A Modelling Framework for Estimating the Risk of Importation of a Novel Disease

Antonio Parrella

September 29, 2023

Thesis submitted for the degree of
Master of Philosophy
in
Applied Mathematics
at The University of Adelaide
Faculty of Engineering, Computer and Mathematical Sciences
School of Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: Date: September 29, 2023

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors Andrew Black and Jack Maclean. Both have gone well beyond their duties, providing invaluable guidance for both this project and my professional ambitions. I greatly appreciate the work these two have put in to allow me to complete my MPhil.

I would also like to thank the friends and family that have supported me over the past 2 years. In particular, I must mention my brilliant partner and best friend Tara Ardalich, who provided unfaltering support even while completing her own honours degree. Finally, I would like to thank Michael Fairbrother, William Allan, Joshua Watt and Tyson Klingner, people who have provided me with companionship and insight throughout all of my university studies.

Abstract

Sequential Monte Carlo (SMC) methods are vital in fitting models, without a tractable likelihood, to data. When combined with Markov Chain Monte Carlo, SMC allows for full posterior distributions of states and parameters to be estimated. However, for many problems, these methods can be prohibitively computationally expensive. One such class of models with intractable likelihoods are continuous-time Branching Processes (CTBPs). In this thesis, we leverage the unique properties of CTBPs to derive a method that approximates the results of standard SMC methods, with a significant reduction in computation time. We find that under certain conditions the method we have developed can produce highly accurate results in orders of magnitude less time than standard SMC methods.

Continuous-time Branching Processes are often used for epidemic modelling, particularly in the early phases of an outbreak. In light of the COVID-19 pandemic, CTBPs have been used in metapopulation models, where agents are partitioned into subpopulations (usually states or countries) that interact through immigration. In this thesis, we build upon existing work in this area, with a focus on estimating disease importation risk. We show how applying our method to this problem can allow for joint estimation of the parameters mediating disease spread and unobserved cases. Specifically, the speed improvement given by our method allows for full posterior distributions for states, parameters and importation risk to be derived. Furthermore, we find that the increase in speed also allows more parameters to be estimated. Consequently, each subpopulation can have its own parameters. As a result, hierarchical modelling can be employed, meaning that parameter estimates from one subpopulation can inform the estimates of others. We find hierarchical modelling to be vital in estimating importation risk, particularly for counties with low observation probability.

Chapter 1

Introduction

Throughout history, infectious diseases have remained an ever-present danger to human society [1]. Even in the modern age, infectious disease is still a leading cause of death in both high and low-income countries, responsible for $\sim 20\%$ of deaths worldwide in 2019 [2]. A vital aspect of minimising the morbidity and mortality associated with a disease is intelligent public health policy [3, 4]. However, implementing the best interventions requires us to understand the current state of an outbreak and predict future states based on available data. One approach to answering questions of this type is the field of mathematical epidemiology [5, 6]. The importance of mathematical epidemiology was made abundantly clear through the recent SARS-CoV-2 pandemic [3, 4].

On December 29 2019 in Wuhan, China, a small cluster of individuals were reported to be exhibiting atypical pneumonia. Soon after, these symptoms were determined to be caused by a novel coronavirus, named SARS-CoV-2 [7]. Shortly after this, many governments implemented various travel restrictions in an attempt to reduce the rate of importation of infected individuals. On February 1, 2020, the Australian Federal Government placed extensive restrictions on travel from Mainland China to Australia [4]. At this time, Australian authorities had detected 9 cases, all with links to people living in Wuhan [8]. Additionally, there were also a confirmed 120 cases outside of China [7]. Thus, an important question at the time was the probability of importation from countries in the wider Asia-Pacific where travel was still permitted [9, 10, 11]. The key difficulty for this endeavour was that the true number of infected people was not observed, only the *reported* cases. This was problematic as reported cases will almost always underestimate the true number of cases, meaning risks will be underestimated [12, 13, 14, 15]. Consequently, one focus of early modelling work was determining which countries were likely to have high rates of underreporting [10, 11]. De Salazar et al. [10] used data on air travel volume from Wuhan, China to estimate each country's relative detection capabilities. From this work, it was found that Singapore had very high relative detection capabilities. Bhatia et al. [11] implemented a similar model but assumed Singapore had *perfect* detection ca-

pabilities, leveraging this allowed the authors to produce estimates of the total number of undetected cases in other countries. While these works were useful in analysing the then-current risks associated with air travel, they did not explore how these risks changed with time. This problem tends to be more difficult as it often requires a mechanistic model that can be used to simulate disease dynamics. Shearer et al. [9] choose to use a simple branching process model to simulate outbreaks, constructing a modular framework that used the estimates of unobserved cases from Bhatia et al. [11] and De Salazar et al. [10]. These simulated outbreaks were then combined with air-travel volume data to estimate the risk of case importation into Australia. However, due to the severe time constraints, this assessment relied on early parameter estimates from other studies [16, 17] and fitted data to simulated outbreaks independently using an approximate approach. This motivates the primary focus of this work: a flexible (meaning applicable to a broad class of models) framework for early-phase disease modelling that jointly estimates parameters, unobserved cases, and importation risk. The key advantage of the joint estimation approach is that the same model is used for parameter estimation, forecasting, and risk assessment ensuring perfect coherence between these elements. Furthermore, this approach allows for new estimates to be produced every time new data is made available instead of having to wait for other papers to be published. During the early phases of a novel disease outbreak, data is scarce meaning it is vital that all of the available data is being made use of, highlighting the importance of this type of approach.

Chinazzi et al. [4] developed a global model of SARS-CoV-2, jointly estimating the R_0 parameter and the total number of infections. Due to computational demands, the authors were forced to rely on approximate Bayesian computation and were limited to the estimation of a single parameter. In this work, we aim to remedy both of these issues. It is important to be able to estimate many parameters simultaneously, since in general, we expect different countries to have differing R_0 values due to factors like population density and public health infrastructure [18, 19, 20]. Additionally, country-specific R_0 values allow for hierarchical modelling to be employed. Hierarchical modelling more accurately encapsulates our knowledge since we expect R_0 values to be different but still correlated since the same underlying disease is the cause [21]. In Section 4.3 we show how the formation of a hierarchical model can significantly improve parameter estimates if done correctly. Moreover, while approximate Bayesian computation can have benefits in the context of model mispecfication [22], it can lead to significant bias in posterior distributions and loss of information [23]. The key barrier in jointly estimating all these parameters is computation time. For the framework we have developed to be maximally useful, it needs to be able to produce estimates within a day, as to keep up with the daily influx of data. In Section 4.3 we show that our method can achieve this goal for a large metapopulation model.

A key factor in grappling with these computational efficiency issues is the choice of the

underlying epidemic model. In particular, the model needs to be a reasonable approximation of reality but simple enough to perform state and parameter estimation efficiently. For modelling the early phases of the COVID-19 outbreaks, Crump-Mode-Jagers processes have received significant attention due to their ability to model complex phenomena [24, 25, 26]. These models are particularly useful for assessing the efficacy of certain local policies such as contact tracing and quarantining. However, due to the complexity of these models, likelihood estimation can be computationally infeasible, meaning authors often need to derive parameter estimates using cruder methods [25, 26]. Conversely, simpler CTMC models, for which likelihood estimates can be feasibly obtained through particle filters, have also been explored [27, 28]. However, these methods tend to rely upon particle Markov chain Monte Carlo methods which incur a significant computational demand [29]. Alternatively, Romano et al. [30] and Carcione et al. [31] have implemented deterministic epidemic models to estimate parameters and predict outbreak trajectories. Deterministic models require far less computation time. However, the stochasticity of the transmission dynamics has a significant effect on infection counts when case counts are low. Since we are focused on the early phases of an outbreak, where case counts are low, we must capture this stochasticity. Given this we follow Shearer et al. [9], making use of continuous-time branching process approximations of compartmental epidemic models. These models capture the stochasticity of early transmission dynamics but also have useful properties which can be used to improve computational efficiency. Leveraging these properties to conduct joint state and parameter estimation is a key contribution made by this work.

Branching process models have many use cases across a variety of disciplines including epidemiology [32, 33], ecology [34], biology [35, 36] and economics [37]. Our focus in Chapter 3 will be on the specific problem of inferring states and parameters of continuous-time branching processes (CTBPs) based on incomplete data. Specifically, we consider data generated by noisy and partial observation of the state at discrete points in a time series, typically thought of as a hidden Markov model [38]. The problem of joint state and parameter can be solved by combining the Metropolis-Hastings algorithm with a bootstrap particle filter (PF), called pseudo-marginal Metropolis-Hastings (pmMH) [39]. The PF essentially numerically integrates over all the unobserved sample paths by simulating from the model directly [29]. Given this, using the PF as part of a pmMH scheme for a CTBP can become computationally expensive when the number of agents gets large [40]. Xu and Minin [36] and Xu et al. [41] have developed approximate methods that improve computational efficiency. However, these improvements are far from sufficient to estimate states and parameters for complex single-population epidemic models within a reasonable time frame. Moreover, while in Chapter 3 we focus on single population models in Chapter 4 we extend this method to multi-population models, which incur further computational demands. This brings us to a primary contribution of this the-

sis: a hybrid algorithm derived in Section 3.2.4 that can approximate the results of an expensive particle filter, but which runs significantly faster. This allows for the estimation of full Bayesian posteriors for the states and parameters for complex epidemic models.

The key observation that makes this algorithm work is the tendency for the state posterior distributions to become approximately Gaussian over time as the agent counts get larger, established in Sections 3.2.2 and 3.3. Under these conditions, we find that a Gaussian approximation method, similar to a Kalman filter, can be used to obtain a reasonable approximation for the PF. Moreover, we show that this observation can be effectively taken advantage of by switching between a Gaussian approximation and a bootstrap particle filter depending on the agent count (Section 3.2.4). We find that for time steps with high agent counts (when the exact simulation of CTBP is very slow) the Gaussian approximation is the most accurate. Complementing this, when the agent count is low (when the exact simulation of CTBP is very fast) the Gaussian approximation is the least accurate. Hence, switching between filtering methods allows our method to run quickly while maintaining as much accuracy as required (Section 3.2.4).

In Chapter 4 we build on the Gaussian approximation method developed in Chapter 3 by applying it to an epidemic model with a meta-population structure (see Section 4.1.3). In particular, we consider a model where sub-populations interact through immigration. This requires an extra layer of approximation which we analyse via a simulation study. This simulation study indicates that accurate results can be obtained in situations relevant to early-phase epidemic modelling. Additionally, through a series of further simulation studies, we show that our Gaussian approximation method can be used to produce full posterior distributions for the states and parameters for meta-population models with around 20 subpopulations, a significant departure from what is feasible with a standard particle filter (see Section 4.3). Next, we explore the effect of hierarchical modelling, finding that it can be used to improve parameter estimates (see Section 4.3.1). Putting everything together, we describe how the method we have developed can be adapted into a framework for estimating the probability of the importation of disease from one sub-population to another (see Section 4.3.3). Further, we show that for large models (~ 50 parameters) joint parameter estimates can be obtained within a day of computation on a standard laptop processor, using 3 cores.

In summary, this thesis has 2 main contributions. Firstly, an adaptive sequential Bayesian filtering method for CTBPs that can approximate the output of the more computationally expensive bootstrap particle filter. This method can be used for efficient joint state and parameter estimation for any multi-type CTBP. Second, we adapt this method to a hierarchical metapopulation epidemic model. This adapted method can produce full posterior distributions for states and parameters, allowing for detailed quantification of the risks associated with travel from different countries.

Chapter 2

Technical Background

2.1 Continuous Time Markov Chains

In this section, we introduce continuous time Markov chains following the treatment in Durrett [42]. A continuous time Markov Chain (CTMC) is a stochastic process $\{\mathbf{z}_t\}_{t\geq 0}$ that satisfies the Markov property. The Markov property or *memoryless property* postulates that the state of the process at some time is only dependent on the most recent knowledge of the state prior to this time [42]. Thus, a stochastic process $\{\mathbf{z}_t\}_{t\geq 0}$ on a countable state space $\mathcal{S}$ is a CTMC if and only if for all finite sets of times $t_1 < \cdots < t_n < t_{n+1}$

$$\mathbb{P}\left(\mathbf{z}_{t_{n+1}} = i \,\middle|\, \bigcap_{k=1}^{n}\{\mathbf{z}_{t_k} = j_k\}\right) = \mathbb{P}\left(\mathbf{z}_{t_{n+1}} = i \,\middle|\, \mathbf{z}_{t_n} = j_n\right), \qquad \forall i, j_1, j_2, \ldots, j_n \in \mathcal{S}. \tag{2.1}$$

The CTMCs we deal this in the thesis will be time homogeneous[1] meaning

$$p_{i,j}(t) := \mathbb{P}(\mathbf{z}_{t+s} = j \mid \mathbf{z}_s = i) = \mathbb{P}(\mathbf{z}_t = j \mid \mathbf{z}_0 = i), \quad \forall i, j \in \mathcal{S}. \tag{2.2}$$

The transition probabilities $p_{i,j}(t)$ are often collected in a matrix $\mathbf{P}(t) \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ defined as

$$\mathbf{P}(t) = (p_{i,j}(t)), \quad \forall i, j \in \mathcal{S}.$$

It is often impractical to define a CTMC model in terms of the transition probabilities. Moreover, deriving the transition probabilities for an already defined CTMC is infeasible, aside from the simplest of cases. To resolve these issues CTMCs are often defined and

[1] Since all CTMCs we deal with in the thesis are time-homogeneous we will here thereafter refer to them as CTMCs with time homogeneity being implied.

analysed in terms of the infinitesimal generator matrix $\mathbf{Q} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$. The infinitesimal generator matrix is defined as the right time derivative of the transition matrix at $t = 0$.

$$\mathbf{Q} := \lim_{h\to 0^+} \frac{\mathbf{P}(h) - \mathbf{P}(0)}{h} = \lim_{h\to 0^+} \frac{\mathbf{P}(h) - \mathbf{I}}{h}.$$

The negative diagonal elements of $\mathbf{Q}$

$$-q_{i,i} = \sum_{j\in\mathcal{S}\setminus\{i\}} q_{i,j}$$

are important in analysing the holding for each state, defined as $T_i := \inf\{t \geq 0 : \mathbf{z}_t \neq i, \mathbf{z}_0 = i\}$. Intuitively T_i can be thought of as the time the process spends in state i before leaving. By definition $\mathbf{z}_t$ satisfies (2.1) meaning the distribution of T_i is exponentially distributed if $q_{i,i} \neq 0$

$$\mathbb{P}(T_i > t|\mathbf{z}_0 = i) = e^{q_{i,i}t}. \tag{2.3}$$

In the case that $q_{i,i} = 0$, state i is an absorbing state meaning once it is entered it never leaves, meaning $T_i = \infty$ [42]. The probability that the process moves to state j when leaving state i at $t = T_i$ can also be written in terms of the elements of $\mathbf{Q}$

$$\mathbb{P}(\mathbf{z}_{T_i} = j|\mathbf{z}_0 = i) = -\frac{q_{i,j}}{q_{i,i}}, \quad i \neq j. \tag{2.4}$$

Equations (2.3) and (2.4) are both conditioned on $\mathbf{z}_0$. However, due to the memoryless property and time homogeneity (2.3) and (2.4) can be used to determine the time until the next state change and the probability of each change conditioned on any state at any time. As a result, $\mathbf{Q}$ gives a full description of a CTMC avoiding the need to define $\mathbf{P}(t)$ explicitly.

Taken together (2.3) and (2.4) leads us to interpret $\mathbf{Q}$ as a rate matrix. Specifically, the off-diagonal elements $q_{i,j}$ are the rate at which the process moves to state j when in state i. Correspondingly, the negative diagonal elements, $-q_{i,i}$ are the total rate at which the process leaves state i. This interpretation can be used as a framework for formulating models as rates of change are common in understanding most phenomena. In particular, for epidemic modelling the *rate* at which a new individual becomes infectious can be mechanistically understood based on the current number of infectious individuals [43, 44].

For CTMCs, modelling most biological or physical processes, it can be inelegant to directly define them in terms of their infinitesimal generator. This problem is particularly pertinent for models with multi-dimensional state spaces as each element of the state space must be mapped to a natural number corresponding to the index of the $\mathbf{Q}$ matrix.

Instead, it can be useful to define the CTMC in terms of the possible *events* and the rates that these events occur as a function of the state. In this context, when event i occurs the state is updated by adding the corresponding state change $\mathbf{e}_i$ to the current state. The rate that event i occurs is defined as a function of the current state $a_i : \mathcal{S} \to \mathbb{R}_{\geq 0}$. Formulating the state changes $\mathbf{e}_i$ and the event rates a_i fully defines $\mathbf{Q}$ and thus the CTMC. The SIR model, which we define properly in Section 2.4, is generally formulated in this way [43]. The state $\mathbf{z}_t = (S_t, I_t)$ represents the number of susceptible and infectious individuals. Table 2.1 shows the state changes, $\mathbf{e}_i$, and event rate functions $a_i(\mathbf{z}_t)$ for the recovery and infection events. In this thesis, this is the approach we take to define CTMCs.

Event	State Change	Event Rate
Infection	$(-1, 1)$	$I_t \cdot \frac{\beta S_t}{N-1}$
Removal	$(0, -1)$	$I_t \gamma$

Table 2.1: State changes and event rates for the SIR CTMC.

2.1.1 Simulation

Equations (2.3) and (2.4) describe the time until the next event and the probability of each event occurring; this gives a natural way of simulating CTMCs called the stochastic simulation algorithm[2] (SSA) [45, 46]. Algorithm 1 details the SSA, the output of which can be used to reconstruct a full realisation within a given time interval. The computational complexity of the SSA is linearly dependent on the number of events that occur within the simulation. Consequently, simulation of CTMCs with consistently large event rates can be computationally expensive, a fact that will be salient throughout this thesis[3] [49].

2.2 Continuous Time Branching Processes

Continuous time branching processes (CTBPs) are a special class of CTMCs that have useful properties that can be leveraged to analyse them more easily than a general CTMC. For a CTBP $\{\mathbf{z}_t\}_{t\geq 0}$ the i^{th} element of the state vector $[\mathbf{z}_t]_i$ represents the population of agents of type i. In particular, for a model with r different types of agents the state space is[4]

$$\mathcal{S} = \mathbb{N}^r.$$

[2] This algorithm is also referred to as the Doob—Gillespie method.

[3] It must me noted that other methods of exact CTMC simulation have been developed. In some cases these method are faster than the SSA [47, 48]. However, these methods are still poorly suited to cases with fast growing agent populations (the focus of this work).

[4] Here and throughout this thesis we are using the convention that the natural numbers include zero.

Algorithm 1 Stochastic Simulation Algorithm

Input: Initial State $\mathbf{z}_0$, time horizon T, event rates $a_j(.)$ and state changes $\mathbf{e}_i$.

Set $i \leftarrow 0$, $t_0 \leftarrow 0$, $s \leftarrow$ TRUE.
while $s =$ TRUE **do**
    Calculate the rates for each of the E events

$$a_j \leftarrow a_j(\mathbf{z}_i), \quad j = 1, 2, \ldots, E.$$

    Calculate rate parameter for holding time

$$a_0 \leftarrow \sum_{i=1}^{E} a_j.$$

    Simulate holding time

$$\tau \sim \text{Exp}(a_0).$$

    if $t_i + \tau < T$ **then**
        Simulate event

$$k \sim \text{Categorical}\left(\frac{a_1}{a_0}, \frac{a_2}{a_0}, \ldots, \frac{a_E}{a_0}\right).$$

        Update State

$$\mathbf{z}_{i+1} \leftarrow \mathbf{z}_i + \mathbf{e}_k.$$

        Update time $t_{i+1} \leftarrow t_i + \tau$.
        Update index $i \leftarrow i + 1$.
    else
        $s \leftarrow$ FALSE
    end if
end while

Output: Times $\{t_\ell\}_{\ell=0}^{i}$, states $\{\mathbf{z}_\ell\}_{\ell=0}^{i}$.

For CTBPs, events only occur upon the death of an agent, meaning to align with the properties of a CTMC each agent type has exponentially distributed lifetimes. The key property of CTBPs that distinguishes them from general CTMCs is that individual agents behave *independently*. Specifically, both the lifetimes of individual agents and the progeny they produce are independent of all the other agents [50].

We write the rate parameters for the lifetimes of each agent as $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_r)$. When an agent of type i dies they produce $j_1, \ldots, j_r$ agents of type $1, \ldots, r$, respectively, with probability $p_{i,\mathbf{j}}$. Note here $\mathbf{j} = (j_1, \ldots, j_r)$ and is assumed to be fixed and known. The new agents that are produced are referred to as the progeny of the dying agent. Consequently, a CTBP can be fully defined by the behaviour of a single agent of each type. This information is encapsulated with the progeny generating function [50, p. 2–3]

$$P_i(\mathbf{s}) = \sum_{\mathbf{j} \in \mathbb{N}^r} p_{i,\mathbf{j}} \prod_{\ell=1}^{r} s_\ell^{j_\ell}, \quad i \in \{1, \ldots, r\},$$

where $\mathbf{s} = (s_1, \ldots, s_r)$.

Immigration can also be incorporated into CTBPs by modelling the arrival of new agents as a Poisson process. Immigration rates for agent type i at time t are represented by $[\boldsymbol{\alpha}_t]_i$. In this thesis, we allow the immigration rates to change over time but constrain them to be piece-wise constant, i.e. only changing at discrete points in time.

