic Approach.” *Ecology Letters* 5 (1): 20–29. <https://doi.org/10.1046/j.1461-0248.2002.00268.x>.
- . 2011. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- Keeling, Matt J., Pejman Rohani, and Bryan T. Grenfell. 2001. “Seasonally Forced Disease Dynamics Explored as Switching Between Attractors.” *Physica D: Nonlinear Phenomena* 148 (3): 317–35. [https://doi.org/10.1016/S0167-2789\(00\)00187-1](https://doi.org/10.1016/S0167-2789(00)00187-1).
- Keith, David R., John D. Sterman, and Jeroen Struben. 2017. “Supply Constraints and Waitlists in New Product Diffusion.” *System Dynamics Review* 33 (3-4): 254–79. <https://doi.org/10.1002/sdr.1588>.
- Kermack, William O., and A. G. McKendrick. 1927. “A Contribution to the Mathematical Theory of Epidemics.” *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115 (August): 700–721. <https://doi.org/10.1098/rspa.1927.01>

18.

- Kimeldorf, George S., and Grace Wahba. 1970. "A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines." *The Annals of Mathematical Statistics* 41 (2): 495–502. <https://doi.org/10.1214/aoms/1177697089>.
- King, Aaron A., Matthieu Domenech de Cellès, Felicia M. G. Magpantay, and Pejman Rohani. 2015. "Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola." *Proceedings of the Royal Society B: Biological Sciences* 282 (1806): 20150347. <https://doi.org/10.1098/rspb.2015.0347>.
- King, Aaron A., Dao Nguyen, and Edward L. Ionides. 2016. "Statistical Inference for Partially Observed Markov Processes via the R Package pomp." *Journal of Statistical Software, Articles* 69 (12): 1–43. <https://doi.org/10.18637/jss.v069.i12>.
- Krylova, Olga, and David J. D. Earn. 2013. "Effects of the infectious period distribution on predicted transitions in childhood disease dynamics." *Journal of The Royal Society Interface* 10 (July): 20130098. <https://doi.org/10.1098/rsif.2013.0098>.
- Kucharski, Adam. 2020. *The Rules of Contagion: Why Things Spread - and Why They Stop*. Profile Books.
- Kuczynski, R. R. 1928. *The Balance of Births and Deaths. Volume 1: Western and Northern Europe*. New York: Macmillan.
- Lambert, Ben. 2018. *A Student's Guide to Bayesian Statistics*. Los Angeles: SAGE.
- Li, Qun, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, et al. 2020. "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia." *New England Journal of Medicine* 382 (13): 1199–1207. <https://doi.org/10.1056/NEJMoa2001316>.
- Linka, Kevin, Mathias Peirlinck, and Ellen Kuhl. 2020. "The Reproduction Number of COVID-19 and Its Correlation with Public Health Interventions." *Computational Mechanics* 66 (4): 1035–50. <https://doi.org/10.1007/s00466-020-01880-8>.
- Liu, Lujun, Xiao-Qiang Zhao, and Yicang Zhou. 2010. "A Tuberculosis Model with Seasonality." *Bulletin of Mathematical Biology* 72 (4): 931–52. <https://doi.org/10.1007/s11538-009-9477-8>.
- Liu, Xinzhi, and Peter Stechlinski. 2012. "Infectious Disease Models with Time-Varying Parameters and General Nonlinear Incidence Rate." *Applied Mathematical Modelling* 36 (5): 1974–94. <https://doi.org/10.1016/j.apm.2011.08.019>.