2.2.1 Mean and Variance Calculation

Dorman et al. [50, p. 14–22] has derived equations for the mean and variance of a CTBP with immigration. In this section, we catalogue the salient results for mean and variance calculation. An important quantity in deriving the mean and variance is the characteristic matrix

$$[\boldsymbol{\Omega}]_{ij} = \begin{cases} \omega_i f_{ij}, & i \neq j, \\ \omega_i (f_{ij} - 1), & i = j, \end{cases} \tag{2.5}$$

where $f_{ij} = \frac{\partial}{\partial s_j} P_i(\mathbf{1})$, is the expected number of agents of type j produced from an agent of type i.

Mean Calculation

Making use of (2.5) we can write the matrix defining the mean of the internal dynamics as

$$[\mathbf{F}]_{i,j} := \mathbb{E}[z_i(t+1) | \mathbf{z}_t = \mathbf{u}_j, \boldsymbol{\alpha}_t = \mathbf{0}] = [e^{\boldsymbol{\Omega}}]_{i,j}, \tag{2.6}$$

where $\mathbf{u}_i \in \mathbb{R}^r$ is a vector with 1 in the i^{th} entry and 0 elsewhere, and $\boldsymbol{\alpha}_t = \mathbf{0}$ removes the effect of immgration. Similarly, assuming a unit rate of immigration, the mean of the progeny of immigrants throughout a unit time step is

$$[\mathbf{B}]_{i,j} := \mathbb{E}[z_i(t+1)|\mathbf{z}_t = \mathbf{0}, \boldsymbol{\alpha}_t = \mathbf{u}_j] = \left[\int_0^1 e^{\boldsymbol{\Omega}(1-\tau)}\,\mathrm{d}\tau\right]_{i,j}. \tag{2.7}$$

Given the independence between agents, the mean, incorporating both internal and immigration dynamics can be written as

$$\mathbb{E}[\mathbf{z}_{t+1}|\mathbf{z}_t] = \mathbf{F}\mathbf{z}_t + \mathbf{B}\boldsymbol{\alpha}_t. \tag{2.8}$$

Variance Calculation

To define the variance matrix we consider the variance $\mathbf{V}_i(t)$ associated with a single agent of type i at time t

$$\mathbf{V}_i(t) := \mathrm{Var}(\mathbf{z}_t|\mathbf{z}_0 = \mathbf{u}_i, \boldsymbol{\alpha}_t = \mathbf{0}). \tag{2.9}$$

It is convenient to define $\mathrm{Vec}[\mathbf{V}(t)]$ and $\mathrm{Vec}[\mathbf{C}]$ as the resulting vector when stacking $\mathbf{V}_i(t)$ and $\mathbf{C}_i$ for all agent types into one vector. Using this setup we have

$$\mathrm{Vec}[\mathbf{V}(t)] = \int_0^1 e^{\tau\boldsymbol{\Omega}} \otimes e^{(t-\tau)\boldsymbol{\Omega}^*} \otimes e^{(t-\tau)\boldsymbol{\Omega}^*}\,\mathrm{d}\tau \mathrm{Vec}[\mathbf{C}],$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{\Omega}^*$ denotes the conjugate transpose, and

$$\begin{aligned}
\mathbf{C}_i &:= \mathbf{V}_i'(0) = \omega_i \left(\mathbf{G}_i + \mathbf{u}_i\mathbf{u}_i^T + \mathbf{f}_i\mathbf{f}_i^T - \mathbf{u}_i\mathbf{f}_i^T - \mathbf{f}_i\mathbf{u}_i^T\right), \\
\mathbf{G}_i &:= \mathbf{H}_{P_i}(\mathbf{1}) + \mathrm{diag}(\mathbf{f}_i) - \mathbf{f}_i\mathbf{f}_i^T, \\
\mathbf{f}_i &:= \nabla P_i(\mathbf{1}),
\end{aligned}$$

given $\mathbf{H}_{P_i}$ is the Hessian matrix of $P_i(\mathbf{s})$. The variance associated with immigration is

$$\mathbf{W}_i(t) := \mathrm{Var}(\mathbf{z}_{t+1}|\mathbf{z}_t = \mathbf{0}, \boldsymbol{\alpha}_t = \mathbf{u}_i). \tag{2.10}$$

Further, making use of the Vec[.] operator the matrix can be calculated with

$$\mathrm{Vec}[\mathbf{W}(t)] = \int_0^t \left(\mathrm{Vec}[\mathbf{V}(t-\tau)] + e^{(t-\tau)\boldsymbol{\Omega}^*} \otimes e^{(t-\tau)\boldsymbol{\Omega}^*}\mathrm{Vec}[\mathbf{u}_i\mathbf{u}_i^T]\right)\mathrm{d}\tau,$$

Since we only make use of unit time steps it is useful to define $\mathbf{V}_i(1) := \mathbf{V}_i$ and $\mathbf{W}_i(1) = \mathbf{W}_i$. Using the independence of agents the overall variance is

$$\mathrm{Var}(\mathbf{z}_{t+1}|\mathbf{z}_t) = \sum_{i=1}^r [\mathbf{z}_t]_i \mathbf{V}_i + \sum_{i=1}^r [\boldsymbol{\alpha}_t]_i \mathbf{W}_i. \tag{2.11}$$

2.2.2 Example

For clarity, we consider a simple example of a CTBP. Consider a simple population model with 2 agent types: adolescents and adults, with counts denoted by $z_1(t)$ and $z_2(t)$. Suppose adolescents move into adulthood in an exponentially distributed amount of time with rate δ, and adults are removed after an exponentially distributed amount of time with rate μ. Adults produce new children independently as a Poisson process with rate λ. Since, in this framework, events only occur when agents die this reproduction event is formally encoded as an adult dying and producing both a child and a new adult. The progeny generating function for this model is

$$P_1(\mathbf{s}) = s_2, \tag{2.12a}$$

$$P_2(\mathbf{s}) = \frac{\lambda}{\lambda+\mu} s_1 s_2 + \frac{\mu}{\lambda+\mu}. \tag{2.12b}$$

The corresponding $\mathbf{\Omega}$ matrix (2.5) is

$$\mathbf{\Omega} = \begin{bmatrix} -\delta & \delta \\ \dfrac{\lambda(\lambda+\mu)}{\lambda+\mu} & (\lambda+\mu)\left(\dfrac{\lambda}{\lambda+\mu} - 1\right) \end{bmatrix}. \tag{2.13}$$

For a given set of parameters, we can use (2.8) and (2.11) to derive the mean and variance matrix of the process at different points in time. Figure 2.1 illustrates how this method can be used to describe the mean and variance for the model defined above. These constructions will be useful later in developing approximations of CTBPs that remove the need for simulation.

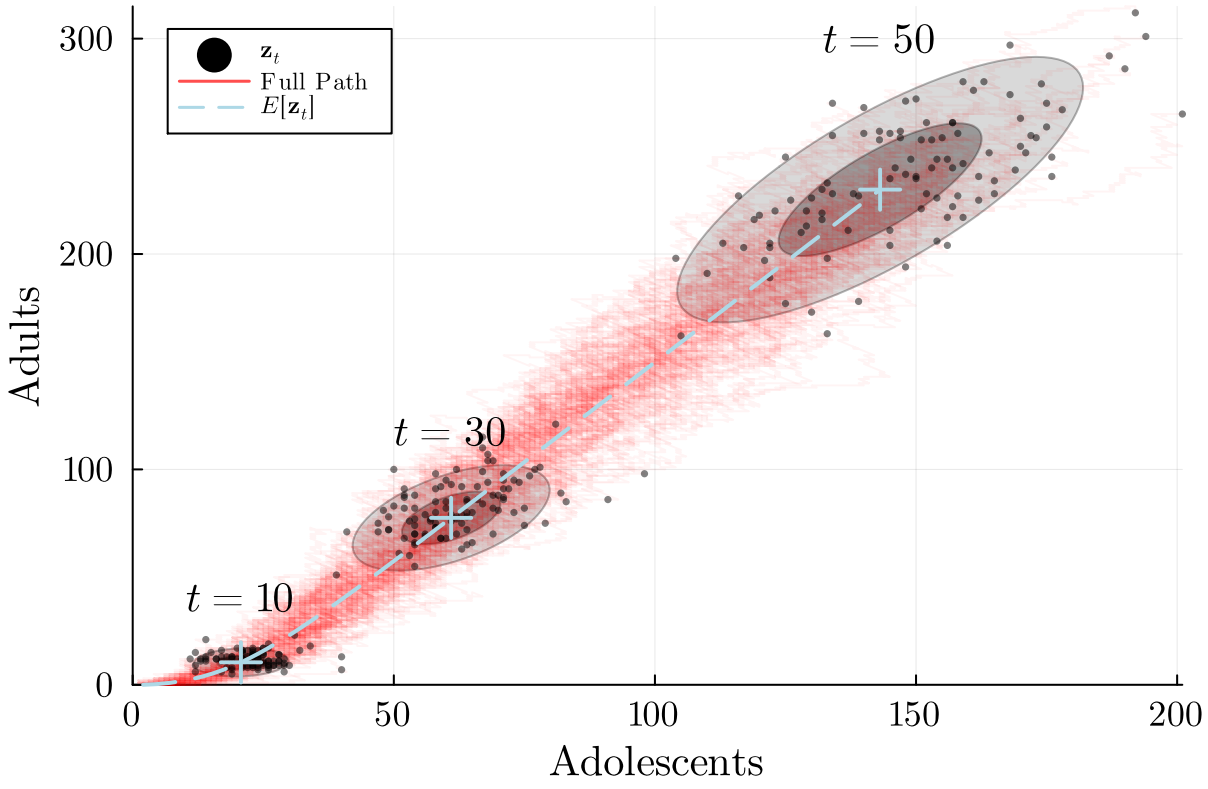


Figure 2.1: Plot showing 100 realisations of the simple population model, with covariance ellipses for the state of the process at $t = 10, 30, 50$. Inner ellipses are for 1 standard deviation and outer ellipses are for 2 standard deviations. The positions and shapes of ellipses are calculated using (2.8) and (2.11). The red lines show the full realisations, with the black dots indicating the state of the realisations at $t = 10, 30, 50$. The parameters used for simulation were $\delta = 1/20$, $\mu = 1/70$, $\lambda = 2/50$ and $\boldsymbol{\alpha} = (0, 1)^T$.

2.3 Bayesian Inference

The term *Bayesian* owes its origin to the 18th-century mathematician and philosopher Thomas Bayes, the namesake of Bayes' theorem

$$\mathbb{P}(\mathbf{X} \mid \mathbf{Y}) = \frac{\mathbb{P}(\mathbf{Y} \mid \mathbf{X})\,\mathbb{P}(\mathbf{X})}{\mathbb{P}(\mathbf{Y})}, \quad \mathbb{P}(\mathbf{Y}) \neq 0. \tag{2.14}$$

The framework that evolved from Bayes' work is now known as Bayesian inference [51]. A vital concept in Bayesian inference is a subjective interpretation of probability, differing starkly from the empirical frequentist interpretation. In particular, Bayesian probability represents the *degree of belief* of a rational subject and will be dependent on the knowledge/data available to this subject [52]. Consequently, even if an event $\mathbf{X}$ in reality has

occurred we can still reason about the probability of $\mathbf{X}$ in the context of the belief of a subject without access to the knowledge that $\mathbf{X}$ has occurred. In contrast, a frequentist may assume $\mathbf{X}$ has not occurred and reason about the consequences of this assumption in relation to our knowledge [53].

This subjective interpretation of probability is particularly useful in quantifying our knowledge about a parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, given we have observed some data $\mathcal{D}$. For a frequentist, parameters are a fixed quantity. As a result, inferences would generally centre around assuming $\boldsymbol{\theta}$ satisfies some set of inequalities and detaining how reasonable this assumption is in the context of the data. In contrast, parameters in the Bayesian framework are interpreted as random variables with distributions dependent on the knowledge of a subject. This knowledge is encapsulated by the distribution of $\boldsymbol{\theta}$ given $\mathcal{D}$, called the *posterior distribution* and written as $p(\boldsymbol{\theta} \mid \mathcal{D})$.

In the context of this thesis, the parameters we are interested in generally relate to the disease dynamics; for example, the mean infection and recovery rate. Parameters relating to the observation of the disease are also important to this work, including the case observation probability and mean reporting delay. The data we use to estimate these parameters is the daily reported case count. In this context, our inferences will take the form of the posterior distribution of the parameters given the reported case counts.

Using Bayes' theorem we can decompose the posterior into a *likelihood* $p(\mathcal{D} \mid \boldsymbol{\theta})$ and a *prior* $p(\boldsymbol{\theta})$

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathcal{D})} = \frac{p(\mathcal{D} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\displaystyle\int_{\boldsymbol{\Theta}} p(\mathcal{D} \mid \tilde{\boldsymbol{\theta}})p(\tilde{\boldsymbol{\theta}})\,\mathrm{d}\tilde{\boldsymbol{\theta}}}. \tag{2.15}$$

Calculating the likelihood requires formulation of a model which mathematically describes the process by which the data is generated [54]. The key challenge in model formulation is achieving two often directly opposed goals simultaneously: capturing real-world phenomena with sufficient accuracy and maintaining tractability in the likelihood.

The prior describes any information we have before observing our data [54]. While there is much philosophical debate surrounding priors, in practice they are often used to represent knowledge that is difficult to quantify mathematically [55]. In particular, prior inferences will often come from modelling a process that is well understood scientifically. One pertinent example of this scientific prior inference would be the knowledge from epidemiology relating to the parameters that mediate the spread of disease through a population. In general, this method of prior construction will be the approach that we take in this work.

2.3.1 Markov Chain Monte Carlo

Evaluation of (2.15) can often prove to be prohibitively difficult due to the normalisation constant

$$p(\mathcal{D}) = \int_{\boldsymbol{\Theta}} p(\mathcal{D} \mid \tilde{\boldsymbol{\theta}}) p(\tilde{\boldsymbol{\theta}}) \, \mathrm{d}\tilde{\boldsymbol{\theta}}. \tag{2.16}$$

This difficulty comes from the required integration of likelihood over the entire parameter space. For many models, the integral is intractable. Given this difficulty, evaluation of the normalisation constant is often avoided by leveraging the fact that we can still calculate the posterior up to proportionality (with respect to $\boldsymbol{\theta}$)

$$p(\boldsymbol{\theta} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{2.17}$$

In particular, Markov chain Monte Carlo (MCMC) methods use this construction to design a Markov chain $\{\boldsymbol{\theta}_t\}_{t \in \mathbb{N}}$ with a stationary distribution that is the same as the posterior distribution [56]. Thus, simulated sample paths from the Markov Chain will follow the desired posterior distribution as long as the Markov Chain has been simulated for long enough for convergence to the stationary distribution to have been achieved. Note that this means that samples simulated before the Markov Chain has converged cannot be expected to come from the posterior. These samples are called *burn-in* and should be removed before conducting any analysis [57]. This brings us to another issue, namely that for most Markov chain constructions $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_{t+\tau}$ will be correlated. Hence, even if the chain has converged, $\{\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1} \dots \boldsymbol{\theta}_T\}$ will not make up a set up *independent* samples from $p(\boldsymbol{\theta}|\mathcal{D})$ despite the marginal distribution of each individual sample being the posterior. In practice, this means that samples from our chain will contain less information than a set of independent samples of the same size. This notion is captured by the effective sample size statistic [58]

$$\mathrm{ESS} = \frac{n}{1 + 2 \sum_{\tau=1}^{\infty} \hat{\rho}(\tau)}, \tag{2.18}$$

where $\hat{\rho}(\tau)$ is an estimate of $\mathrm{cor}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+\tau})$. Here the ESS is an estimate of the number of independent samples that would contain the same amount of information as the correlated samples from the chain. Consequently, chains with high $\mathrm{cor}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+\tau})$ will require many samples to make useful inferences about $\boldsymbol{\theta}$. This fact adds to the many computational challenges associated with MCMC [57].

There are a wide variety of ways to construct an appropriate Markov chain, however, due to the complexity of the likelihood the only relevant method for this work is the Metropolis-Hastings algorithm [59]. The algorithm begins by first sampling $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. Then to obtain the $\boldsymbol{\theta}_t$ we sample a proposed $\boldsymbol{\theta}_t^*$ from some proposal kernel $q(\boldsymbol{\theta}_t^* | \boldsymbol{\theta}_{t-1})$.

This proposed sample is then accepted with probability

$$\alpha(\boldsymbol{\theta}_t^*, \boldsymbol{\theta}_{t-1}) = \min\left\{1, \frac{q(\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_t^*)p(\mathcal{D}|\boldsymbol{\theta}_t^*)p(\boldsymbol{\theta}_t^*)}{q(\boldsymbol{\theta}_t^*|\boldsymbol{\theta}_{t-1})p(\mathcal{D}|\boldsymbol{\theta}_{t-1})p(\boldsymbol{\theta}_{t-1})}\right\}.$$

If the proposed sample is accepted then $\boldsymbol{\theta}_t = \boldsymbol{\theta}_t^*$ otherwise the previous sample is repeated $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$. Note that this method requires the exact calculation of the likelihood, however, this is not possible for many epidemic models. Accordingly, we will instead make use of the pseudo-marginal Metropolis-Hastings algorithm (pmMH) [60].

Pseudo-Marginal Metropolis-Hastings algorithm

Pseudo-Marginal Metropolis-Hastings only requires us to calculate an unbiased estimate of the likelihood $\hat{p}(\boldsymbol{\theta}|\mathcal{D})$. Moreover, for the epidemic models we are concerned with this can be achieved using particle filters, which are discussed in the next section [61]. The pmMH algorithm works very similarly to the standard Metropolis-Hastings algorithm: the only difference is that $p(\mathcal{D}|\boldsymbol{\theta}_t^*)$ is replaced with this unbiased estimate of the likelihood

$$\alpha(\boldsymbol{\theta}_t^*, \boldsymbol{\theta}_{t-1}) = \min\left\{1, \frac{q(\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_t^*)\hat{p}(\mathcal{D}|\boldsymbol{\theta}_t^*)p(\boldsymbol{\theta}_t^*)}{q(\boldsymbol{\theta}_t^*|\boldsymbol{\theta}_{t-1})\hat{p}(\mathcal{D}|\boldsymbol{\theta}_{t-1})p(\boldsymbol{\theta}_{t-1})}\right\}. \tag{2.19}$$

It is not immediately obvious that substituting the exact likelihood with an unbiased estimate in (2.19) will maintain the distributional accuracy of the resulting chain. However, as demonstrated by Andrieu et al. [39] the stationary distribution of this Markov chain will be the target posterior (see Algorithm 2)[5]. In this thesis, we make use of sequential Bayesian filtering to generate unbiased likelihood estimates.

2.3.2 Sequential Bayesian Filtering

Branching process models can be formulated as a hidden Markov model (HMM). HMMs can be written as some Markovian process $\{\mathbf{z}_t\}_{t\in\mathbb{N}}$ which at each discrete point in time generates data $\mathbf{y}_t$ dependant only on $\mathbf{z}_t$ [62, p. 67–71]

$$\mathbf{z}_0 \sim p(\mathbf{z}_0), \tag{2.20a}$$
$$\mathbf{z}_t \sim p(\mathbf{z}_t \mid \mathbf{z}_{t-1}), \tag{2.20b}$$
$$\mathbf{y}_t \sim p(\mathbf{y}_t \mid \mathbf{z}_t). \tag{2.20c}$$

Our data $\mathcal{D}$, therefore, takes the form of an ordered discrete set of observations $\mathbf{y}_{1:T} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$. In the context of epidemic modelling these observations refer to daily confirmed cases. The hidden process $\{\mathbf{z}_t\}_{i=1}^T$ corresponds to the observed states of the

[5] Saving the likelihood in the fashion shown in line 9 is not only more computationally efficient but also required for the Markov chain to converge to the correct distribution [39].

Algorithm 2 Pseudo-Marginal Metropolis-Hastings

Input: Starting state $\boldsymbol{\theta}_0$, total samples n, unbiased likelihood estimator $\hat{p}(\mathcal{D}|\boldsymbol{\theta})$, proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$.

$\mathcal{L}_{\text{current}} \leftarrow \hat{p}(\mathcal{D}|\boldsymbol{\theta}_0)$
for $t \in 1:n$ **do**
    Propose a new sample from the proposal distribution

$$\boldsymbol{\theta}_t^* \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}_{t-1})$$

    Obtain a likelihood estimate for the proposed sample

$$\mathcal{L}_{\text{prop}} \leftarrow \hat{p}(\mathcal{D}|\boldsymbol{\theta}_t^*)$$

    Calculate acceptance probability

$$\alpha \leftarrow \left\{1, \frac{q(\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_t^*)p(\boldsymbol{\theta}_t^*)\mathcal{L}_{\text{prop}}}{q(\boldsymbol{\theta}_t^*|\boldsymbol{\theta}_{t-1})p(\boldsymbol{\theta}_{t-1})\mathcal{L}_{\text{current}}}\right\}$$

    Sample $u \sim \mathcal{U}(0,1)$
    if $\alpha \geq u$ **then**
        $\boldsymbol{\theta}_t = \boldsymbol{\theta}_t^*$
        $\mathcal{L}_{\text{current}} \leftarrow \mathcal{L}_{\text{prop}}$
    else
        $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$
    end if
end for

Output: Sample path $\{\boldsymbol{\theta}_t\}_{t=1}^n$

outbreak e.g. the *actual* number of infectious/exposed individuals.