- Lloyd, Alun L. 2001a. “Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics.” *Theoretical Population Biology* 60 (1): 59–71. <https://doi.org/10.1006/tpbi.2001.1525>.
- . 2001b. “The Dependence of Viral Parameter Estimates on the Assumed Viral Life Cycle: Limitations of Studies of Viral Load Data.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268 (1469): 847–54. <https://doi.org/10.1098/rspb.2000.1572>.
- . 2009. “Sensitivity of Model-Based Epidemiological Parameter Estimation to Model Assumptions.” In *Mathematical and Statistical Estimation Approaches in Epidemiology*, edited by Gerardo Chowell, James M. Hyman, Luis M. A. Bettencourt, and Carlos Castillo-Chavez, 123–41. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-2313-1_6.
- Ma, Junling, and David J. D. Earn. 2006. “Generality of the Final Size Formula for an Epidemic of a Newly Invading Infectious Disease.” *Bulletin of Mathematical Biology* 68 (3): 679–702. <https://doi.org/10.1007/s11538-005-9047-7>.
- Macdonald, George. 1952. “The Analysis of Equilibrium in Malaria.” *Tropical Diseases Bulletin* 49 9: 813–29.
- McElreath, R. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. A Chapman & Hall Book. CRC Press LLC.
- Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. “Equation of State Calculations by Fast Computing Machines.” *The Journal of Chemical Physics* 21 (6): 1087–92. <https://doi.org/10.1063/1.1699114>.
- Mills, Christina E., James M. Robins, and Marc Lipsitch. 2004. “Transmissibility of 1918 Pandemic Influenza.” *Nature* 432 (7019): 904–6. <https://doi.org/10.1038/nature03063>.
- Mollison, D. 1995. “The Structure of Epidemic Models.” In *Epidemic Models: Their Structure and Relation to Data*, 17–33. Cambridge University Press.
- Monnahan, Cole C., James T. Thorson, and Trevor A. Branch. 2017. “Faster Estimation of Bayesian Models in Ecology Using Hamiltonian Monte Carlo.” *Methods in Ecology and Evolution* 8 (3): 339–48. <https://doi.org/10.1111/2041-210X.12681>.
- Monto, Arnold S., and Keiji Fukuda. 2019. “Lessons From Influenza Pandemics of the Last 100 Years.” *Clinical Infectious Diseases* 70 (5): 951–57. <https://doi.org/10.1093/cid/ciz803>.
- Monto, Arnold S., and Robert G. Webster. 2013. “Influenza Pandemics:

- History and Lessons Learned.” In *Textbook of Influenza*, 20–34. <https://doi.org/10.1002/9781118636817.ch2>.
- Morens, David M., Gregory K. Folkers, and Anthony S. Fauci. 2009. “What Is a Pandemic?” *The Journal of Infectious Diseases* 200 (7): 1018–21. <http://www.jstor.org/stable/27794175>.
- Mossong, Joël, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, et al. 2008. “Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases.” *PLOS Medicine* 5 (3): 1–1. <https://doi.org/10.1371/journal.pmed.0050074>.
- Msemburi, William, Ariel Karlinsky, Victoria Knutson, Serge Aleshin-Guendel, Somnath Chatterji, and Jon Wakefield. 2023. “The WHO Estimates of Excess Mortality Associated with the COVID-19 Pandemic.” *Nature* 613 (7942): 130–37. <https://doi.org/10.1038/s41586-022-05522-2>.
- Neal, Radford. 1995. “Suppressing Random Walks in Markov Chain Monte Carlo Using Ordered Overrelaxation.” <https://arxiv.org/abs/bayesian/9506004>.
- . 2011. “MCMC Using Hamiltonian Dynamics.” In *Handbook of Markov Chain Monte Carlo*, 116–62. CRC Press.
- Nelder, J. A., and R. Mead. 1965. “A Simplex Method for Function Minimization.” *The Computer Journal* 7 (4): 308–13. <https://doi.org/10.1093/comjnl/7.4.308>.
- Nishiura, Hiroshi, and Gerardo Chowell. 2009. “The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends.” In *Mathematical and Statistical Estimation Approaches in Epidemiology*, edited by Gerardo Chowell, James M. Hyman, Luis M. A. Bettencourt, and Carlos Castillo-Chavez, 103–21. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-2313-1_5.
- Ogi-Gittins, I, WS Hart, J Song, RK Nash, J Polonsky, A Cori, EM Hill, and RN Thompson. 2023. “A Simulation-Based Approach for Estimating the Time-Dependent Reproduction Number from Temporally Aggregated Disease Incidence Time Series Data.” *medRxiv*. <https://doi.org/10.1101/2023.09.13.23295471>.
- Oliva, Rogelio. 2003. “Model Calibration as a Testing Strategy for System Dynamics Models.” *European Journal of Operational Research* 151 (3): 552–68. [https://doi.org/10.1016/S0377-2217\(02\)00622-7](https://doi.org/10.1016/S0377-2217(02)00622-7).
- Oliva, Rogelio, and John D. Serman. 2001. “Cutting Corners and Working Overtime: Quality Erosion in the Service Industry.” *Management Science* 47 (7): 894–914. <https://doi.org/10.1287/mnsc.47.7.894.9807>.
- Osgood, Nathaniel, and Juxin Liu. 2015. “Combining Markov Chain Monte

- Carlo Approaches and Dynamic Modeling.” In *Analytical Methods for Dynamic Modelers*, 125–70. MIT Press.
- Park, Sang Woo, and Benjamin M. Bolker. 2020. “A Note on Observation Processes in Epidemic Models.” *Bulletin of Mathematical Biology* 82 (3): 37. <https://doi.org/10.1007/s11538-020-00713-2>.
- Patterson, KD, and GF Pyle. 1991. “The Geography and Mortality of the 1918 Influenza Pandemic.” *Bulletin of the History of Medicine* 65 (1): 4–21. <http://europepmc.org/abstract/MED/2021692>.
- Pawitan, Y. 2013. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. OUP Oxford.
- Petersen, Eskild, Marion Koopmans, Unyeong Go, Davidson H Hamer, Nicola Petrosillo, Francesco Castelli, Merete Storgaard, Sulien Al Khalili, and Lone Simonsen. 2020. “Comparing SARS-CoV-2 with SARS-CoV and Influenza Pandemics.” *The Lancet Infectious Diseases* 20 (9): e238–44. [https://doi.org/10.1016/S1473-3099\(20\)30484-9](https://doi.org/10.1016/S1473-3099(20)30484-9).
- Pierson, Kawika, and John D. Sterman. 2013. “Cyclical Dynamics of Airline Industry Earnings.” *System Dynamics Review* 29 (3): 129–56. <https://doi.org/10.1002/sdr.1501>.
- Powell, M. J. D. 1964. “An efficient method for finding the minimum of a function of several variables without calculating derivatives.” *The Computer Journal* 7 (2): 155–62. <https://doi.org/10.1093/comjnl/7.2.155>.
- Priestley, M. B. 1980. “STATE-DEPENDENT MODELS: A GENERAL APPROACH TO NON-LINEAR TIME SERIES ANALYSIS.” *Journal of Time Series Analysis* 1 (1): 47–71. <https://doi.org/10.1111/j.1467-9892.1980.tb00300.x>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rasmussen, David A., Oliver Ratmann, and Katia Koelle. 2011. “Inference for Nonlinear Epidemiological Models Using Genealogies and Time Series.” *PLOS Computational Biology* 7 (8): 1–11. <https://doi.org/10.1371/journal.pcbi.1002136>.
- Renshaw, E. 1993. *Modelling Biological Populations in Space and Time*. Cambridge Studies in Mathematical Biology. Cambridge University Press.
- Ritchie, Hannah, Esteban Ortiz-Ospina, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, Cameron Appel, Lucas Rodés-Guirao, and Max Roser. 2020. “Coronavirus Pandemic (COVID-19).” *Our World in Data*.

- Robert, C. 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Texts in Statistics. Springer New York.
- Robert, C., and G. Casella. 2005. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer New York.
- . 2010. *Introducing Monte Carlo Methods with R*. Use r! Springer.
- . 2011. “A Short History of MCMC: Subjective Recollections from Incomplete Data.” In *Handbook of Markov Chain Monte Carlo*, 49–66. CRC Press.
- Rohani, Pejman, Matthew J. Keeling, and Bryan T. Grenfell. 2002. “The Interplay Between Determinism and Stochasticity in Childhood Diseases.” *The American Naturalist* 159 (5): 469–81. <https://doi.org/10.1086/339467>.
- Ross, Ronald. 1910. *The Prevention of Malaria*. J. Murray.
- Ross, Ronald, and Hilda P. Hudson. 1917. “An Application of the Theory of Probabilities to the Study of a Priori Pathometry.—Part III.” *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 93 (650): 225–40. <https://doi.org/10.1098/rspa.1917.0015>.
- Royston, Patrick. 2007. “Profile Likelihood for Estimation and Confidence Intervals.” *The Stata Journal* 7 (3): 376–87. <https://doi.org/10.1177/1536867X0700700305>.
- Saleh, Mohamed, Rogelio Oliva, Christian Erik Kampmann, and Pål I. Davidsen. 2010. “A Comprehensive Analytical Approach for Policy Analysis of System Dynamics Models.” *European Journal of Operational Research* 203 (3): 673–83. <https://doi.org/10.1016/j.ejor.2009.09.016>.
- Sanche, Steven, Yen Ting Lin, Chonggang Xu, Ethan Romero-Severson, Nick Hengartner, and Ruian Ke. 2020. “High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2” 26 (7): 1470. <https://doi.org/10.3201/eid2607.200282>.
- Sartwell, Philip. 1995. “The distribution of incubation periods of infectious disease.” *American Journal of Epidemiology* 141 (5): 386–94. <https://doi.org/10.1093/oxfordjournals.aje.a117440>.
- Sklar, A. 1996. “Random Variables, Distribution Functions, and Copulas: A Personal Look Backward and Forward.” *Lecture Notes-Monograph Series* 28: 1–14. <http://www.jstor.org/stable/4355880>.
- Sterman, John D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill Higher Education. Irwin/McGraw-Hill.