Sequential Bayesian filtering allows us to obtain two useful quantities:

- An estimate of the distribution of the states given the data and the parameters $\hat{p}(\mathbf{z}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$. This distribution can be used to estimate the full state and parameter posterior $p(\mathbf{z}_{1:T}, \boldsymbol{\theta} \mid \mathbf{y}_{1:T})$.
- An unbiased estimate of the likelihood $\hat{p}(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})$. The data likelihood is used in (2.19) to estimate the parameters.

The filtering procedure works by using the estimate of $p(\mathbf{z}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$ to obtain $p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$ (the predictive distribution) and then using this to find $p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t+1}, \boldsymbol{\theta})$ (the filtering distribution). This step is repeated inductively for $t = 1, 2, \ldots, T-1$. At each step of the process we can also estimate $p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$ which allows for calculation of the overall data likelihood $p(\mathbf{y}_{1:T}, \boldsymbol{\theta})$. In the following section, we present the standard derivation for the general filtering procedure as well as the specific implementation of the bootstrap particle filter [29].

General Filtering Procedure

For notational convenience, we suppress the dependence of $\boldsymbol{\theta}$. Suppose inductively that we have some estimate of $p(\mathbf{z}_t \mid \mathbf{y}_{1:t})$. Using this, $p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t})$ can be calculated by integrating Equation (2.20b) over $\mathbf{z}_t$

$$p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}) = \int_{\mathcal{S}} p(\mathbf{z}_{t+1} \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{y}_{1:t}) \, \mathrm{d}\mathbf{z}_t, \tag{2.21}$$

where $\mathcal{S}$ is the state space of the process. Next, we calculate the likelihood associated with $\mathbf{y}_{t+1}$, analytically written as

$$p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}) = \int_{\mathcal{S}} p(\mathbf{y}_{t+1} \mid \mathbf{z}_{t+1}) p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}) \, \mathrm{d}\mathbf{z}_{t+1}. \tag{2.22}$$

The filtering distribution at $t+1$ can then be written as

$$p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t+1}) \quad \propto \quad p(\mathbf{y}_{t+1} \mid \mathbf{z}_{t+1}) p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}). \tag{2.23}$$

Equations (2.21), (2.22) and (2.23) can be applied inductively for $t \in \{1, 2, \ldots, T-1\}$, allowing for the full likelihood to be derived

$$p(\mathbf{y}_{1:T}) \quad = \quad p(\mathbf{y}_1) \prod_{i=1}^{T-1} p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}). \tag{2.24}$$

Note that, the filtering distributions at $t \in \{2, \ldots, T\}$ are also derived in this process. For a CTBP Equation 2.21 can not be computed explicitly as the exact transition density is intractable for most models [63]. This problem can be overcome with a bootstrap particle filter. This method makes use of simulation to approximate (2.21), (2.22) and (2.23).

Bootstrap Particle Filter

For the bootstrap particle filter (PF) the filtering and predictive distributions are represented by a set of n samples $\left\{\mathbf{z}_t^{(j)}\right\}_{j=1}^{n}$. These samples are called *particles* [62, p. 129–165]. In some applications, particles include a weight representing the likelihood of the corresponding sample. For the specific implementation used in this thesis, all samples are produced in such a way as to have equal likelihood. The starting point for the PF is a set of samples $\left\{\mathbf{z}_t^{(j)}\right\}_{j=1}^{n}$ from the filtering distribution at time t. To draw samples from the predictive distribution (2.21) each $\mathbf{z}_t^{(j)}$ is simulated forward one time step giving $\left\{\tilde{\mathbf{z}}_{t+1}^{(j)}\right\}_{j=1}^{n}$. The predictive distribution can then be estimated

$$\hat{p}(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}) = \frac{1}{n}\sum_{j=1}^{n} \delta(\mathbf{z}_{t+1} - \tilde{\mathbf{z}}_{t+1}^{(j)}). \tag{2.25}$$

To approximate Equation (2.23) numerical integration is used by taking the mean of the observation likelihood over the samples from our predictive distribution.

$$\hat{p}(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}) = \frac{1}{n}\sum_{j=1}^{n} p(\mathbf{y}_{t+1} \mid \tilde{\mathbf{z}}_{t+1}^{(j)}). \tag{2.26}$$

The filtering distribution (2.23) is approximated by resampling the particle from the predictive distribution based on their likelihood given the observed data at $t+1$. In particular, each particle is assigned a normalised weight

$$w^{(j)} = \frac{p(\mathbf{y}_{t+1} \mid \tilde{\mathbf{z}}_{t+1}^{(j)})}{\sum_{i=1}^{n} p(\mathbf{y}_{t+1} \mid \tilde{\mathbf{z}}_{t+1}^{(i)})}. \tag{2.27}$$

Samples from the filtering distribution can be obtained by independently sampling n indices $\left\{a_{t+1}^{j}\right\}_{i=1}^{n}$ from a discrete distribution over $\{1, 2, \ldots, n\}$ with PMF $(w^{(1)}, w^{(2)}, \ldots, w^{(n)})$. These resampling indices correspond to the selected samples meaning

$$\left\{\mathbf{z}_{t+1}^{(j)}\right\}_{j=1}^{n} = \left\{\tilde{\mathbf{z}}_{t+1}^{(a_{t+1}^{j})}\right\}_{j=1}^{n}$$

is from the filtering distribution. The above procedure can then be repeated for $t \in \{1, 2, \ldots, T-1\}$. The overall data likelihood (2.24) can then be estimated

$$\hat{p}(\mathbf{y}_{1:T}) = \hat{p}(\mathbf{y}_1) \prod_{i=1}^{T-1} \hat{p}(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}). \tag{2.28}$$

At this stage, we must note that this method will require a significant amount of forward simulation of the CTBP. Moreover, exact forward simulation of a CTBP has time complexity $\mathcal{O}(\mathcal{E})$, where $\mathcal{E}$ is the total number of events that occur [40]. Hence, once agent counts get very large, the forward simulation becomes very slow. This motivates the derivation of an approximation in Chapter 3 which allows us to analytically solve Equations (2.21), (2.22) and (2.23), for these time steps where many events are expected to occur.

Backwards Smoothing

In modelling contexts where state estimates are important, the goal is to obtain the smoothed distributions $p(\mathbf{z}_t \mid \mathbf{y}_{1:T})$ for each time step. The bootstrap particle filter only obtains the filtering distributions $p(\mathbf{z}_t \mid \mathbf{y}_{1:t})$. Since smoothed distributions are based on the full data-set, rather than the data up to t, they provide a better estimate of the state. In general, this is done by inductively solving/estimating

$$p(\mathbf{z}_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) = \int_{\mathcal{S}} p(\mathbf{z}_t \mid \mathbf{z}_{t+1}) p(\mathbf{z}_{t+1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) \, \mathrm{d}\mathbf{z}_{t+1}, \quad t \in \{T-1, \ldots, 2, 1\}. \tag{2.29}$$

The starting point of the induction is $p(\mathbf{z}_T \mid \boldsymbol{\theta}, \mathbf{y}_{1:T})$. For the bootstrap particle filter, (2.29) can be estimated numerically. A sample $\left\{\hat{\mathbf{z}}_{t+1}^{(j)}\right\}_{j=1}^{N}$ from $p(\mathbf{z}_{t+1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T})$ is obtained inductively from previous steps. If the resampling indices a_t^j are recorded, then for each $\hat{\mathbf{z}}_{t+1}^{(j)}$ we have the corresponding $\mathbf{z}_t^{(a_t^j)}$. More clearly, for each $\hat{\mathbf{z}}_{t+1}^{(j)}$ we have a sample $\mathbf{z}_t^{(a_t^j)}$ from $p(\mathbf{z}_t \mid \hat{\mathbf{z}}_{t+1}^{(j)})$. Given (2.29) the sample $\left\{\mathbf{z}_t^{(a_t^j)}\right\}_{j=1}^{N}$ which we write as $\left\{\hat{\mathbf{z}}_t^{(j)}\right\}_{j=1}^{N}$ follows the desired distribution

$$\hat{\mathbf{z}}_t^{(j)} \sim p(\mathbf{z}_t | \boldsymbol{\theta}, \mathbf{y}_{1:T}), \quad j = 1, 2, \ldots, N.$$

Thus, starting with the sample from the filtering distribution at $t = T$, which comes from the bootstrap particle filter, we can inductively obtain samples from $p(\mathbf{z}_t | \boldsymbol{\theta}, \mathbf{y}_{1:T})$ for $t = T-1, T-2, \ldots, 1$. In practice, the sampled paths from a single run will be highly correlated, often having all samples being identical for earlier time steps [62, p. 189–227]. However, as we will in the next section this will be repeated for many sets of parameters alleviating this issue [64].

Obtaining Joint State and Parameter Posteriors with pmMH

The joint state and parameter posteriors $p(\mathbf{z}_t, \boldsymbol{\theta}|\mathbf{y}_{1:T})$ can be sampled from with pmMH, for $t = 1, 2, \ldots, T$. Note that each time the bootstrap particle filter is used to estimate the likelihood, backward smoothing can be used to obtain samples from $p(\mathbf{z}_t|\boldsymbol{\theta}, \mathbf{y}_{1:T})$, for $t = 1, 2, \ldots, T$. Since

$$p(\mathbf{z}_t, \boldsymbol{\theta}|\mathbf{y}_{1:T}) \propto p(\mathbf{z}_t|\boldsymbol{\theta}, \mathbf{y}_{1:T})p(\boldsymbol{\theta}|\mathbf{y}_{1:T}),$$

saving the samples from $p(\mathbf{z}_t|\boldsymbol{\theta}, \mathbf{y}_{1:T})$ for accepted $\boldsymbol{\theta}$ will give samples from the joint posterior. In practice, it is best to only take one sample from $p(\mathbf{z}_t|\boldsymbol{\theta}, \mathbf{y}_{1:T})$ for each $\boldsymbol{\theta}$ to ensure each sample is independent [64].

2.4 Epidemic Modelling

2.4.1 Compartmental Models

In this thesis the model of the *true* disease dynamics $\{\mathbf{z}_t\}_{t\geq 0}$ will be a compartmental epidemic model. Compartmental models assign each individual in a population a type that determines their behaviour. One of the foundational compartmental models for the spread of infectious disease throughout a population is the susceptible-infectious-removed (SIR) model [65]. While there are many more complex compartmental epidemics models SIR serves as a useful introduction. The SIR model partitions individuals into susceptible, infectious, or removed and then describes the rate at which infectious individuals infect others and the rate at which they are removed. Susceptible individuals are vulnerable to infection, infectious individuals can infect others and removed individuals have already been infected and are thus assumed to be immune. Individuals begin as susceptible, becoming infectious after transmission and are eventually removed becoming immune to future infection.

Although the model was originally formulated as a deterministic model, it can be easily reformulated as a CTMC, which can more effectively model the inherent variability present in the early stages of a disease outbreak [66]. For a population of size N the state space of the CTMC is simply the number of susceptible S_t and infectious I_t individuals in the population

$$\mathcal{S} = \{(S_t, I_t)|S_t + I_t \leq N, S_t \in \mathbb{N}, I_t \in \mathbb{N}\}.$$

The SIR model assumes homogeneous mixing, meaning all individuals interact equally with each other, hence, individuals that an infectious individual interacts with are sampled randomly and uniformly from the population. An infectious contact refers to an interaction that would cause transmission of the disease if one individual is infectious and the other is susceptible. For a single infectious individual, infectious contacts are modelled as a Poisson process with rate β. Consequently, the rate of infection, for a

single infectious individual is simply β multiplied by the probability of interacting with a susceptible individual. Thus, the overall rate of infection is

$$a_1((S_t, I_t)) = I_t \cdot \frac{\beta S_t}{N-1}.$$

Infectious individuals are assumed to be removed[6] at a rate of γ, which assumes an exponentially distributed infectious period with mean $1/\gamma$. Table 2.2 shows the CTMC formulation of the SIR model.

Event	State Change	Event Rate
Infection	$(-1, 1)$	$I_t \cdot \frac{\beta S_t}{N-1}$
Removal	$(0, -1)$	$I_t\gamma$

Table 2.2: State changes and event rates for the SIR CTMC.

An important parameter that can be derived from these models is the basic reproduction number R_0, which is the mean number of secondary infections one infected person will cause, where all other individuals are susceptible to infection. Calculation of R_0 is crucial as it determines whether an outbreak is likely to die out or spread throughout the population [44, 67]. In particular, if $R_0 > 1$ the disease will likely have a sustained spread throughout the population and infect a significant proportion of the population. However, if $R_0 < 1$ the disease is likely to be eliminated quickly without infecting a large proportion of the population [68]. For the SIR model, assuming all individuals are susceptible, infections caused by an infectious individual are simply a Poisson process with rate β which continues for the duration of the infectious period thus

$$R_0 = \mathbb{E}[i\tau], \quad i|\tau \sim \text{Pois}(\beta\tau), \quad \tau \sim \text{Exp}(\gamma),$$

where τ is the length of the infectious period and i is the number of infections caused. Hence, for the SIR model, the basic reproduction number is

$$R_0 = \frac{\beta}{\gamma}$$

While the SIR model is very simplistic, the more complex compartmental epidemic models that we make use of in this thesis operate in the same basic framework. One extension we make use of is the $\text{SE}^{(n)}\text{I}^{(m)}\text{R}$ model [69, 70]. This model includes a latent period where infected individuals are not infectious for a period of time after their infection. Another key feature of this model is shown in Figure 2.2, this being the multiple stages

[6] In this context removal refers to death or recovery.

for the infectious and latent period. Consequently, the latent and infectious periods have generalised Erlang distributions [71]. This flexibility is important as the exponential distribution tends to be a poor fit for infectious and latent periods observed in the real world [69]. The key feature that makes an exponential distribution, assumed by the SIR model, a poor fit to the real-world data is the memoryless property, something we certainly would not expect for an infectious or latent period.

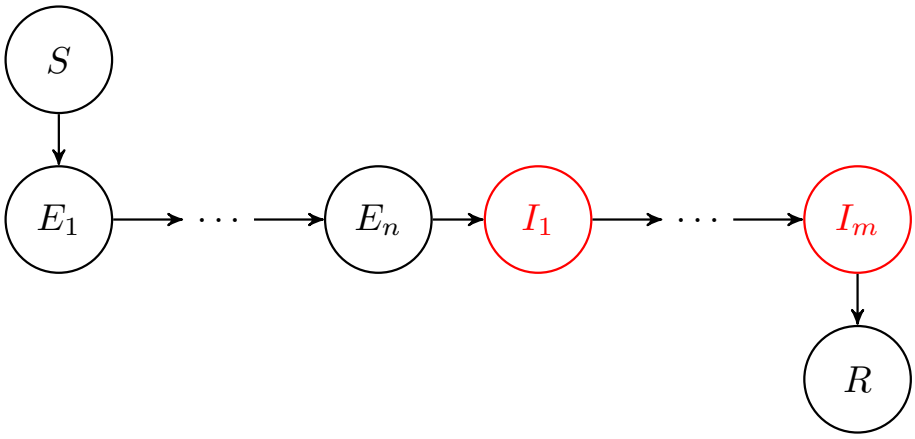


Figure 2.2: States of an individual for the $\mathrm{SE}^{(n)}\mathrm{I}^{(m)}\mathrm{R}$ model. Red states are infectious and will produce new agents of type E_1.

2.4.2 Branching Process Approximations

The SIR model variants are not CTBPs since the infection rate for an individual agent is dependent on the proportion of susceptible people in the population. During the early phases of an outbreak when the vast majority of the population is susceptible this ratio is close to 1

$$\frac{S_t}{N-1} \approx 1, \tag{2.30}$$

making the rate of infection for an individual infectious person approximately β, removing the dependence on other agents [63]. Table 2.3 shows the formulation of the branching process approximation of the SIR model. Similar approximations can be made for the more complex variants of the SIR model making use of (2.30) to remove the inter-agent dependence.

Ball and Donnelly [72] has shown strong convergence of these types of branching process approximations to their corresponding CTMC models for large susceptible populations. As a result, it may be reasonable to use a branching process approximation during the early phases of an outbreak when a very small proportion of the population is

infected or recovered. Figure 2.3 illustrates this phenomenon. In the early phases of the outbreak (when the vast majority of the population is susceptible) the branching process approximation corresponds closely with the SIR model. However, as more of the population becomes non-susceptible the branching process approximation becomes less accurate.

Making these types of approximations can be useful since CTBPs are simpler and easier to analyse. In particular, exact analytical calculation of the mean and variance of $\mathbf{z}_{t+1}|\mathbf{z}_t$ is far easier for a CTBP (see Section 2.2.1) than a general CTMC.

Event	Rate (SIR model)	Rate (CTBP Approximation)
Infection	$I_t \cdot \frac{\beta S_t}{N-1}$	$I_t\beta$
Recovery	$I_t\gamma$	$I_t\gamma$

Table 2.3: Table comparing the event rates for the SIR model and the CTBP approximation.

The branching process model only has one agent type: infectious individual. Thus, the progeny generating function is simply

$$P(s) = \frac{\beta}{\beta+\gamma}s^2 + \frac{\gamma}{\beta+\gamma}. \tag{2.31}$$

Furthermore, the corresponding $\mathbf{\Omega}$ is a 1×1 matrix

$$\mathbf{\Omega} = \left[\beta - \gamma\right]. \tag{2.32}$$

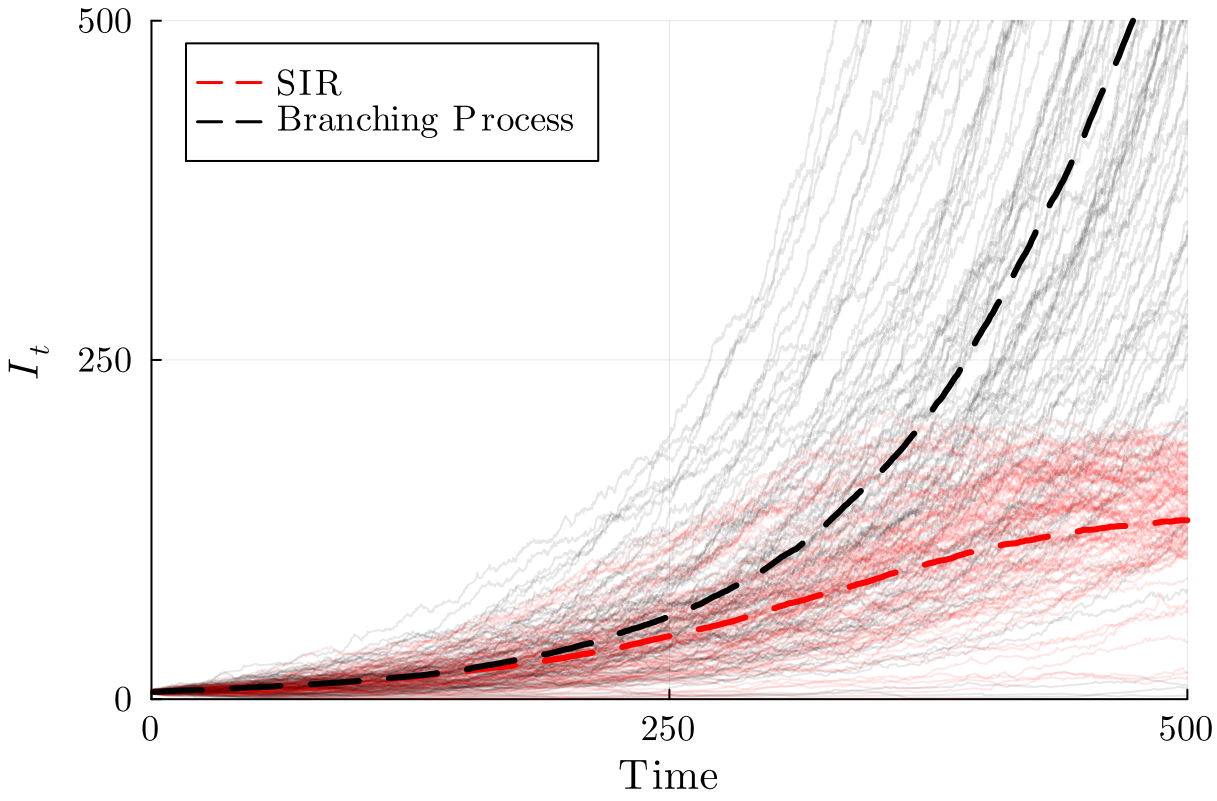


Figure 2.3: Plot showing 100 simulations of the standard SIR model and the branching process approximation. Solid lines show sample paths and dashed lines show the means of the respective simulations. The parameters used for the simulations were $\beta = 0.02$, $\gamma = 0.01$ and $N = 1000$.

Chapter 3

Inference for Partially Observed Continuous Time Branching Processes

In this section, we consider the problem of inference for continuous time multitype branching processes (CTBPs) with partial observation. In particular, we develop an algorithm for state and parameter estimation. The algorithm avoids the computational cost of standard sequential Monte Carlo methods, by exploiting a near-Gaussian property of some components of CTBPs. Essentially the method we propose is an approximation of the bootstrap particle filter. This work is motivated by the specific goal of this project: multi-population early-phase epidemic modelling. However, the method described here can be applied to any multi-type branching process. As a result, we describe our inference method concerning a general branching process model.