- . 2018. “System Dynamics at Sixty: The Path Forward.” *System Dynamics Review* 34 (1-2): 5–47. <https://doi.org/10.1002/sdr.1601>.
- Strogatz, Steven. 2018. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Second edition. A Chapman & Hall Book. Boca Raton London New York: CRC Press Taylor & Francis Group.
- Struben, Jeroen, John Sterman, and David Keith. 2015. “Parameter Estimation Through Maximum Likelihood and Bootstrapping Methods.” In *Analytical Methods for Dynamic Modelers*, 3–38. MIT Press.
- Sturniolo, Simone, William Waites, Tim Colbourn, David Manheim, and Jasmina Panovska-Griffiths. 2021. “Testing, Tracing and Isolation in Compartmental Models.” *PLOS Computational Biology* 17 (3): 1–28. <https://doi.org/10.1371/journal.pcbi.1008633>.
- Svensson, Åke. 2007. “A Note on Generation Times in Epidemic Models.” *Mathematical Biosciences* 208 (1): 300–311. <https://doi.org/10.1016/j.mbs.2006.10.010>.
- Talts, Sean, Michael Betancourt, Daniel Simpson, Aki Vehtari, and Andrew Gelman. 2018. “Validating Bayesian Inference Algorithms with Simulation-Based Calibration.” arXiv. <https://doi.org/10.48550/ARXIV.1804.06788>.
- Thompson, R. N., J. E. Stockwin, R. D. van Gaalen, J. A. Polonsky, Z. N. Kamvar, P. A. Demarsh, E. Dahlqwert, et al. 2019. “Improved Inference of Time-Varying Reproduction Numbers During Infectious Disease Outbreaks.” *Epidemics* 29: 100356. <https://doi.org/10.1016/j.epidem.2019.100356>.
- Valderrama-Bahamóndez, Gloria I., and Holger Fröhlich. 2019. “MCMC Techniques for Parameter Estimation of ODE Based Models in Systems Biology.” *Frontiers in Applied Mathematics and Statistics* 5: 55. <https://doi.org/10.3389/fams.2019.00055>.
- van den Driessche, Pauline. 2017. “Reproduction Numbers of Infectious Disease Models.” *Infectious Disease Modelling* 2 (3): 288–303. <https://doi.org/10.1016/j.idm.2017.06.002>.
- Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2021. “Rank-Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of MCMC.” *Bayesian Analysis*, 1–28. <https://doi.org/10.1214/20-BA1221>.
- Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon. 2009. “Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized

- Subspace Sampling.” *International Journal of Nonlinear Sciences and Numerical Simulation* 10 (3): 273–90. <https://doi.org/doi:10.1515/IJNS.NS.2009.10.3.273>.
- Vynnycky, E., and W. J. Edmunds. 2008. “Analyses of the 1957 (Asian) influenza pandemic in the United Kingdom and the impact of school closures.” *Epidemiology and Infection* 136 (2): 166–79. <https://doi.org/10.1017/S0950268807008369>.
- Vynnycky, E., Amy Trindall, and Punam Mangtani. 2007. “Estimates of the reproduction numbers of Spanish influenza using morbidity data.” *International Journal of Epidemiology* 36 (4): 881–89. <https://doi.org/10.1093/ije/dym071>.
- Vynnycky, E., and R. White. 2010. *An Introduction to Infectious Disease Modelling*. Oxford University Press.
- Wale, Nina, Matthew J. Jones, Derek G. Sim, Andrew F. Read, and Aaron A. King. 2019. “The contribution of host cell-directed vs. parasite-directed immunity to the disease and dynamics of malaria infections.” *Proceedings of the National Academy of Sciences* 116 (44): 22386–92. <https://doi.org/10.1073/pnas.1908147116>.
- Wallinga, Jacco, and Marc Lipsitch. 2007. “How Generation Intervals Shape the Relationship Between Growth Rates and Reproductive Numbers.” *Proceedings of the Royal Society B: Biological Sciences* 274 (February): 599–604. <https://doi.org/10.1098/rspb.2006.3754>.
- Wallinga, Jacco, and Peter Teunis. 2004. “Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures.” *American Journal of Epidemiology* 160 (6): 509–16. <https://doi.org/10.1093/aje/kwh255>.
- Wearing, Helen J, Pejman Rohani, and Matt J Keeling. 2005. “Appropriate Models for the Management of Infectious Diseases.” *PLoS Medicine* 2 (July): e174. <https://doi.org/10.1371/journal.pmed.0020174>.
- WHO. 2017. “Pandemic Influenza Risk Management: A WHO Guide to Inform and Harmonize National and International Pandemic Preparedness and Response.” Technical documents. World Health Organization.
- Wiersema, U. F. 2008. *Brownian Motion Calculus*. Wiley Finance. Wiley.
- World Health Organization. 2019. *Non-Pharmaceutical Public Health Measures for Mitigating the Risk and Impact of Epidemic and Pandemic Influenza*. Geneva: World Health Organization. <https://apps.who.int/iris/handle/10665/329438>.