3.1 Hidden Markov Models

Here we describe the set of models to which our inference method is applicable too. Consider a hidden Markov model, where the dynamic *hidden* process is a CTBP $\{\mathbf{z}_t\}_{t\geq 0}$, with the immigration, as described in Section 2.2. In general, this method can accommodate a Gaussian observation process of the form

$$\mathbf{y}_t \mid \mathbf{z}_t \sim \mathcal{N}(\mathbf{H}\mathbf{z}_t, \mathbf{R}_t), \tag{3.1}$$

where $\mathbf{y}_t$ is column vector of size d, $\mathbf{H}$ is a $d \times r$ matrix & $\mathbf{R}_t$ is a $d \times d$ matrix. In this thesis, the data we observe is generally discrete, consequently, this observation process represents an approximation. In Section 3.2.3 we see that this approximation is necessary. Additionally, exact observation of $\mathbf{H}\mathbf{z}_t$ can also be incorporated by setting $\mathbf{R}_t$ to the $d \times d$ zero matrix. In this case, the observation distribution has Dirac delta density centred at

$\mathbf{H}\mathbf{z}_t$. In practice our method assumes (3.1) meaning for the results to be accurate (3.1) must be reasonably close to the *true* observation process.

This formulation appears restrictive, in that the observation noise $\mathbf{R}_t$ is not state-dependent. In the context of an epidemic model, we would expect the variance in the observation of reported cases to increase as the number of real cases increases. However, as we will see in Section 3.1.1 this issue can be avoided by modelling part of the observation process in the CTBP.

3.1.1 Example

For clarity, we begin with a simple example of the type of problem we are interested in. Consider a simple population model with 4 agent types, two of these representing actual individuals: adolescents & adults, with counts denoted by $z_1(t)$ & $z_2(t)$. The other two agent types are event counters for partially observed births & deaths, with counts denoted by $z_3(t)$ & $z_4(t)$. Suppose adolescents move into adulthood, and adults die after an exponentially distributed amount of time with rates δ & μ, respectively. Adults produce new children independently as a Poisson process with rate λ. Moreover, reproduction events will produce a birth counter agent with probability p_1 and death events will produce a death counter agent with probability p_2. The progeny-generating function is

$$\begin{aligned} P_1(\mathbf{s}) &= s_2, \\ P_2(\mathbf{s}) &= \frac{(1-p_1)\lambda}{\lambda+\mu}s_1s_2 + \frac{p_1\lambda}{\lambda+\mu}s_1s_2s_3 + \frac{(1-p_2)\mu}{\lambda+\mu} + \frac{p_2\mu}{\lambda+\mu}s_4, \\ P_3(\mathbf{s}) &= s_3, \\ P_4(\mathbf{s}) &= s_4. \end{aligned}$$

The corresponding $\mathbf{\Omega}$ (see Equation (2.5) for details) is

$$\mathbf{\Omega} = \begin{bmatrix} -\delta & \delta & 0 & 0 \\ \frac{\lambda(\lambda+\mu)}{\lambda+\mu} & (\lambda+\mu)\left(\frac{\lambda}{\lambda+\mu}-1\right) & \frac{\lambda p_1(\lambda+\mu)}{\lambda+\mu} & \frac{\mu p_2(\lambda+\mu)}{\lambda+\mu} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.3}$$

Section 2.2.1 includes results from Dorman et al. [50, p. 17–22] showing how this matrix can be used to determine the mean and variance of the progeny produced by a single agent of a given type, at a fixed time, which will be important in approximating the branching process.

Suppose adolescents & adults also immigrate independently as Poisson processes with rates α_1 & α_2 respectively. Section 2.2.1 also gives the mean and variance of the progeny of immigrants over a time-step.

The observation process represents daily case counts rather than cumulative counts. Hence, we will generally set $z_3(t)$ & $z_4(t)$ to zero at the beginning of each time step. Thus, $z_3(t)$ & $z_4(t)$ would represent new daily observed cases. In particular, $z_3(t)$ and $z_4(t)$ are binomial observation processes of the birth and death events. This technique allows for state-dependent observation noise, without explicitly including it in the observation model.

Implicitly we have aligned time units with the observation frequency, hence we can write our data as $\mathbf{y}_{1:t} := \{\mathbf{y}_s\}_{s=1}^{t}$. Recall $z_3(t)$ & $z_4(t)$ refer to counts of observed births and deaths, however in some cases, we may want to add some extra noise to the data model that is not accounted for by partial observation. This extra noise can be modelled with multivariate Gaussian distribution

$$\mathbf{y}_t \mid \mathbf{z}_t \sim \mathcal{N}\left(\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \\ z_4(t) \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right). \tag{3.4}$$

Further, σ_1^2 & σ_2^2 represent the variance in the observation of births & deaths, respectively and we can also allow for a correlation in the observation with $\rho > 0$. Putting this together, the set of parameters for this toy model is

$$\boldsymbol{\theta} := (\lambda, \mu, \delta, \alpha_1, \alpha_2, p_1, p_2, \sigma_1^2, \sigma_1^2, \rho).$$

Our goal is then to infer the parameters $\boldsymbol{\theta}$ and the states $\{\mathbf{z}_t\}_{t=1}^{T}$ given the data set $\{\mathbf{y}_t\}_{t=1}^{T}$.

3.2 Sequential Bayesian Filtering Equations

As described in Chapter 2, we are interested in using the MCMC method to estimate model parameters $\boldsymbol{\theta}$. In this context, the purpose of the sequential Bayesian filtering procedure is to estimate the likelihood $\hat{p}(\mathbf{y}_{1:T}|\boldsymbol{\theta})$ and the marginal distribution of the states given the data and parameters $\hat{p}(\mathbf{z}_t|\mathbf{y}_{1:t}, \boldsymbol{\theta})$. As discussed in Chapter 2 this goal can be achieved by solving

$$p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}) = \int_{\mathcal{S}} p(\mathbf{z}_{t+1} \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{y}_{1:t}) \, \mathrm{d}\mathbf{z}_t, \tag{3.5}$$

$$p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}) = \int_{\mathcal{S}} p(\mathbf{y}_{t+1} \mid \mathbf{z}_{t+1}) p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}) \, \mathrm{d}\mathbf{z}_{t+1}, \tag{3.6}$$

$$p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t+1}) \propto p(\mathbf{y}_{t+1} \mid \mathbf{z}_{t+1}) p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}), \tag{3.7}$$

for $t = 1, \ldots T-1$ and

$$p(\mathbf{y}_{1:T}) = p(\mathbf{y}_1) \prod_{i=1}^{T-1} p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}). \tag{3.8}$$

Aside from the simplest of CTBPs, the transition density $p(\mathbf{z}_{t+1} \mid \mathbf{z}_t)$ is not tractable [50]. As a result, the bootstrap particle filter (PF)[1] is often used. This method numerically estimates (3.5), (3.6) & (3.7) by simulating instances of $p(\mathbf{z}_{t+1} \mid \mathbf{z}_t)$ using the SSA [62, p. 129–165].

As discussed in Section 2.3.2, the SSA can become computationally expensive as the number of events increases. For a CTBP the rate that events occur is linearly dependent on the number of agents [50]. Hence, once agent counts get very large, forward simulation becomes very slow and the PF becomes infeasible. This problem is compounded as we intend to use the PF as part of the pmMH scheme, which requires repeated application of the PF. This issue motivates the method we develop in this chapter. In particular, we make use of an asymptotically valid approximation for the transition density and use this to analytically solve (3.5), (3.6) and (3.7), avoiding any simulation.

3.2.1 Approximation of the Transition Density

We turn our focus to Equation (3.5), in particular the state transition density $p(\mathbf{z}_{t+1} \mid \mathbf{z}_t)$. For a CTBP the transition density is discrete since the state space $\mathbb{N}^r$ is discrete. In this section, we approximate this state space with its continuous analogue in $\mathbb{R}^r$. Since $\mathbf{z}_{t+1}$ is simply the sum of the progeny of each agent at time t and the progeny evolve independently, a central limit argument approximation can be used [73, p. 82–88]. If each $[\mathbf{z}_t]_i$ is sufficiently large for all $i = 1, \ldots, r$, then

$$\mathbf{z}_{t+1} \mid \mathbf{z}_t \dot{\sim} \mathcal{N}(m_t(\mathbf{z}_t), v_t(\mathbf{z}_t)), \tag{3.9}$$

where $m_t(\mathbf{z}_t)$ and $v_t(\mathbf{z}_t)$ are the mean and variance respectively, given by

$$m_t(\mathbf{z}_t) := \mathbb{E}[\mathbf{z}_{t+1} \mid \mathbf{z}_t] = \mathbf{F}\mathbf{z}_t + \mathbf{B}\boldsymbol{\alpha}_t, \tag{3.10}$$

$$v_t(\mathbf{z}_t) := \mathrm{Var}(\mathbf{z}_{t+1} \mid \mathbf{z}_t) = \sum_{i=1}^{r} [\mathbf{z}_t]_i \mathbf{V}_i + \sum_{i=1}^{r} [\boldsymbol{\alpha}_t]_i \mathbf{W}_i, \tag{3.11}$$

[1] The bootstrap particle filter is discussed in more detail in Section 2.3.2.

where $\mathbf{F}$, $\mathbf{B}$, $\mathbf{V}_i$ and $\mathbf{W}_i$ are fixed, computable $r \times r$ matrices (see Section 2.2.1 for definition & calculation). Recall that $\boldsymbol{\alpha}_t$ is the vector immigration rates for each agent type over the time interval $[t, t+1)$

3.2.2 Approximation of Predictive Distribution

Our goal in this section is to derive a Gaussian approximation of the predictive distribution $p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t})$, assuming that (3.9) holds at time t. A second goal is to identify the conditions under which the Gaussian approximation holds. For convenience let $\boldsymbol{\mu}_{n|m}$ and $\boldsymbol{\Sigma}_{n|m}$ denote the mean and variance of the state at time n given data upto time m. We begin by deriving the mean and variance of the predictive distribution

$$\boldsymbol{\mu}_{t+1|t} = \mathbb{E}[\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}] = \mathbf{F}\boldsymbol{\mu}_{t|t} + \mathbf{B}\boldsymbol{\alpha}_t \tag{3.12}$$

and variance

$$\boldsymbol{\Sigma}_{t+1|t} = \mathrm{Var}(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}) = \mathbf{F}\boldsymbol{\Sigma}_{t|t}\mathbf{F}^T + v_t(\boldsymbol{\mu}_{t|t}). \tag{3.13}$$

Making use of Equation (3.9), for sufficiently large $[\mathbf{z}_t]_i$, for $i = 1, \ldots, r$ we can rewrite Equation (3.7) as

$$p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}) \approx \int_{\mathcal{S}} \phi(\mathbf{z}_{t+1}; m_t(\mathbf{z}_t), v_t(\mathbf{z}_t)) p(\mathbf{z}_t \mid \mathbf{y}_{1:t}) \, \mathrm{d}\mathbf{z}_t, \tag{3.14}$$

where $\phi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the pdf of a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, evaluated at $\mathbf{x}$. Equation (3.14) furthers our goal of analytical evaluation of this integral. However, what we would like to say is that this whole integral reduces to a Gaussian density.

The main obstacle in formulating a Gaussian approximation of the predictive distribution is that the variance of the state transitions $v_t(\mathbf{z}_t)$ is state dependent. In particular, if the $v_t(\mathbf{z}_t)$ differs significantly across samples from $\mathbf{z}_t \mid \mathbf{y}_{1:t}$ then the distribution can become highly skewed. However, when this state dependence is inconsequential this issue is mitigated. The severity of the state dependence is modulated by the relative size of $[\mathbf{z}_t]_i$ compared with $[\boldsymbol{\Sigma}_{t|t}]_{i,i}^{\frac{1}{2}}$. To see this effect, consider the variance of a sample from the filtering distribution at $\mathbf{z}^{(j)}(t) \sim \mathbf{z}_t \mid \mathbf{y}_{1:t}$,

$$\begin{aligned} v_t(\mathbf{z}^{(j)}(t)) &= v_t(\mathbf{z}_{t|t} + \boldsymbol{\epsilon}_j) \\ &= \sum_{i=1}^{r} [\boldsymbol{\mu}_{t|t}]_i \mathbf{V}_i + \sum_{i=1}^{r} [\boldsymbol{\alpha}_t]_i \mathbf{W}_i + \sum_{i=1}^{r} [\boldsymbol{\epsilon}_j]_i \mathbf{V}_i \\ &\approx v_t(\boldsymbol{\mu}_{t|t}), \end{aligned}$$

when

$$[\boldsymbol{\mu}_{t|t}]_i \gg |[\boldsymbol{\epsilon}_j]_i|, \quad \forall i \in \{1, 2, \ldots, r\}\,. \tag{3.15}$$

The size of $|[\boldsymbol{\epsilon}_j]_i|$ is influenced by $[\boldsymbol{\Sigma}_{t|t}]^{\frac{1}{2}}_{i,i}$, meaning the approximation will work best when the mean of the filtering distribution is significantly larger than the corresponding element-wise standard deviation. Figure 3.1 illustrates this phenomenon; when $\boldsymbol{\mu}_{t|t}$ is small the difference between the variance transition densities is significant making the real predictive distribution skewed. However, when $\boldsymbol{\mu}_{t|t}$ is large these differences are less significant and the predictive distribution is well approximated with a Gaussian distribution.

Given these conditions are sufficiently well met the transition density in the integrand of (3.14) becomes

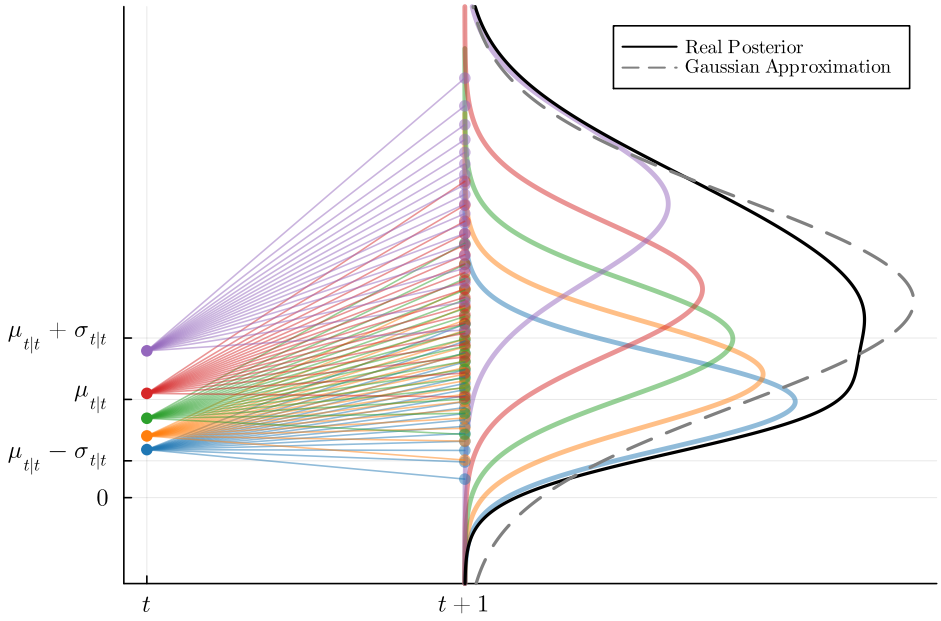
$$\phi(\mathbf{z}_{t+1}; m_t(\mathbf{z}_t), v_t(\mathbf{z}_t)) \approx \phi(\mathbf{z}_{t+1}; m_t(\mathbf{z}_t), v_t(\boldsymbol{\mu}_{t|t})). \tag{3.16}$$

Given this, we can write the predictive distribution as a sum of a Gaussian random variable and an affine transformation of a sample from the filtering distribution at t

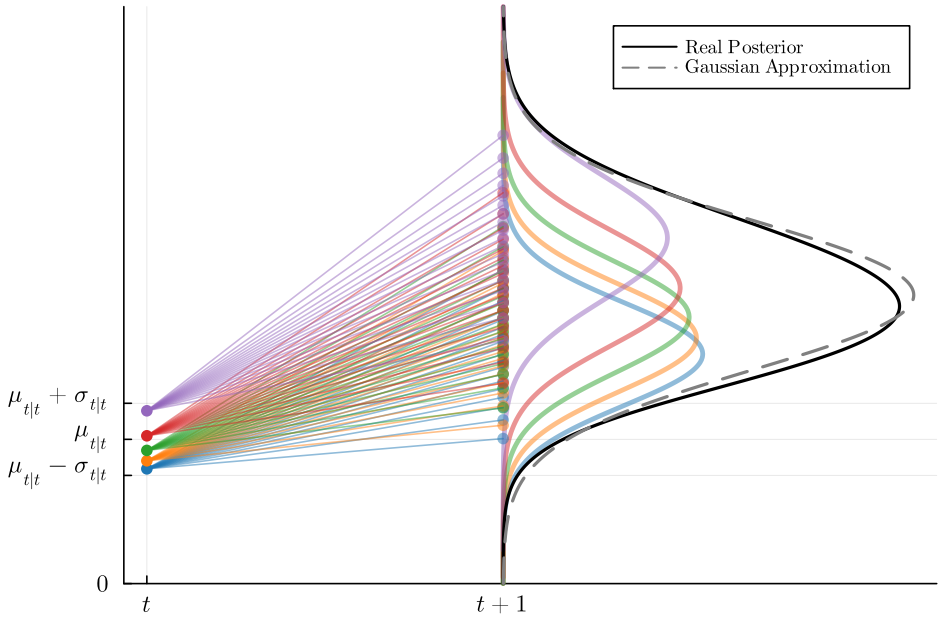
$$\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t} = \mathcal{N}(\mathbf{0}, v_t(\boldsymbol{\mu}_{t|t})) + m_t(\mathbf{z}_t^{(j)}), \quad \mathbf{z}_t^{(j)} \sim p(\mathbf{z}_t \mid \mathbf{y}_{1:t}). \tag{3.17}$$

Equation (3.17) is valid whenever (3.15) holds, illustrating how the predictive distributions can become approximately Gaussian. Note that if $p(\mathbf{z}_t \mid \mathbf{y}_{1:t})$ is Gaussian the predictive distribution becomes Gaussian. Consequentially, all the proceeding predictive and filtering distributions are also Gaussian. This explains why these predictive and filtering distributions appear Gaussian in our simulation studies as $\boldsymbol{\mu}_{t|t}$ gets larger (Section 3.3). Replacing $m_t(\mathbf{z}_t)$ and $v_t(\mathbf{z}_t)$ with the notation introduced before (3.14), we obtain the Gaussian predictive distribution

$$\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t} \dot{\sim} \mathcal{N}(\boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t}). \tag{3.18}$$



(a) $\boldsymbol{\mu}_{t|t} = 5$, $\sigma_{t|t} = 5$



(b) $\boldsymbol{\mu}_{t|t} = 20$, $\sigma_{t|t} = 5$

Figure 3.1: Illustrative, plots comparing the Gaussian approximation of the predictive distribution $p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$ with the real distribution, with a small (a) and large (b) posterior mean at t. Here, the black shaded curves represent the transition densities for each of the samples form $\mathbf{z}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}$. The samples at t are drawn from an exponential distribution and the transition density is assumed to be Gaussian.

3.2.3 Filtering Distribution

Assume the Gaussian approximation (3.18) is satisfied, and consider the observation process given in (3.1) we can write the filtering distribution as

$$p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t+1}) \propto \phi(\mathbf{z}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t})\phi(\mathbf{y}_{1:t+1}; \mathbf{H}\mathbf{z}_{t+1}, \mathbf{R}_{t+1})$$

which is identical to the filtering step of a standard Kalman Filter (see Appendix 6), meaning

$$p(\mathbf{z}_{t+1} \mid \mathbf{y}_{1:t+1}) = \phi(\mathbf{z}_{t+1}; \boldsymbol{\mu}_{t+1|t+1}, \boldsymbol{\Sigma}_{t+1|t+1}), \tag{3.19}$$

where

$$\begin{aligned} \mathbf{K}_t &:= \boldsymbol{\Sigma}_{t+1|t}\mathbf{H}(\mathbf{H}\boldsymbol{\Sigma}_{t+1|t}\mathbf{H}^T + \mathbf{R}_{t+1})^{-1}, && (3.20)\\ \boldsymbol{\mu}_{t+1|t+1} &= \boldsymbol{\mu}_{t+1|t} + \mathbf{K}_t(\mathbf{y}_{t+1} - \mathbf{H}\boldsymbol{\mu}_{t+1|t}), && (3.21)\\ \boldsymbol{\Sigma}_{t+1|t+1} &= (\mathbf{I} - \mathbf{K}_t\mathbf{H})\boldsymbol{\Sigma}_{t+1|t}. && (3.22) \end{aligned}$$

This brings us to the marginal likelihood calculation (3.6) which will also reduce to a Gaussian density

$$p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}) = \phi(\mathbf{y}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t}). \tag{3.23}$$

These equations constitute one step of the Gaussian approximation, moving from time t to $t+1$.

3.2.4 Switching between Filtering Methods

In this section, we introduce a method that, based on a thresholding step from (3.15), adaptively chooses whether to use the Gaussian approximation or the PF based on the estimated agent count at each observation time. The derivation in Section 3.2.3 rests on (3.1) holding at time $t+1$ and (3.9) and (3.16) holding at time t. Equation (3.9) will not hold in many common modelling situations, in particular in the early phases of some epidemic models, where the agent count is low, so the generic situation is one in which we begin filtering with a PF and later switch to a Gaussian filter.

We define the set $\mathcal{T}$ as the set of time steps where these assumptions are all reasonable. Meaning, if $t \in \mathcal{T}$ we can derive the filtering distribution and likelihood estimate for t using the Gaussian approximation. We discuss practical estimation of $\mathcal{T}$ later (see Equations (3.28) and (3.29)). Suppose $\tau \notin \mathcal{T}$ but $\tau + 1 \in \mathcal{T}$, meaning we want to use the Gaussian approximation to obtain posteriors and likelihoods for $t = \tau + 1$ but we used a PF for $t = \tau$. In this case our state posterior for t exists in the form of a set of samples $\{\mathbf{z}^{(1)}(\tau), \ldots, \mathbf{z}^{(n)}(\tau)\}$ from $\mathbf{z}_t \mid \mathbf{y}_{1:t}$. To make use of the Gaussian approximation for the

next step we simply obtain an unbiased estimate of the mean and variance of our state posterior for $t = \tau$

$$\hat{\boldsymbol{\mu}}_{\tau|\tau} := \frac{1}{n}\sum_{k=1}^{n}\mathbf{z}^{(k)}(\tau), \quad \hat{\boldsymbol{\Sigma}}_{\tau|\tau} := \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{z}_\tau^{(k)} - \hat{\boldsymbol{\mu}}_{\tau|\tau})(\mathbf{z}_\tau^{(k)} - \hat{\boldsymbol{\mu}}_{\tau|\tau})^T, \tag{3.24}$$

$$p(\mathbf{z}_\tau|\mathbf{y}_{1:\tau}) = \phi(\mathbf{z}_\tau; \hat{\boldsymbol{\mu}}_{\tau|\tau}, \hat{\boldsymbol{\Sigma}}_{\tau|\tau}) \tag{3.25}$$

Given this Gaussian density, we can use the method described in Section 3.2.2 to solve Equations (3.7) and (3.6), producing the necessary state and likelihood estimates for $t = \tau + 1$. More rarely, we may want to switch from the Gaussian approximation to the PF, and in this case, we can simply draw n rounded and censored samples from $\mathcal{N}(\boldsymbol{\mu}_{\tau|\tau}, \boldsymbol{\Sigma}_{\tau|\tau})$, assign them equal weights, and continue with the PF. Finally, we have fully defined the Gaussian approximation allowing us to specify the last of our filtering equations

$$p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) = p(\mathbf{y}_1 \mid \boldsymbol{\theta}) \prod_{t\in\mathcal{T}} \phi(\mathbf{y}_t; \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}) \prod_{t\in\{1,2,\ldots,T\}\setminus\mathcal{T}} \hat{p}(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}), \tag{3.26}$$

where $\hat{p}(\mathbf{y}_t \mid \mathbf{y}_{1:t-1})$ is the likelihood estimate obtained from the t^{th} step of the PF (2.26) [62, p. 129–165]. We summarise the Gaussian approximation algorithm in Algorithm 3.

Defining which time steps to use the PF and the Gaussian approximation is a trade-off between computational time and accuracy. The time complexity of a single step of the Gaussian filter is independent of the number of samples being used and the total agent population, unlike the PF. Thus, the overall time complexity of the Gaussian approximation (as a function of the total number of events) is well approximated by the complexity of the steps when the PF is employed

$$\mathcal{O}\left(n\sum_{t\in\mathcal{T}}\mathcal{E}_t\right),$$

where n is the total number of PF samples and $\mathcal{E}_t$ is the number of events that occur over the interval $[t-1, t]$. Consider that, $\mathcal{E}_t$ is highly correlated with the total number of agents at time t, a quality which modulates the accuracy of Equation (3.9). Further, the variance in the likelihood estimate will also be lower if the Gaussian approximation is used more often, as no simulation is taking place during these steps. This effect is of great importance when using this method as part of a pseudo-Marginal Metropolis-Hastings scheme, where lower variance estimators improve the mixing of chain up to a point [74, 61]. In practice, the n required to obtain sufficiently low variance in the likelihood estimate will

Algorithm 3

Gaussian approximation for obtaining a likelihood estimate and state posterior distribution. Here the function PF(.) denotes a single step of the bootstrap particle filter with resampling [62, p. 129–165].

Input: Observations $\mathbf{y}_{1:T}$, parameters $\boldsymbol{\theta}$.

Sample $\mathcal{Z}_0 \sim p(\mathbf{z}_0 \mid \boldsymbol{\theta})$.
Initialise log-likelihood $\ell \leftarrow 0$.
for $t+1 \in \{1, 2, \ldots, T-1\}$ **do**
  if $t \in \mathcal{T}$ **then**
    if $t \notin \mathcal{T}$ **then**
      Calculate mean $\boldsymbol{\mu}_{t|t}$ and covariance estimate $\boldsymbol{\Sigma}_{t|t}$ with (3.24).
    end if
    Update $\boldsymbol{\mu}_{t+1|t}$ and $\boldsymbol{\Sigma}_{t+1|t}$ with (3.12) and (3.13).
    Update $\boldsymbol{\mu}_{t+1|t+1}$ and $\boldsymbol{\Sigma}_{t+1|t+1}$ with (3.20) and (3.22).
    Update ℓ with (3.23) in log form.
  else
    if $t \in \mathcal{T}$ **then**
      Sample $\mathcal{Z}_t \sim \mathcal{N}(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})$.
    end if
    $\mathcal{Z}_{t+1}, \ell \leftarrow \mathrm{PF}(\mathcal{Z}, \mathbf{y}_{t+1}, \ell)$.
  end if
end for

Output: The log-likelihood ℓ, the filtering distribution means $\left\{\boldsymbol{\mu}_{t|t}\right\}_{t\in\mathcal{T}}$ and variances $\left\{\boldsymbol{\Sigma}_{t|t}\right\}_{t\in\mathcal{T}}$ and the filtering distribution samples $\{\mathcal{Z}_t\}_{t\in\mathcal{T}^c}$.

be lower for a switching criterion that uses the Gaussian approximation more often.

At any given time we have access to state estimate $\mathcal{Z}_{t-1}$, from the previous step and the observation $\mathbf{y}_t$ from the current step. Hence, we will define a switching criterion A as an indicator function of these quantities

$$\mathcal{T} := \{t \in 1, 2, \ldots, T : A(\mathcal{Z}_t, \mathbf{y}_t) = 1\}. \tag{3.27}$$

The simplest way of defining A is based on the observed data as this can give insight into where we expect the agent counts to be at any given time

$$A(\mathcal{Z}_{t-1}, \mathbf{y}_t) = \begin{cases} 1, & f(\mathbf{y}_t) \geq s, \\ 0, & f(\mathbf{y}_t) < s. \end{cases} \tag{3.28}$$

for a threshold s. This rule is convenient as $\mathcal{T}$ can be derived before running the filter, making the computational time easy to estimate. A more justified way of defining A based on the smallest component of the mean of the state estimate at t

$$A(\mathcal{Z}_{t-1}, \mathbf{y}_t) = \begin{cases} 1, & \min(\mathbb{E}[\mathcal{Z}_t]) \geq s, \\ 0, & \min(\mathbb{E}[\mathcal{Z}_t]) < s. \end{cases} \tag{3.29}$$

This rule is more justified as it is more closely related to our assumptions in Section 3.2.2. However, making A dependent on state estimates could cause issues if the Gaussian approximation is significantly biased in the calculation of the likelihood, as this could cause certain state estimates to be artificially favoured.

In modelling contexts where agent counts are high for a large portion of the time series, it may be reasonable to use Gaussian approximation throughout the entire time series (Section 3.3.2), setting $s = 0$. Using the Gaussian approximation at every time step gives a likelihood estimate with zero variance, which can greatly improve the efficiency of the chain [61]. Consequently, this protocol can be used for more complex parameter estimation problems.

3.2.5 Smoothed State Estimates

To obtain a likelihood estimate that can be used for pmMC we only need to obtain the filtering distribution at each time step $p(\mathbf{z}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$. In modelling contexts where state estimates are important, the goal is to obtain the smoothed distributions $p(\mathbf{z}_t \mid \mathbf{y}_{1:T}, \boldsymbol{\theta})$ for each time step. Since smoothed distributions are based on the full data-set, rather than the data up to t, they provide a better estimate of the state. Fortunately, smoothed distributions can be obtained from the output of the Gaussian approximation. For $t \in$

$\mathcal{T}$ the filtering distributions are Gaussian meaning we can step backwards through the estimates using the Rauch-Tung-Striebel smoother [29]. For $t \notin \mathcal{T}$, we need to keep track of the resampling indexes and update the particle weights stepping backwards through time (see Section 2.3.2). This means that we can also obtain posteriors for the states using the Gaussian approximation.

3.3 Results

In this section we will compare the efficiency, state estimates and parameter estimates produced by the PF and Gaussian approximation for various thresholds. The PF gives distributionally accurate state and parameter estimates. Hence, the Gaussian approximation can be validated by comparing its estimates with those produced by the PF. The purpose of the Gaussian approximation is to produce a faster alternative to the PF with minimal losses in accuracy. The following results illustrate the degree to which this purpose is achieved. In Sections 3.3.1 and 3.3.2 we used simulated data from branching process approximations of the SEIR (Susceptible-Infectious-Recovered) and $\mathrm{SE}^{(2)}\mathrm{I}^{(2)}\mathrm{R}$ epidemic models [64]. Section 3.3.3 uses real data from the Second COVID-19 Wave in Victoria, Australia.

3.3.1 SEIR Branching Process

For the SEIR model branching process the state of our process at a given time is $\mathbf{z}_t = (E(t), I(t), C(t))$, these elements representing the number of exposed, infected and observed symptom onset events respectively. Table 3.1 describes the event rates and corresponding state updates. Note that, symptom onset events correspond to observed symptoms onset events with probability p.

Event	State Update	BP Rate
Infection	$(1, 0, 0)$	$\beta \mathbf{z}_2$
Symptoms onset (observed)	$(-1, 1, 1)$	$p\delta \mathbf{z}_1$
Symptoms onset (unobserved)	$(-1, 1, 0)$	$(1-p)\delta \mathbf{z}_1$
Removal	$(0, 0, -1)$	$\lambda \mathbf{z}_2$
Exposed Immigration	$(1, 0, 0)$	$\boldsymbol{\alpha}_1$
Infectious Immigration	$(0, 1, 0)$	$\boldsymbol{\alpha}_2$

Table 3.1: Table of events and rates for the SEIR branching process approximation

We assume our data, $\mathbf{y}_t$, consists of the number of observed symptom onset events over the period $(t-1, t]$, with Gaussian noise

$$p(\mathbf{y}_t \mid \mathbf{z}_t) = \mathcal{N}(C^*(t), \sigma^2) \tag{3.30}$$

Note that since we are only focused on errors associated with the newly observed symptom onset events we need to define a new variable $C^*(t) = C(t) - C(t-1)$, representing the new cases for the day. The observation variance, σ^2, can be interpreted as extra noise associated with data collection, aside from the binomial observation of symptomatic onsets. Additionally, this variance can act as a regularisation parameter.

To generate, $\mathbf{y}_{1:T}$ we simply fix parameters (values given in Table 3.2) and simulate from our branching process model using the SSA, drawing an observation at each time step (see Figure 3.2). In particular, we choose a slow-growing epidemic with a basic reproduction number of 1.2, an initial state of $\mathbf{0}$ and a low immigration rate. This setup leads to a large portion of the time series having a low agent population, hence probing the worst-case scenario for the Gaussian approximation.

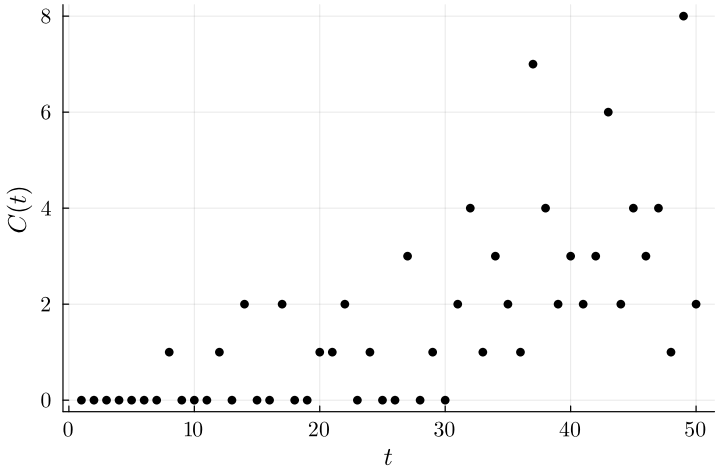


Figure 3.2: Plot of observed cases produced from SEIR simulation.

To analyse the Gaussian approximation accuracy, we fix the parameters $\boldsymbol{\theta}$ at the values used for simulation and compare filtering distributions at different points in the time series. We use the switching criteria from Equation (3.29) with thresholds of 0, 20 and ∞ (note that $s = \infty$ is a bootstrap particle filter which will give the distributionally correct estimates of filtering distribution). Figure 3.3, shows that as agent counts become larger, the Gaussian approximation becomes more accurate, consistent with our analysis in Sections 3.2.1 and 3.2.2. For $t = 8$ the $s = 20$ method and PF are identical as the switch has not occurred, however, the $s = 0$ method is capturing the mean and variance but fails to capture the shape of the distribution. Moving to $t = 35$, we can see that the PF looks significantly more Gaussian but still with some skew, moreover, the $s = 20$ method and $s = 0$ method are similar but the former appears to be a better approximation of the PF. At $t = 50$, all 3 methods are very similar. The positive skew of the early filtering

distributions is a result of the censoring of agent counts at zero. As a result, this skew is reduced as the filtering distributions move further from zero.

Figure 3.3 also indicates that methods using the Gaussian approximation more often give slightly lower estimates for the agent counts. We suspect this effect is caused by not accounting for the skew in the filtering distributions. In particular, the mean and variance of the predictive distribution are correct, however, when calculating the mean of the filtering distribution we assume is predictive distribution is Gaussian. Since the *real* predictive distribution is generally positively skewed, this assumption may cause a systematic error in which the means of the filtering distributions are systemically underestimated.

To examine the accuracy of the parameter estimation we conduct pmMH on our simulated data using the Gaussian approximation at thresholds of $s = 0, 10, 20$ and ∞ and compare the posteriors for the parameters. Figure 3.4 shows the posteriors for R_0, illustrating the importance of using the PF for time steps with low agent populations. As expected, increasing the threshold produces an increase in accuracy. Specifically, it appears that lower thresholds lead to lower estimates for R_0. This underestimation is likely caused by the aforementioned systematic agent count underestimation.

Other parameters were estimated (see Table 3.2), however, the data is not very informative meaning estimates for all methods were visually indistinguishable. Moreover, we see a similar pattern of results for joint posteriors of R_0 and T_E (see Figure 3.5).

Figure 3.6 illustrates the relationship between computation time and the log-likelihood variance for different switching thresholds. For each threshold we simulate 1000 runs, calculate the sample variance and the time taken for each run, repeating this for different numbers of samples. The box plots represent the sample of observed computation times for that number of samples and methods. From these results, it is clear that reducing the threshold can significantly reduce both computation time and the variance of the likelihood estimate. These effects together lead to an overall increase in the efficiency of the pmMH, meaning less time is needed to estimate posteriors to a sufficient detail [61].

Parameter	True value	Prior	Standard Formulation
Basic reproduction number (R_0)	1.2	Gamma$(4.4, 2)$	$R_0 = \beta/\lambda$
Mean latency period (T_E)	5	Gamma$(100, 20)$	$T_E = 1/\delta$
Mean infectious period (T_I)	14	Gamma$(30, 2)$	$T_E = 1/\lambda$
Immigration rate ($\boldsymbol{\alpha}$)	$(0.5, 0.5)^T$	fixed	-
Observation probability (p)	0.8	fixed	-
Observation variance (σ^2)	1	fixed	-
Initial State ($\mathbf{z}_0$)	$(0, 0, 0)^T$	fixed	-

Table 3.2: Parameters and priors used for SEIR simulation study.

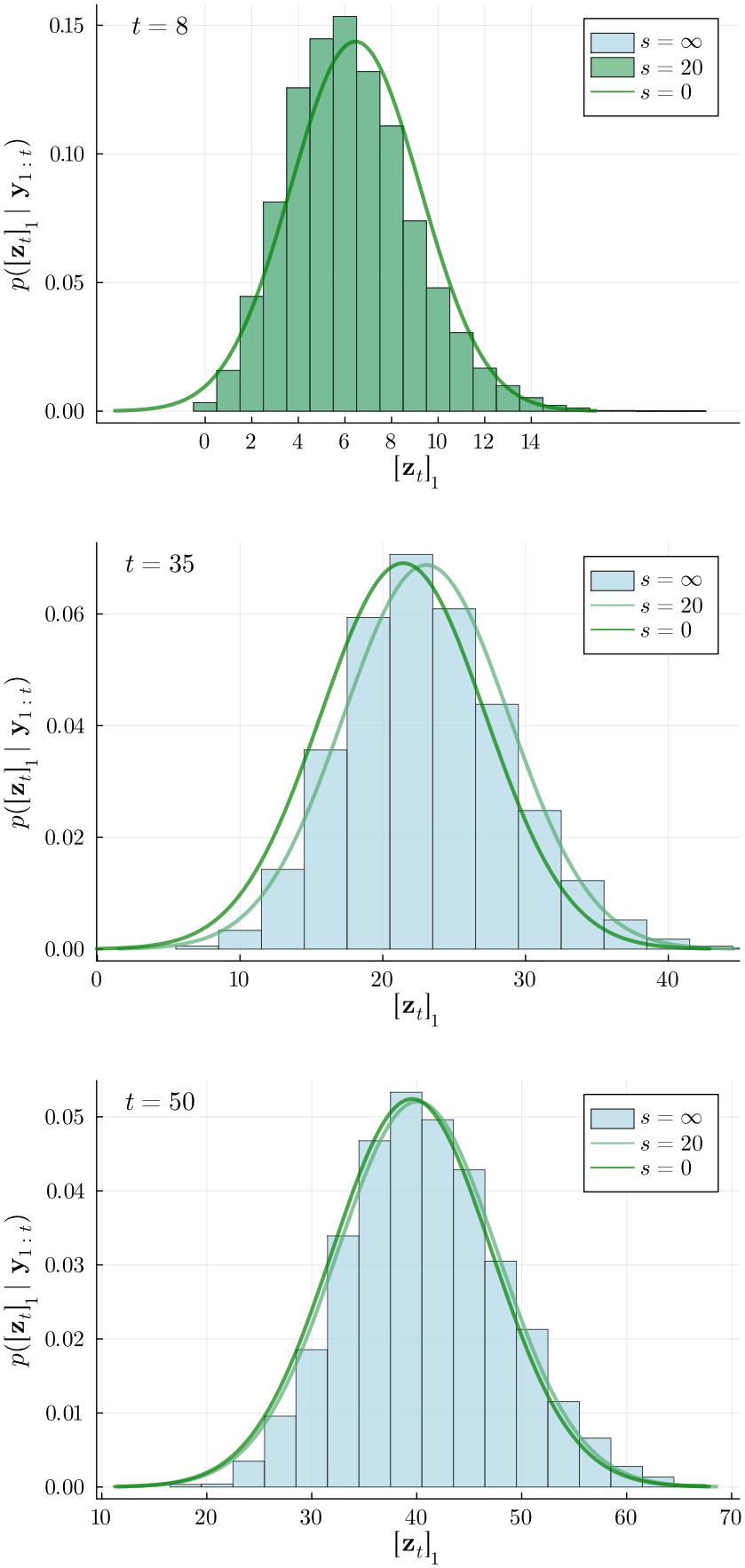


Figure 3.3: Filtering distributions for $\mathbf{z}_2(t)$ for $t = 8, 35, 50$ using the PF and Gaussian approximation for $s = 0, 20$.

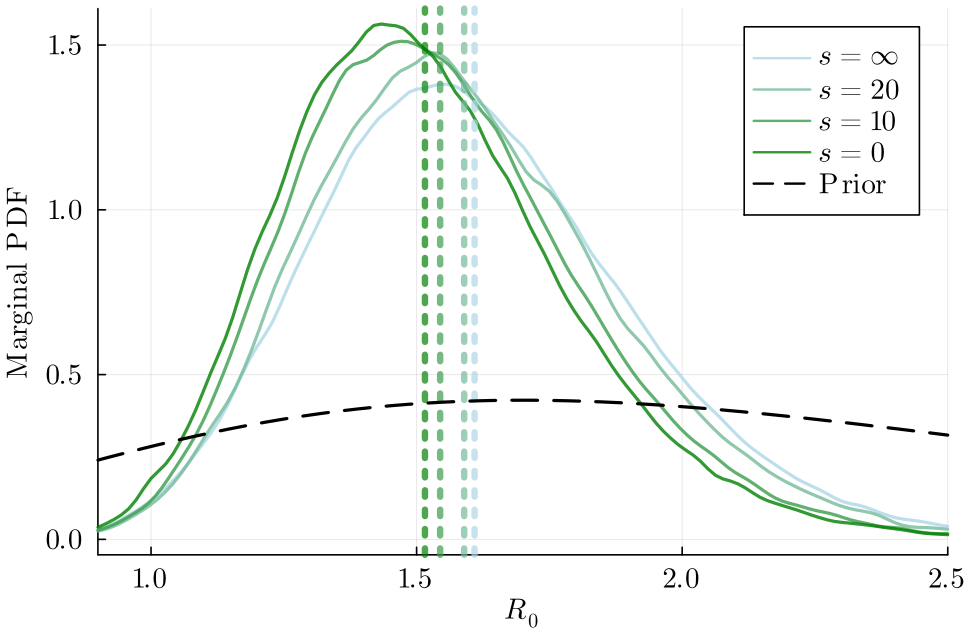


Figure 3.4: Plot of kernel density estimates of the marginal posterior for R_0. Dashed lines indicate the posterior mean. Here the dashed line represents the corresponding posterior mean. As expected from our analysis, all methods produce similar estimates, with the higher thresholds being accurate (closer to the estimate produced by the PF).

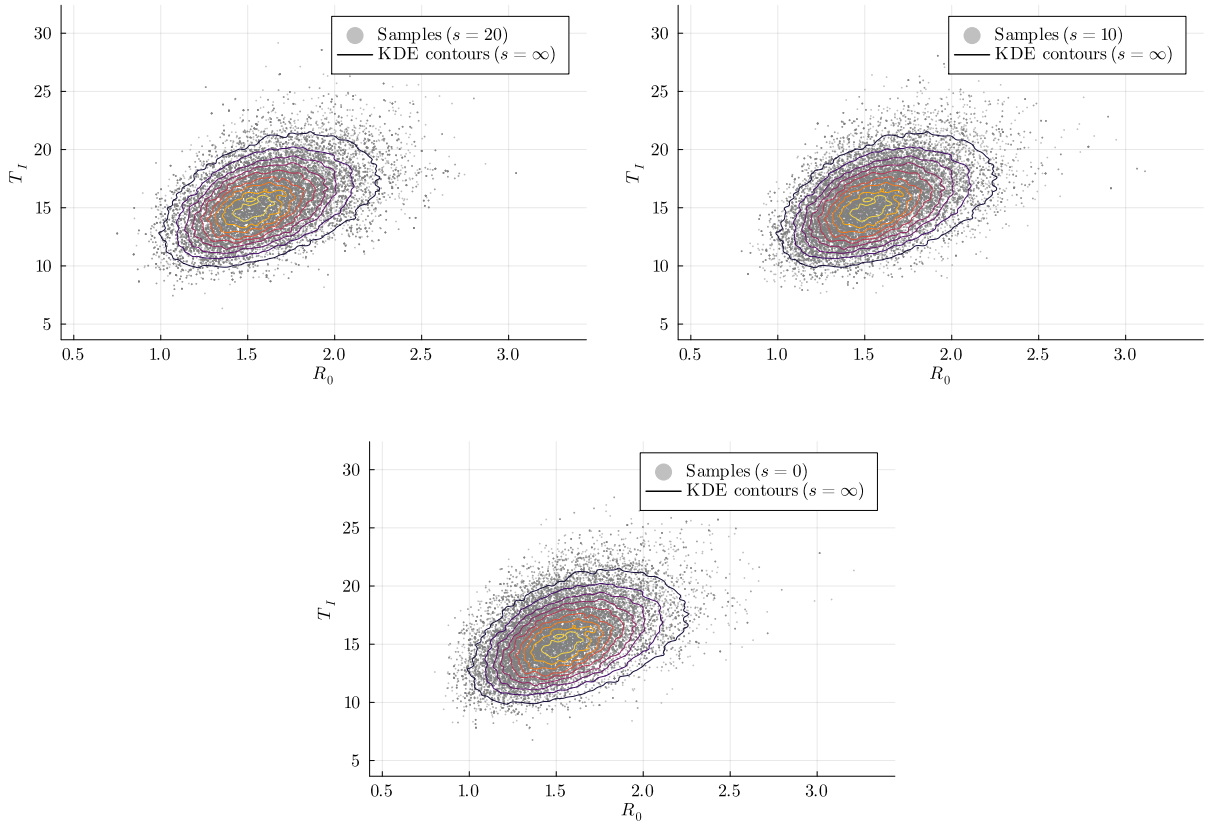


Figure 3.5: Comparisons of the joint posterior $p(R_0, T_I|\mathbf{y}_{1:T})$ produced by the PF with each of the Gaussian approximation estimates. The PF posteriors are represented by the contours of the kernel density estimate (KDE), while the Gaussian approximation posterior are shown as samples.

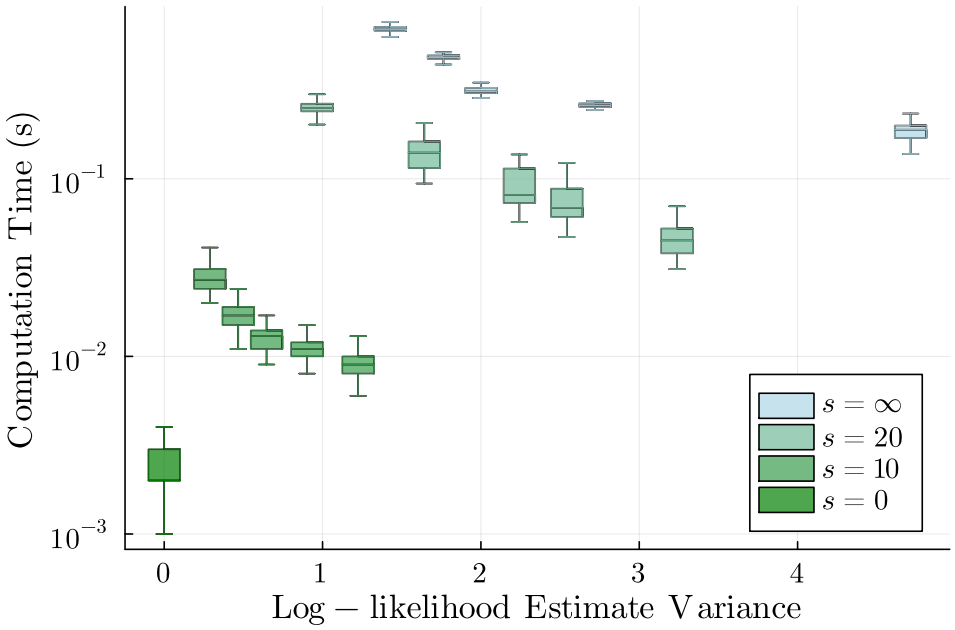


Figure 3.6: Box plot comparing the computational time and variance of the log-likelihood estimates produced by the Gaussian approximation at different thresholds.

3.3.2 $SE^{(2)}I^{(2)}R$ Branching Process

In this section, we repeat the experiment described in Section 3.3.1 but instead with the more complex $SE^{(2)}I^{(2)}R$ model. The state vector is $\mathbf{z}_t = (E_1(t), E_2(t), I_1(t), I_2(t), C(t))$, with $E_i(t)$ and $I_i(t)$ being the number of exposed and infectious individuals in stage i at time t. Here we have 2 stages for the exposed and infectious states, this allows for non-exponentially distributed latent and infectious periods [69, 70]. Given this, Table 3.3 describes the model dynamics. Again, our observations correspond to symptom onset events with probability p. As before, we use the observation distribution defined in (3.30).

For this simulation, we have a much higher R_0 and a large starting exposed and infectious population as well as an intervention that reduces the rate of infection by a factor of q at $t = t_q$ causing the case count to lower significantly (See Table 3.3). This setup allows us to examine the other type of switching possible: Gaussian approximation to PF.

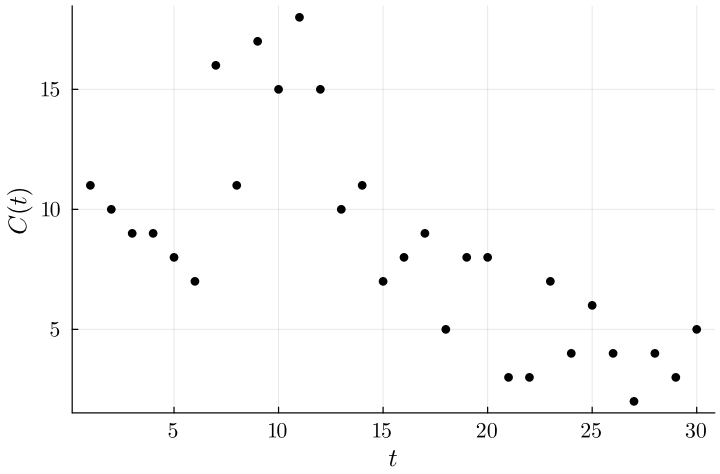


Figure 3.7: Plot of observed cases produced from $SE^{(2)}I^{(2)}R$ simulation used as the data for this experiment.

As before, we begin with the filtering distributions. It appears that for both early in the time series (when case counts are high) and late in the time series (when case counts are lower) the PF is well approximated using Gaussian approximation regardless of threshold. This then translates to similar results for parameter estimation across all case thresholds (see Table 3.4, Figure 3.9 and Figure 3.10). Figure 3.11 compares the computational time and likelihood variance for both methods making use of the same method, as Figure 3.6. Given the highly accurate parameter estimates and superior performance, these results suggest a use case for setting the threshold to 0 and using the Gaussian approximation for the entire time series.

It is clear from this experiment that the Gaussian approximation is far more accurate for this experiment compared with the previous experiment. One reason for this discrepancy is the larger overall case count, which makes the Gaussian approximation perform better. Additionally, starting from a high agent count and moving to a low agent means that the largest inaccuracies do not occur at the beginning and thus, do not propagate throughout the entire time series. Consequently, the systematic underestimation of the filtering distribution means are far less pronounced.

Event	State Update	BP Rate
Infection (1)	$(1,0,0,0,0)$	$q_t\beta\mathbf{z}_1$
Infection (2)	$(1,0,0,0,0)$	$q_t\beta\mathbf{z}_2$
$\text{Exposed}_1 \to \text{Exposed}_2$	$(-1,1,0,0,0)$	$2\delta\mathbf{z}_1$
Symptoms onset (observed)	$(0,-1,1,0,1)$	$2p\delta\mathbf{z}_1$
Symptoms onset (unobserved)	$(0,-1,1,0,0)$	$2(1-p)\delta\mathbf{z}_2$
$\text{Infectious}_1 \to \text{Infectious}_2$	$(0,0-1,1,0)$	$2\lambda\mathbf{z}_3$
Recovery	$(0,0,0,-1,0)$	$2\lambda\mathbf{z}_4$

Table 3.3: Table of events and rates for the $\text{SE}^{(2)}\text{I}^{(2)}\text{R}$ branching process approximation, where $q_t = 1$ for $t < t_q$ and $q_t = q$ otherwise.

Parameter	True value	Prior	Formulation
Basic reproduction number (R_0)	3	Gamma$(2.5, 0.5)$	$R_0 = \beta/\lambda$
Intervention effect (q)	0.1	Beta$(1.1, 1.5)$	-
Intervention day (t_q)	10	Discrete Uniform$(1, 30)$	-
Mean latency period (T_E)	6	Gamma$(60, 10)$	$T_E = 1/\delta$
Mean infectious period (T_I)	12	Gamma$(26, 2)$	$T_I = 1/\lambda$
Immigration rate ($\boldsymbol{\alpha}$)	$(0,0,0,0)^T$	fixed	-
Observation probability (p)	0.8	fixed	-
Observation variance (σ^2)	1	fixed	-
Initial State ($\mathbf{z}_0$)	$(30,30,30,30,0)^T$	fixed	-

Table 3.4: Parameters and priors used for SEIR simulation study.

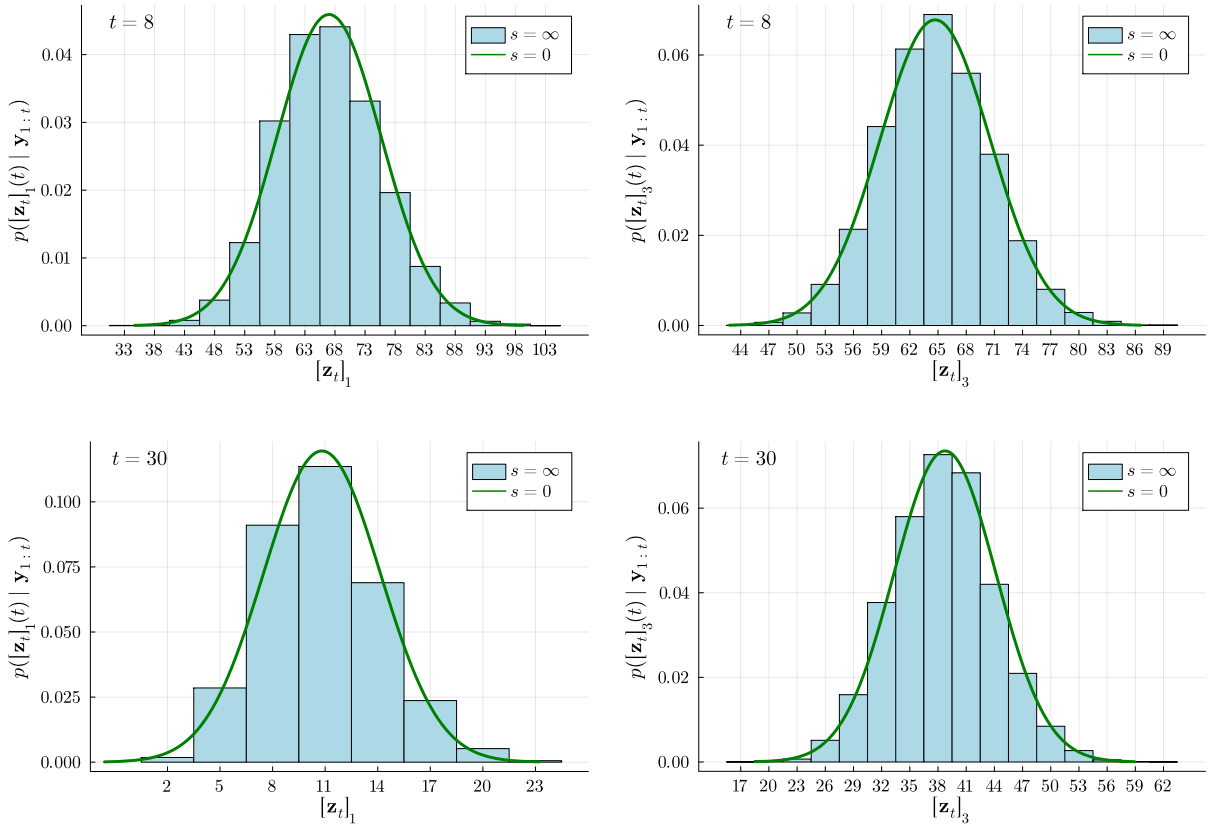


Figure 3.8: State density estimates of $\mathbf{z}_1(t) \mid \mathbf{y}_{1:t}$ and $\mathbf{z}_3(t) \mid \mathbf{y}_{1:t}$ for $t = 8, 30$ using PF, Gaussian approximation ($s = 0$) methods, using Equation 3.29 as the switching criteria. Even the $s = 0$ method accurately estimates the filtering distributions across both time steps. We have not shown any other thresholds as they all give very similar estimates, as expected.

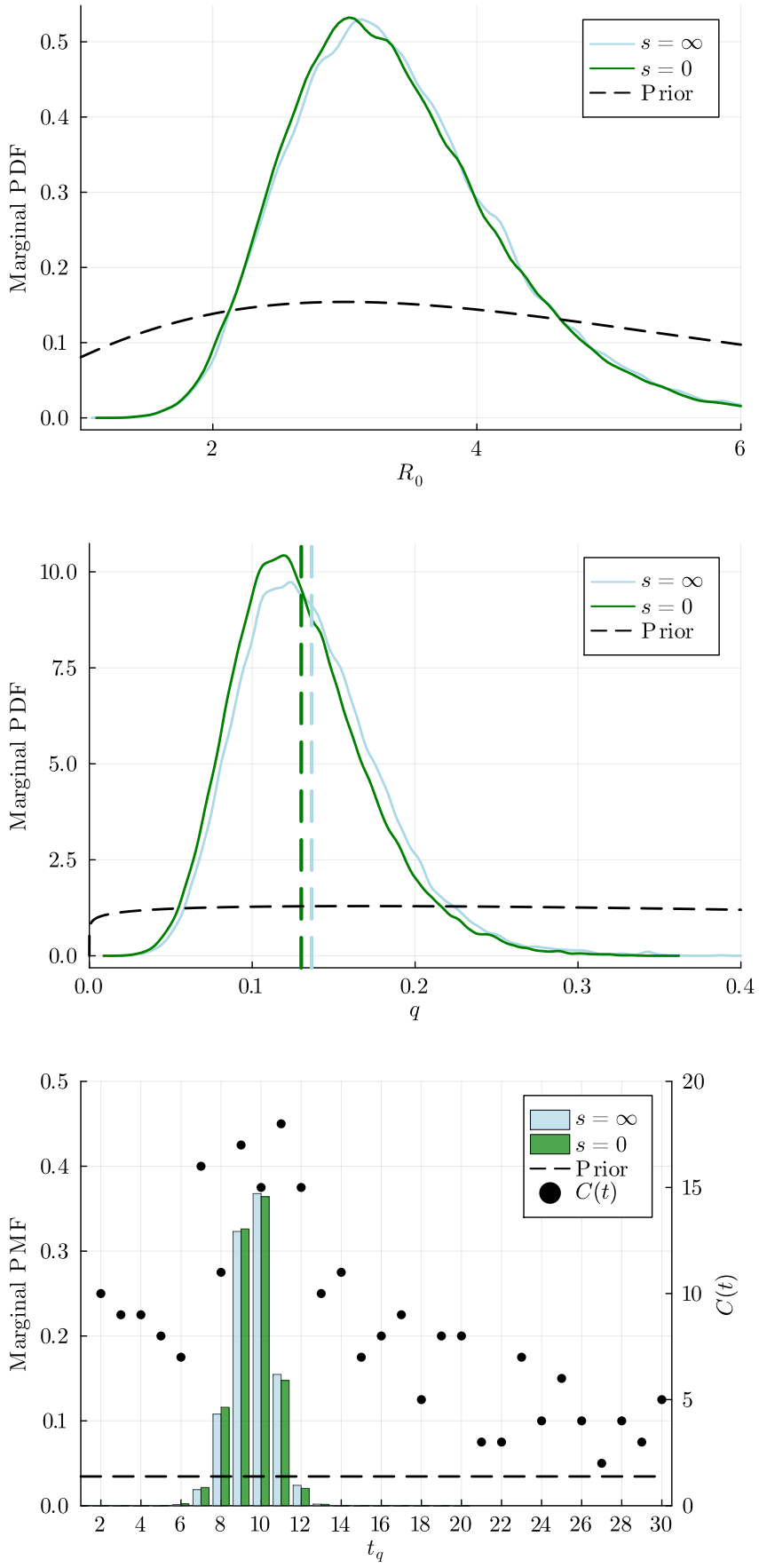


Figure 3.9: Marginal posteriors for R_0, q and t_q, using different thresholds.

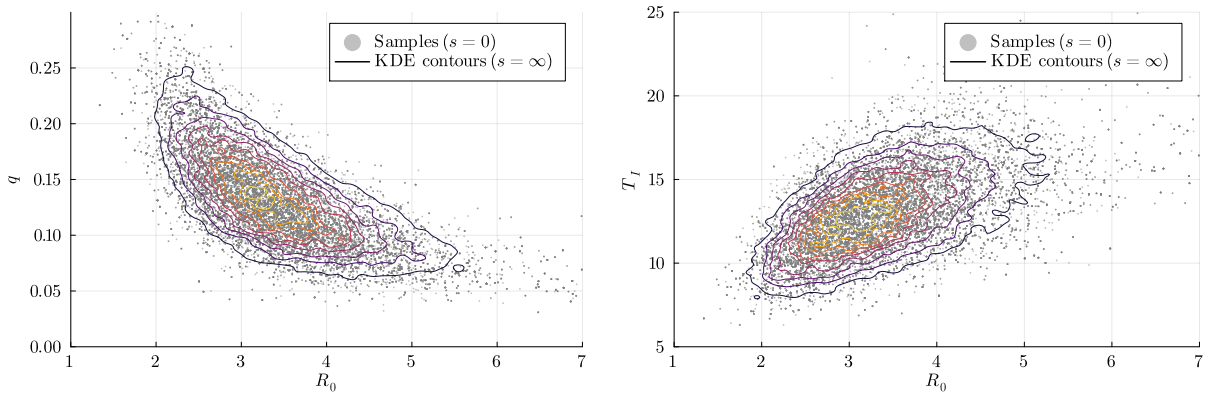


Figure 3.10: Comparisons of the joint posterior $p(R_0, q|\mathbf{y}_{1:T})$ and $p(R_0, T_I|\mathbf{y}_{1:T})$ produced by the PF and the Gaussian appoximation ($s = 0$). The PF posteriors are represented by the contours of the kernel density estimate (KDE), while the $s = 0$ switching threshold estimate is shown as samples.

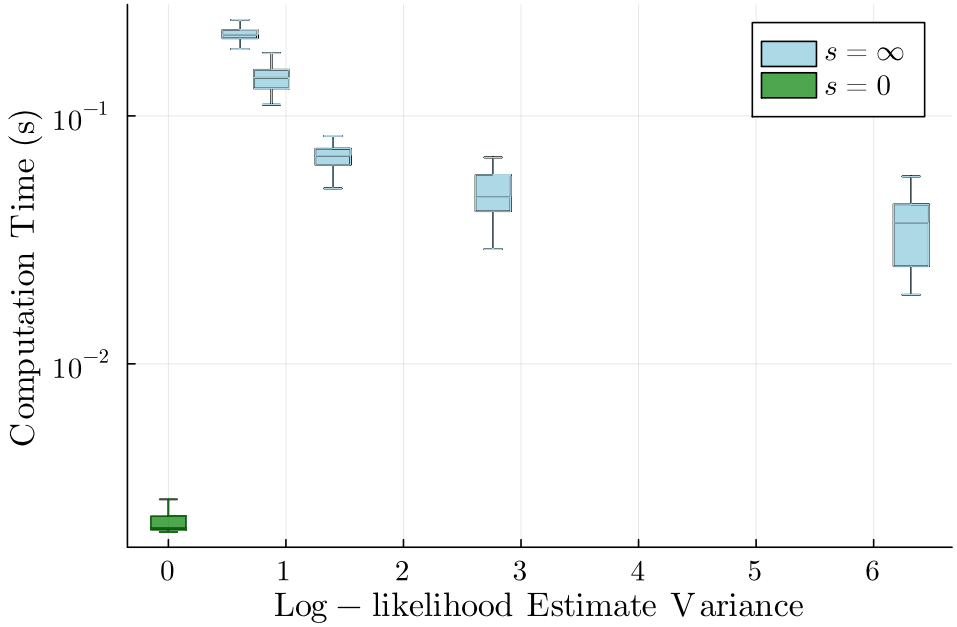


Figure 3.11: Box plot comparing the computational time and variance of the log-likelihood estimates produced by the Gaussian approximation at different thresholds.

3.3.3 Victoria Second COVID-19 Wave

For our final test, we apply the $s = 0$ switching threshold to a significantly more complex and computationally challenging problem. We will use 100 days of real data from the

second COVID-19 wave in Victoria, Australia [75]. During this wave daily case counts exceeded 600 people meaning the PF would be prohibitively slow. The length of the time series means switching between the Gaussian approximation and PF produces high variance log-likelihood estimates, even at low switching thresholds. Adding to this, we assume a time dependant reproduction number changing every 10 days as a Gaussian random walk [76, 77]. The primary purpose of this section is not to give novel insight into the basic reproduction number of COVID in Victoria at this time, but rather to illustrate that this method allows us to fit a relatively complex branching process model to messy real-world data. Given the computational challenges associated with this type of estimation, it would be infeasible with current methods [36, 41].

We treat each R_t as a parameter to be estimated, with a gamma prior on R_1, and a random walk structure enforced by the prior for $R_2, \dots R_{10}$

$$p(R_1, R_2, \dots, R_{10}) = \text{Gamma}(R_1; 1.1, 1) \prod_{t=2}^{10} \phi(R_t; R_{t-1}, 0.25). \tag{3.31}$$

We use the same $\text{SE}^{(2)}\text{I}^{(2)}\text{R}$ model as in the previous tests, but we instead allow for the rate of infection to change every 10 days (see Table 3.5). Details for other parameters can be found in Table 3.6. From Figure 3.4 we see the Gaussian approximation was successful in estimating the posterior distributions of each of the time-varying reproduction numbers. As expected, we see the estimates for R_t drop below 1 as the daily case counts begin to fall. Additionally, we see that earlier R_t estimates have far more variance than later estimates. This occurs as there are fewer infectious individuals at this time, meaning there are fewer opportunities to observe the infection rate of the disease. Moreover, towards the end of the time series, we observe a slight widening of the posterior. This effect is caused by a combination of the lowered case count and that many of the infections that occur over that period are yet to be observed due to the delay between infection and case observation.

The procedure took $\sim$ 14 hours to complete, on 3 cores. The minimum ESS was $\sim$ 10000, after removing 50,000 burn-in samples[2]. The convergence of the chain was validated with the $\hat{R}$ diagnostic [57].

To validate the model we sample parameters from $p(\boldsymbol{\theta} \mid y_{1:T})$ and then use them to simulate daily case counts. Figure 3.13 compares simulated daily case counts to the real data. This comparison shows reasonable adherence of the model and parameter estimates to the real data.

[2] All code was written in Julia v1.8 and the processor used was the Intel(R) Core(TM) i7-8650U CPU.

Event	State Update	BP Rate
Infection (1)	$(1,0,0,0,0)$	$\beta_{\lceil t/10 \rceil}\mathbf{z}_1$
Infection (2)	$(1,0,0,0,0)$	$\beta_{\lceil t/10 \rceil}\mathbf{z}_2$
$\text{Exposed}_1 \to \text{Exposed}_2$	$(-1,1,0,0,0)$	$2\delta\mathbf{z}_1$
Symptoms onset (observed)	$(0,-1,1,0,1)$	$p\delta\mathbf{z}_1$
Symptoms onset (unobserved)	$(0,-1,1,0,0)$	$2(1-p)\delta\mathbf{z}_2$
$\text{Infectious}_1 \to \text{Infectious}_2$	$(0,0-1,1,0)$	$2\lambda\mathbf{z}_3$
Recovery	$(0,0,0,-1,0)$	$2\lambda\mathbf{z}_4$

Table 3.5: Table of events and rates for the model of the COVID-19 outbreak in Victoria.

Parameter	Prior	Standard Formulation
Mean latency period (T_E)	Gamma$(5,1)$	$T_E = 1/\delta$
Mean infectious period (T_I)	Gamma$(4,0.5)$	$T_E = 1/\lambda$
Observation probability (p)	fixed $p = 0.8$	-
Observation variance (σ^2)	fixed $\sigma^2 = 1$	-
Initial State ($\mathbf{z}_0$)	fixed $\mathbf{z}_0 = (10,0,0,0)^T$	-

Table 3.6: Parameters and priors used for the analysis of the COVID-19 outbreak in Victoria.

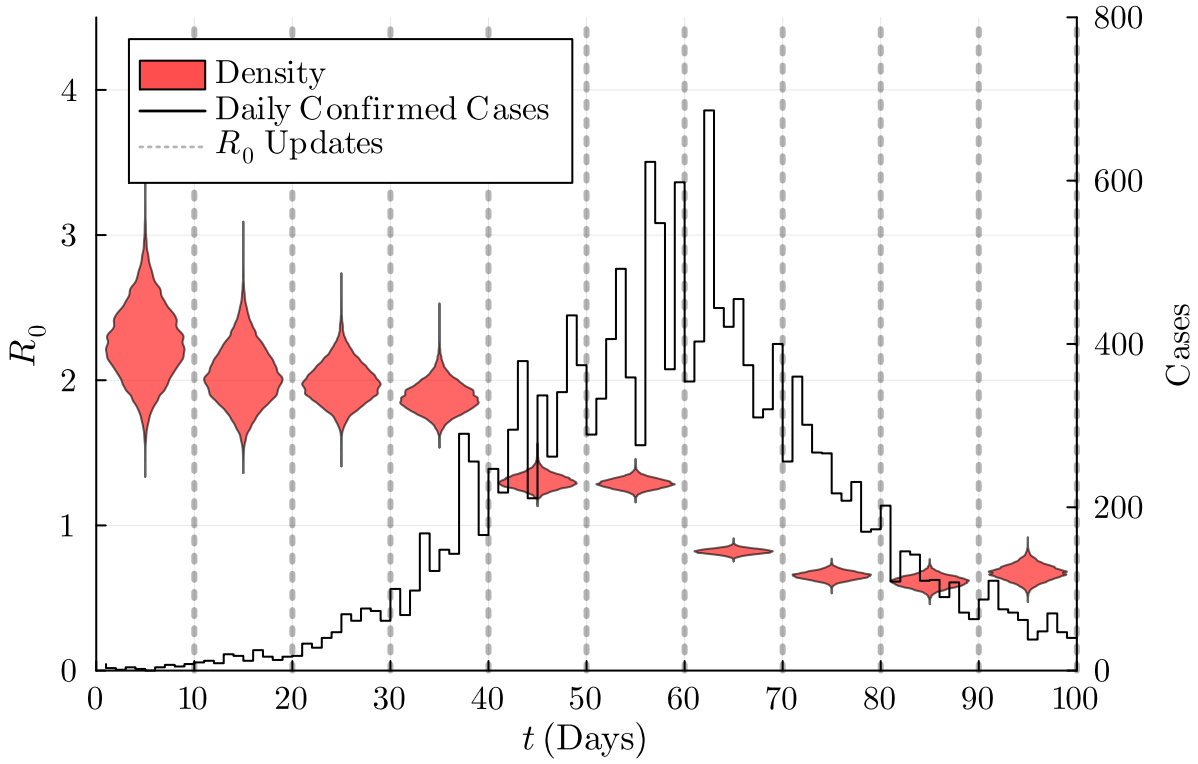


Figure 3.12: Violin Plot showing the posterior densities for time-varying reproduction numbers as well as the daily case counts from the second COVID-19 wave in Victoria [75].

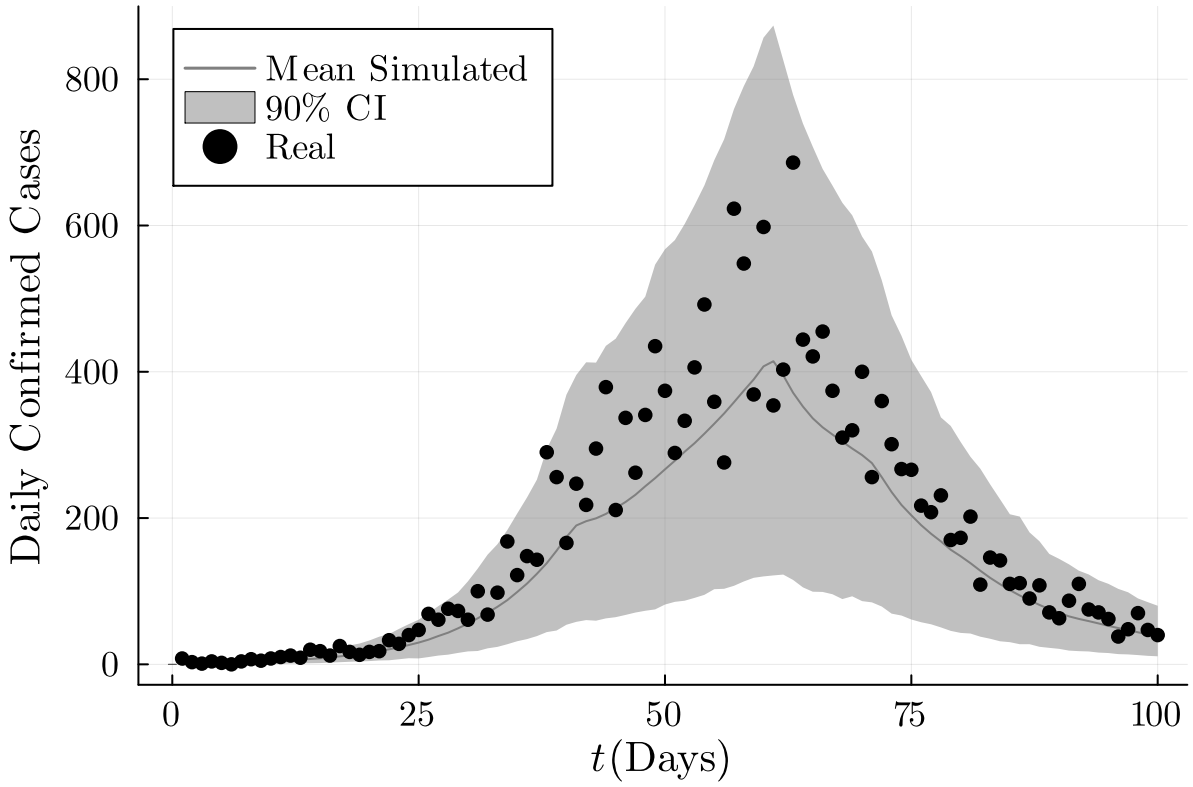


Figure 3.13: 90% credible intervals for the posterior predictive distribution and the real data.

Chapter 4

Metapopulation Model

In this chapter, we extend our work in Chapter 3 to a metapopulation model. A metapopulation model partitions agents into interacting subpopulations. Our focus is modelling the early phases of a novel disease outbreak where our subpopulations represent countries/states that interact through immigration. Building on the work of Shearer et al. [9] our primary goal is the development of a framework for estimating the risk of disease importation from different countries. We will extend the Gaussian filtering method from Chapter 3 to a metapopulation epidemic model. Driven by the computational requirements of large populations, we focus on Gaussian assumptions that enable predictions on global scales. Our array of modelling assumptions are validated in Section 4.2. Moreover, we show the utility of a hierarchical model for the subpopulation R_0 parameters, testing it against a non-hierarchical model. A hierarchical model is well suited to this problem since we expect R_0 parameters to be correlated among different subpopulations. Given the fundamental goal of the thesis is to analyse the outbreak of a *novel* disease we choose to maintain the functionality of our method for an arbitrary continuous-time branching process, which allows for the modelling of a variety of disease dynamics. Different diseases will have differing latent and generation time *distributions* as well as parameters, thus confining ourselves to a particular compartmental model will not suffice.

4.1 Methods

Internal Dynamics

Consider a partition of the population into S subpopulations. The epidemic dynamics of each subpopulation j will be modelled by a multitype continuous-time branching process with r types

$$\mathbf{z}_{j,t} = (z_{1,j,t}, z_{2,j,t}, \dots z_{r,j,t})^T.$$

here these r types represent the agent count for some compartmental epidemic model. This formulation implicitly assumes that subpopulations can only interact through immigration. The parameters of the continuous-time branching process $\left\{\boldsymbol{\theta}_i^{\mathbf{H}}\right\}_{i=1}^{S}$ can differ across subpopulations allowing diversity in important properties such as the rate of infection and recovery. The parameters for the internal dynamics of all the subpopulations are written as

$$\boldsymbol{\theta}^{\mathbf{H}} := (\boldsymbol{\theta}_1^{\mathbf{H}}, \boldsymbol{\theta}_2^{\mathbf{H}}, \ldots, \boldsymbol{\theta}_S^{\mathbf{H}}). \tag{4.1}$$

The data for each subpopulation j is generated by some state-dependant process as described in Chapter 3

$$\begin{aligned} \mathbf{y}_{j,t} &:= (y_{1,j,t}(t), y_{2,j,t}(t), \ldots y_{d,j,t}(t))^T, && (4.2)\\ \mathbf{y}_{j,t} &\sim \mathcal{N}(\mathbf{y}_{j,t}; \mathbf{H}\mathbf{z}_{j,t}, \mathbf{R}_j) && (4.3) \end{aligned}$$

here $\mathbf{y}_{j,t}$ is a column vector of size d which represents the data for subpopulation j at time t. $\mathbf{H}$ is a $d \times r$ matrix, that describes which part of the state space is observed. $\mathbf{R}_j$ represents any extra noise in the observations not accounted for by the hidden process. As discussed in Chapter 3 we can model some of the observation noise as part of the hidden process, allowing for state dependant observation noise. With this, we can write the totality of our parameters as

$$\boldsymbol{\theta} = (\boldsymbol{\theta}^{\mathbf{H}}, \mathbf{R}_1, \ldots, \mathbf{R}_S). \tag{4.4}$$

In the context of an epidemic model $\mathbf{z}_{j,t}$ would represent the underlying *true* process of the disease, this being the numbers of agents that are exposed, infected, asymptomatic, ect[1]., in subpopulation j and at time t. While $\mathbf{y}_{j,t}$ would represent the data generated by this underlying process, in particular, statistics like reported cases, recoveries and deaths for subpopulation j and time step t.

4.1.1 Immigration Network

For our purpose, we will define our subpopulations as countries/states, as this aligns reasonably well with the assumptions of our model:

- Individuals from one subpopulation cannot infect those from another subpopulation.
- Data is available that can give an estimate as to the volume of movement between these subpopulations.

[1] In some cases this state may also keep track the total number of events that have occurred often the total number of deaths/infections.

For a given $\mathbf{z}_{j,t}$ there is a subset of its elements that represent individuals that are capable of movement between subpopulations, we call this subset $\mathcal{I}$. Following Zhang et al. [78] & De Salazar et al. [10] we model the immigration of infected agents of type $i \in \mathcal{I}$ between subpopulations as a time inhomogeneous Poisson process. We assume further that travel is equivalent to randomly sampling $T_{j,k}$ individuals from subpopulation j and duplicating the infected ones in subpopulation k, where $T_{j,k}$ is the estimated travel volume moving from $j \to k$. As a result, when a subpopulation receives an immigrant there is no corresponding removal from another subpopulation. As we will see later this approximation is necessary for estimating the transition densities of each subpopulation independently. Given that the number of infected agents will always be significantly larger than the outward flux of agents this approximation is unlikely to have a significant effect.

With this setup, we can calculate the rate of immigration of type i agents from subpopulation j to subpopulation k at time t is

$$\alpha_{i,j,k,t} = \begin{cases} \frac{T_{j,k}}{N_j} \cdot z_{i,j,t} & i \in \mathcal{I} \\ 0 & i \notin \mathcal{I} \end{cases}, \tag{4.5}$$

$$\boldsymbol{\alpha}_{j,k,t} := (\alpha_{1,j,k,t}, \alpha_{2,j,k,t}, \ldots, \alpha_{r,j,k,t}) \tag{4.6}$$

where N_j is the population of subpopulation j. Thus, the overall rate of immigration into subpopulation k is

$$\boldsymbol{\alpha}_{k,t} = \sum_{j=1}^{S} \boldsymbol{\alpha}_{j,k,t} \tag{4.7}$$

As we will see later it is vital that we can calculate each subpopulation's transmission at a given time step independently of other subpopulations' transmission during that time step. To achieve this while including immigration effects we assume immigration rates at $t + \tau$ for $\tau \in [0, 1)$ are based on the agent count in each subpopulation at time t, making the immigration rates constant over the interval $[t, t + 1)$. Here we are approximating an inhomogeneous Poisson process with a piece-wise homogeneous one,

$$\boldsymbol{\alpha}_j(t + \tau) \approx \boldsymbol{\alpha}_j(t) \quad \tau \in [0, 1).$$

Effectively this means that $\mathbf{z}_{j,t+\tau}$ and $\mathbf{z}_{k,t+\tau}$ are conditionally independent given the agent counts for all the subpopulations at the beginning of the time step $\mathbf{z}_t$. We expect generally that the number of agents in each subpopulation will increase in the interval $[t + 1)$, thus increasing the rate of immigration. Consequently, this approximation will slightly underestimate the rate of immigration most of the time, but we expect this effect to be small as long as the length of the time step is small in comparison with the growth rate of the disease.

The hidden Markov model for each subpopulation is then

$$\mathbf{z}_j(0) \sim p(\mathbf{z}_j(0)|\boldsymbol{\theta}) \tag{4.8}$$
$$\mathbf{z}_{j,t+1} \sim p(\mathbf{z}_{j,t+1}|\mathbf{z}_{j,t}, \boldsymbol{\alpha}_{j,t}, \boldsymbol{\theta}), \tag{4.9}$$
$$\mathbf{y}_{i,t} \sim p(\mathbf{y}_{i,t}|\mathbf{z}_{j,t}, \boldsymbol{\theta}) \tag{4.10}$$

for $j = 1, 2, \ldots, S$ and $t = 1, 2, \ldots T$.

4.1.2 Problem Outline

Our fundamental goal is to produce a method that can jointly estimate states and parameters for partially observed interacting CTBPs across many separate subpopulations. Practically this translates to using daily case data for a disease from different countries to jointly estimate both the number of infected individuals and the parameters mediating the spread of the disease in these countries. Mathematically our estimates are encapsulated by the full joint posterior distribution $p(\mathbf{z}_{1:T}, \boldsymbol{\theta}|\mathbf{y}_{1:T})$, where $\mathbf{y}_{1:T}$ is the daily case counts for all the countries and $\mathbf{z}_{1:T}$ is the number of infectious/exposed individuals in each country. As discussed in Chapter 3, using pmMH with a bootstrap particle filter to estimate $p(\mathbf{z}_{1:T}, \boldsymbol{\theta}|\mathbf{y}_{1:T})$ can be computationally expensive for a single subpopulation. Extending the problem to many subpopulations increases the computational cost in 3 distinct ways:

- The cost of sample path simulation increases linearly with the number of subpopulations [40].
- The variance of the likelihood estimate will increase, requiring more bootstrap samples [74, 61].
- The number of parameters that need to be estimated increases with the number of subpopulations. This increase in the dimension of the parameter space can cause an increase in the required burn-in time for the pmMH Markov chain and the time required to fully explore the sample space [57].

Taken together these issues suggest that estimation on a reasonable scale will require a more efficient method. To remedy this situation we apply the Gaussian approximation of the bootstrap particle filter derived in Chapter 3 to the muli-population case. Given the large computational challenge, we choose to use the Gaussian approximation throughout the entire time series, rather than switching between it and the PF.

4.1.3 Application Of Gaussian Approximation

Applying the Gaussian approximation to the multi-population model is non-trivial, as each subpopulation interacts through immigration. This interaction prevents simply applying the method detailed in Chapter 3 to each subpopulation separately, as we cannot

decompose the full transition density $p(\mathbf{z}_{t+1}|\mathbf{z}_t)$ into the transition density for the individual subpopulations $\prod_{i=1}^{S} p(\mathbf{z}_{i,t+1}|\mathbf{z}_{i,t})$.

In this section, our goal is to derive a reasonable approximation of $p(\mathbf{z}_{t+1}|\mathbf{z}_t)$ that can be decomposed into transition densities that the Gaussian approximation can be applied to. Firstly, since immigration rates remain constant throughout $[t, t+1)$ the state of a given subpopulation at $t+1$ is independent of all the others given the state of all the subpopulations at t thus

$$p(\mathbf{z}_{t+1}|\mathbf{z}_t) = \prod_{i=1}^{S} p(\mathbf{z}_{i,t+1}|\mathbf{z}_{i,t}, \mathbf{z}_{-i,t}), \tag{4.11}$$

where $\mathbf{z}_{-i,t}$ is the state of all other subpopulations apart from i. The dependence on $\mathbf{z}_{-i,t}$ still prevents us from applying the Gaussian approximation: we would have to integrate over all the possible states of the rest of the subpopulations since the immigration rate into subpopulation i is a function of $\mathbf{z}_{-i,t}$. More clearly, the problem arises because we need a known immigration rate to derive the mean (3.10) and variance (3.11) of the transition density. We choose to replace $\mathbf{z}_{-i,t}$ with a fixed value as opposed to a random variable, in order to apply the approximation. The value we choose is the mean estimate of $\mathbf{z}_{t,-i}$ given the data upto time t, written as $\boldsymbol{\mu}_{-i,t|t}$

$$\boldsymbol{\alpha}_{i,t} = \sum_{j \in -i} \frac{T_{j,i}}{N_j} \cdot [\mathbf{z}_{j,t}]_{\mathcal{I}} \approx \sum_{j \in -i} \frac{T_{j,k}}{N_j} \cdot [\boldsymbol{\mu}_{j,t|t}]_{\mathcal{I}}, \tag{4.12}$$

where $[\boldsymbol{\mu}_{j,t|t}]_{\mathcal{I}}$ denotes the elements of $\boldsymbol{\mu}_{j,t|t}$ that correspond to immigrating agent types. Application of this approximation relinquishes the problematic dependence on the states of the other subpopulations, weakly decoupling them

$$p(\mathbf{z}_{t+1}|\mathbf{z}_t) \approx \prod_{i=1}^{S} p(\mathbf{z}_{i,t+1}|\mathbf{z}_{i,t}, \boldsymbol{\mu}_{j,t|t}). \tag{4.13}$$

We say *weakly* here because they are conditionally independent given the mean of the filtering distribution, meaning that one subpopulation having a high case count will still lead to an increased immigration rate for the other subpopulations. As a result, this weak decoupling allows our Gaussian approximation to be used while still capturing the immigration dynamics.

Figure 4.1 illustrates the salient difference between the Gaussian approximation and the bootstrap particle filter. Since the bootstrap particle filter gives distributionally accurate filtering distributions, comparing the two can help in understanding the effect of assuming (4.13). Consider the grey path in subpopulation i, from Figure 4.1, and note

that the immigration into subpopulation i is calculated based on the grey path in subpopulation j. Since the grey path has the largest $z_{1,i,t}$, in subpopulation i it is more likely to be associated with the higher $z_{1,j,t}$ paths in subpopulation j (which it is in our illustration). However, this is not true for the Gaussian approximation as immigration into subpopulation i is simply based on the mean value of subpopulation j (shown as a grey dashed line). Thus, the higher estimates for $z_{1,i,t}$ in the Gaussian approximation will have a lower incoming immigration rate than it should. This occurs, in the same way with the lower estimates for $z_{1,i,t}$, but now with higher immigration than would be correct. This property has the effect of reducing the overall variance of the state estimates. In the context of an epidemic model, immigration rates become insignificant in comparison to local transmission as case counts increase. Since higher case counts are significantly more informative than lower case counts we suspect that the effect of this error will be insignificant.

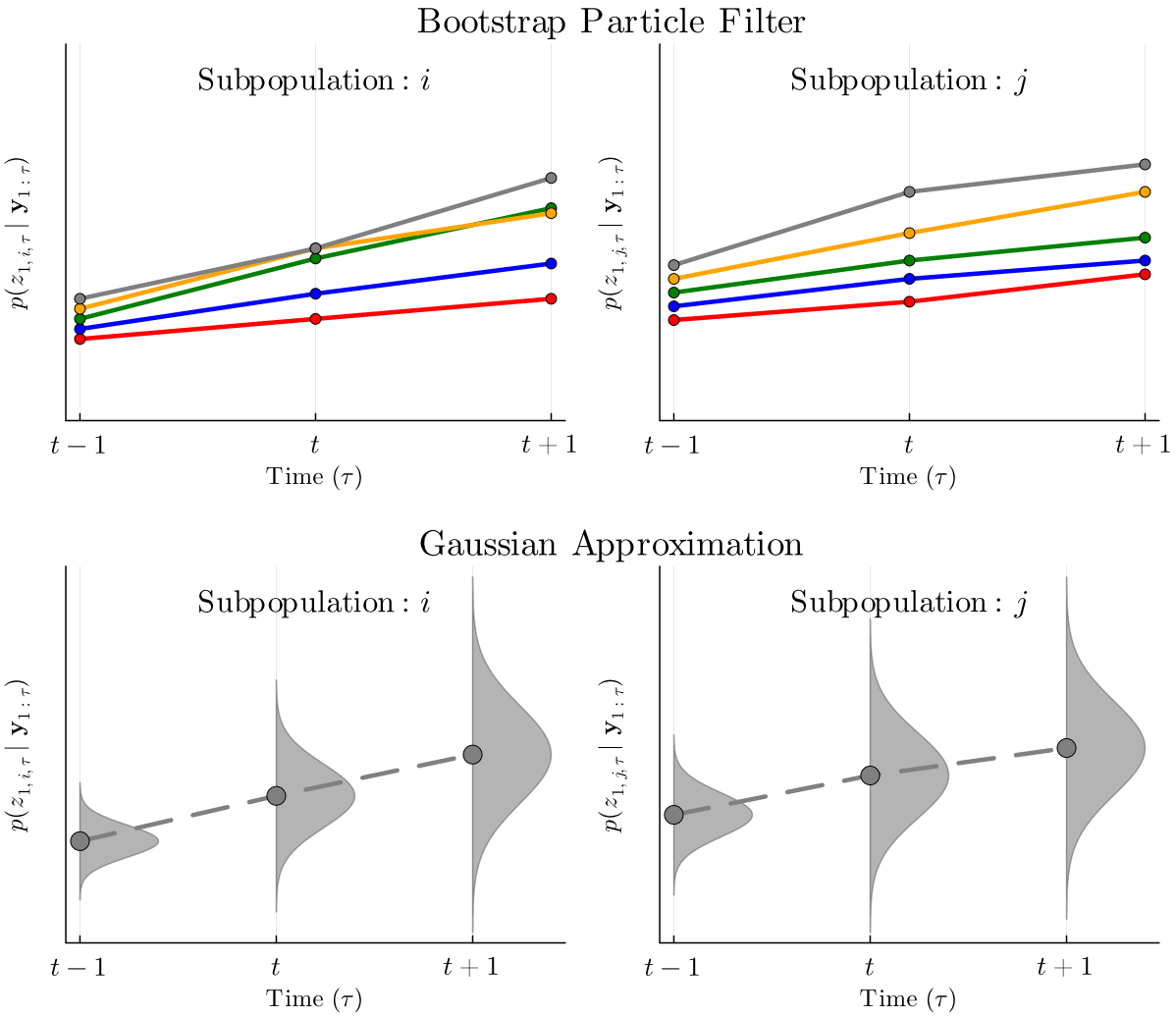


Figure 4.1: Illustration of the difference between the bootstrap particle and our Gaussian approximation. The colours on the bootstrap particle filter denote the corresponding sample paths which make up the particle. We observe that the order of the colours is similar in both figures since sample paths tend to be correlated. This effect is not captured by the Gaussian approximation, but we show in Section 4.2.2 that the loss of this correlation information is insignificant as long the immigration rates are small.

4.1.4 Algorithm Summary

Making use of the single population method described in Chapter 3 and the approximations developed in this chapter we can now fully describe the method. The inputs are:

- Observations $\mathbf{y}_{1:T}$.
- Parameters for all of the subpopulations $\boldsymbol{\theta}$ (See (4.4)).
- Observation matrix $\mathbf{H}$.
- Population sizes $\{N_i\}_{i=1}^S$.
- Travel volumes $\{T_{i,j}\}_{j=1}^S$ for $i \in \{1, 2, \ldots, S\}$.
- The intial mean $\boldsymbol{\mu}_{i,0|0}$ and variance $\boldsymbol{\Sigma}_{i,0|0}$ of $p(\mathbf{z}_{i,0})$ for all $i \in \{1, 2, \ldots, S\}$.

To initialise the algorithm we must use the parameters for each subpopulation to derive their specific generator matrix (see Section 2.2). We write the generator matrix corresponding to subpopulation i as $\boldsymbol{\Omega}_i$. Then using $\boldsymbol{\Omega}_i$ we can calculate the mean matrix $\mathbf{F}_i$ (2.6), the mean immigration matrix $\mathbf{B}_i$ (2.7), the variance matrices $\{\mathbf{V}_{j,i}\}_{j=1}^d$ (2.9) and the immigration variance matrices $\{\mathbf{W}_{j,i}\}_{j=1}^d$ (2.10) (see Section 2.2.1 for calculation). Finally, initialise the log-likelihood $\ell \to 0$. The following steps are then repeated for $t \in \{1, 2, \ldots, T\}$:

1. Calculate immigration rates for all subpopulations $i \in \{1, 2, \ldots, S\}$

$$\boldsymbol{\alpha}_{i,t} \leftarrow \sum_{j \in -i} \frac{T_{j,k}}{N_j} \cdot [\boldsymbol{\mu}_{j,t|t}]_{\mathcal{I}}.$$

2. Using the immigration rates calculate the mean and variance of the predictive distributions for all subpopulations $i \in \{1, 2, \ldots, S\}$

$$\begin{aligned} \boldsymbol{\mu}_{i,t+1|t} &\leftarrow \mathbf{F}_i \boldsymbol{\mu}_{i,t|t} + \mathbf{B}_i \boldsymbol{\alpha}_{i,t}, \\ \boldsymbol{\Sigma}_{i,t+1|t} &\leftarrow \mathbf{F}_i \boldsymbol{\Sigma}_{i,t|t} \mathbf{F}_i^T + \sum_{j=1}^r [\boldsymbol{\mu}_{i,t|t}]_j \mathbf{V}_{j,i} + \sum_{j=1}^r [\boldsymbol{\alpha}_{i,t}]_j \mathbf{W}_{j,i}. \end{aligned}$$

3. Calculate the mean and variance of the filtering distributions for all subpopulations $i \in \{1, 2, \ldots, S\}$

$$\begin{aligned} \mathbf{K}_i &\leftarrow \boldsymbol{\Sigma}_{i,t+1|t} \mathbf{H} (\mathbf{H} \boldsymbol{\Sigma}_{i,t+1|t} \mathbf{H}^T + \mathbf{R}_i)^{-1}, \\ \boldsymbol{\mu}_{i,t+1|t+1} &\leftarrow \boldsymbol{\mu}_{i,t+1|t} + \mathbf{K}_i (\mathbf{y}_{t+1} - \mathbf{H} \boldsymbol{\mu}_{i,t+1|t}), \\ \boldsymbol{\Sigma}_{i,t+1|t+1} &\leftarrow (\mathbf{I} - \mathbf{K}_i \mathbf{H}) \boldsymbol{\Sigma}_{i,t+1|t}. \end{aligned}$$

4. Update the log-likelihood total

$$\ell \leftarrow \ell + \sum_{i=1}^S \log \left[\mathcal{N}(\mathbf{y}_{i,t+1}; \boldsymbol{\mu}_{i,t+1|t}, \boldsymbol{\Sigma}_{i,t+1|t}) \right].$$

The output of the procedure is the means $\left\{\boldsymbol{\mu}_{i,t|t}\right\}_{t=1}^T$ and variances $\left\{\boldsymbol{\Sigma}_{i,t|t}\right\}_{t=1}^T$ of the filtering distributions and the log-likelihood ℓ.

4.2 Validation of Gaussian Approximation

4.2.1 Initial Importation Time

Due to the exponential nature of disease outbreaks, small changes in the time of the initial infection can lead to vastly different infection numbers in the future. For our purposes, the initial infection for most subpopulations will come through immigration. This is problematic for our method since the Poisson immigration process is instead modelled with a Gaussian distribution. While the mean is accurate the skew of the Poisson distribution is not captured by the Gaussian. Additionally, there is the variance reduction issue discussed in Section 4.1.3.

To interrogate this issue we take the sample path generated in Section 3.3.1 and estimate immigration rates and the corresponding imported cases using both the bootstrap particle filter and the Gaussian approximation (GA). Figure 4.2 shows the PMF of the time of the first imported case using both the PF and GA. The GA has less variance, likely caused by using the mean to approximate the immigration rate. Moreover, the expected value is significantly greater, which we suspect is due to a combination of the mean approximation and the lack of skew in the Gaussian approximation.

These findings are problematic for the GA method as this may cause a significant underestimation of cases and an overestimation of R_0. However, since the GA approximation assumes a continuous state space the importance of the first case threshold may be mitigated. This motivates the experiment in the next section where we explore the accuracy of the GA for parameter estimation, testing if this discrepancy is important. It should be noted that this issue is not relevant when calculating importation risks. When importation risks are derived we do not use a Gaussian approximation, they are derived directly from CTBP theory and the immigration rate (see Section 4.3.3). Consequently, the parameter and state estimates are the key issues that need to be resolved in light of this discrepancy.

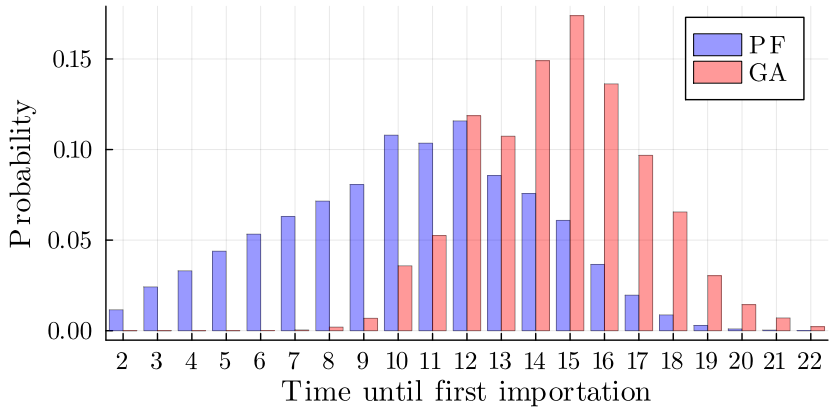


Figure 4.2: Plot of the PMF for the time to first importation using the PF and GA.

4.2.2 Parameter Estimation

In Section 4.1.3 we discussed some of the biases associated with applying the Gaussian approximation to metapopulation models with immigration. In this section, we investigate the severity of these biases in the context of an epidemic model via a simulation study. For simplicity, we make use of the standard SEIR branching process approximation (see Table 4.1). Our immigration model consists of 2 subpopulations with immigration going both ways between them (see Figure 4.3). We choose population sizes and travel volumes that give very exaggerated immigration rates, assuming 1% of the population travels every day. This setup is chosen to exacerbate the effect of the approximation described in Section 4.1.3. The effect is exacerbated because the impact of immigration on the trajectory of the outbreak is increased.

For the simulation study, we fix a set of parameters and simulate from the CTMC model using the SSA. Immigration is accounted for in the simulation using the procedure described in Section 4.1.1. We then take the observed cases from the simulation, which can be obtained from the simulated sample paths (see Section 3.3.1) and use this as the dataset for the simulation study (see Figure 4.4).

Table 4.2 describes the parameters used for the simulation as well as the priors used for estimating the parameters. We then use a standard bootstrap particle filter (PF) for likelihood estimation as part of a pmMH scheme to estimate all the non-fixed parameters (see Table 4.2). Once converged, the parameter samples produced will follow the correct posterior distribution [39]. We then repeat this process using the Gaussian approximation method (GA), described in the previous section. Recall that the PF will give samples from the true posterior meaning if the GA is working well the parameter estimates should be similar for both methods.

Figure 4.5 compares the parameter estimates produced by the PF and GA method. We see good adherence between the posteriors produced by the PF and GA. In particular, we do not see significantly narrower posteriors for the GA which is what we would expect if the variance of the filtering distributions were being significantly underestimated, due to the immigration approximation. Additionally, we do not observe higher R_0 estimates due to the discrepancies in initial case importation. This supports our assertion, described in Section 4.1.3, that immigration is only significant when the case count is very low or zero (when the data is the least informative). It should be noted that in this test we have avoided the issues with low agent populations as we have already explored this in Section 3.3.1. Moreover, for our specific application subpopulations with low agent counts (infected individual counts) are less important when estimating the importation risk.

Event	State Update	BP Rate
Infection	$(1, 0, 0)$	$\beta_j \mathbf{z}_2$
Symptoms Development (observed)	$(-1, 1, 1)$	$p_j \delta \mathbf{z}_1$
Symptoms Development (unobserved)	$(-1, 1, 0)$	$(1 - p_j)\delta \mathbf{z}_1$
Removal	$(0, 0, -1)$	$\lambda \mathbf{z}_2$
Immigration of exposed	$(1, 0, 0)$	$\boldsymbol{\alpha}_1$
Immigration of infectious	$(0, 1, 0)$	$\boldsymbol{\alpha}_2$

Table 4.1: Table of events and rates for the SEIR branching process approximation in subpopulation j.

Parameter	True value	Prior	Formulation
Basic reproduction number 1: $R_0^{(1)}$	2.5	$\mathcal{N}_{[0,10]}(3, 25)$	$R_0 = \beta_1/\lambda$
Basic reproduction number 2: $R_0^{(2)}$	3	$\mathcal{N}_{[0,10]}(3, 25)$	$R_0 = \beta_2/\lambda$
Mean latency period: T_E	3.6	$\mathcal{N}_{[0,10]}(3.6, 0.5)$	$T_E = 1/\delta$
Mean infectious period: T_I	5.2	Gamma$(5.2, 0.5)$	$T_I = 1/\lambda$
Observation probabilities : p_1, p_2	0.6	Beta$(24, 16)$	-
Observation variance: σ^2	1	fixed	-

Table 4.2: Parameters and priors used for SEIR simulation study.

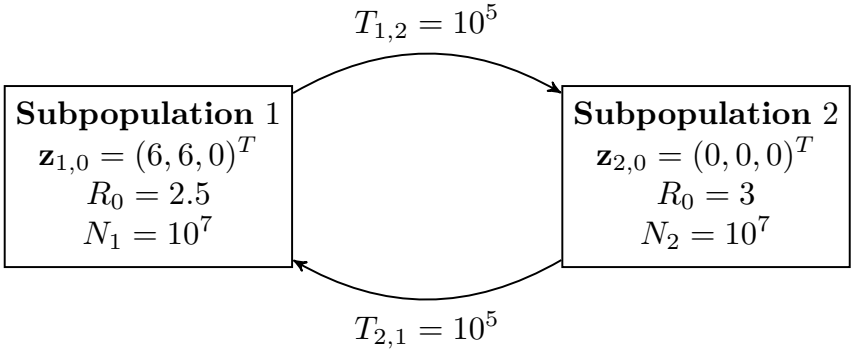


Figure 4.3: Immgration network diagram between 2 subpopulations. Arrows denote the allowed directions of immigration.

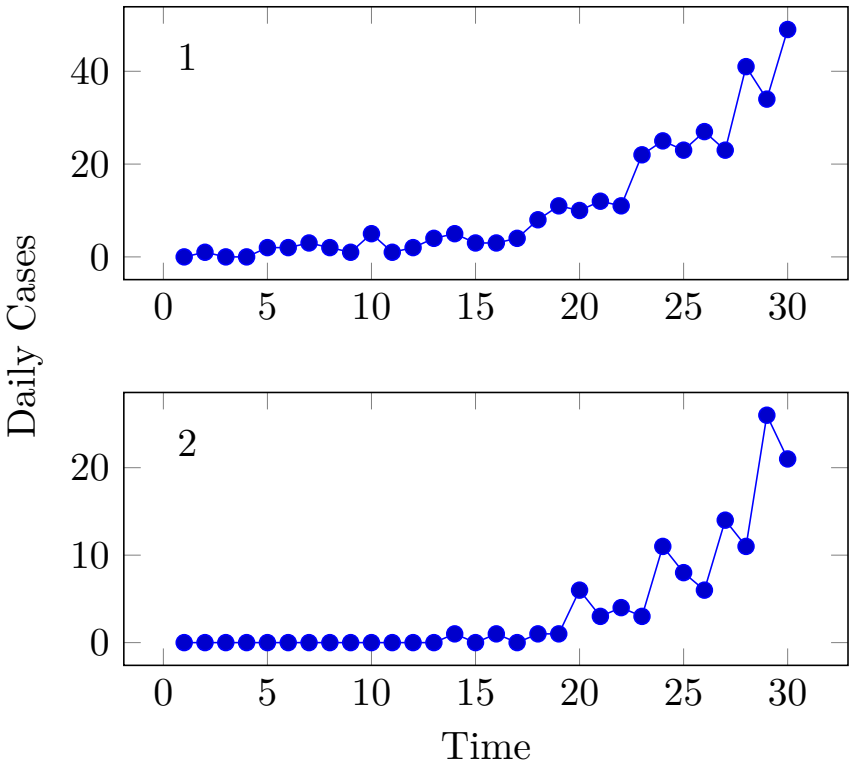


Figure 4.4: Simulated daily case counts for each subpopulation. The top left corner of each plot identifies the subpopulation.

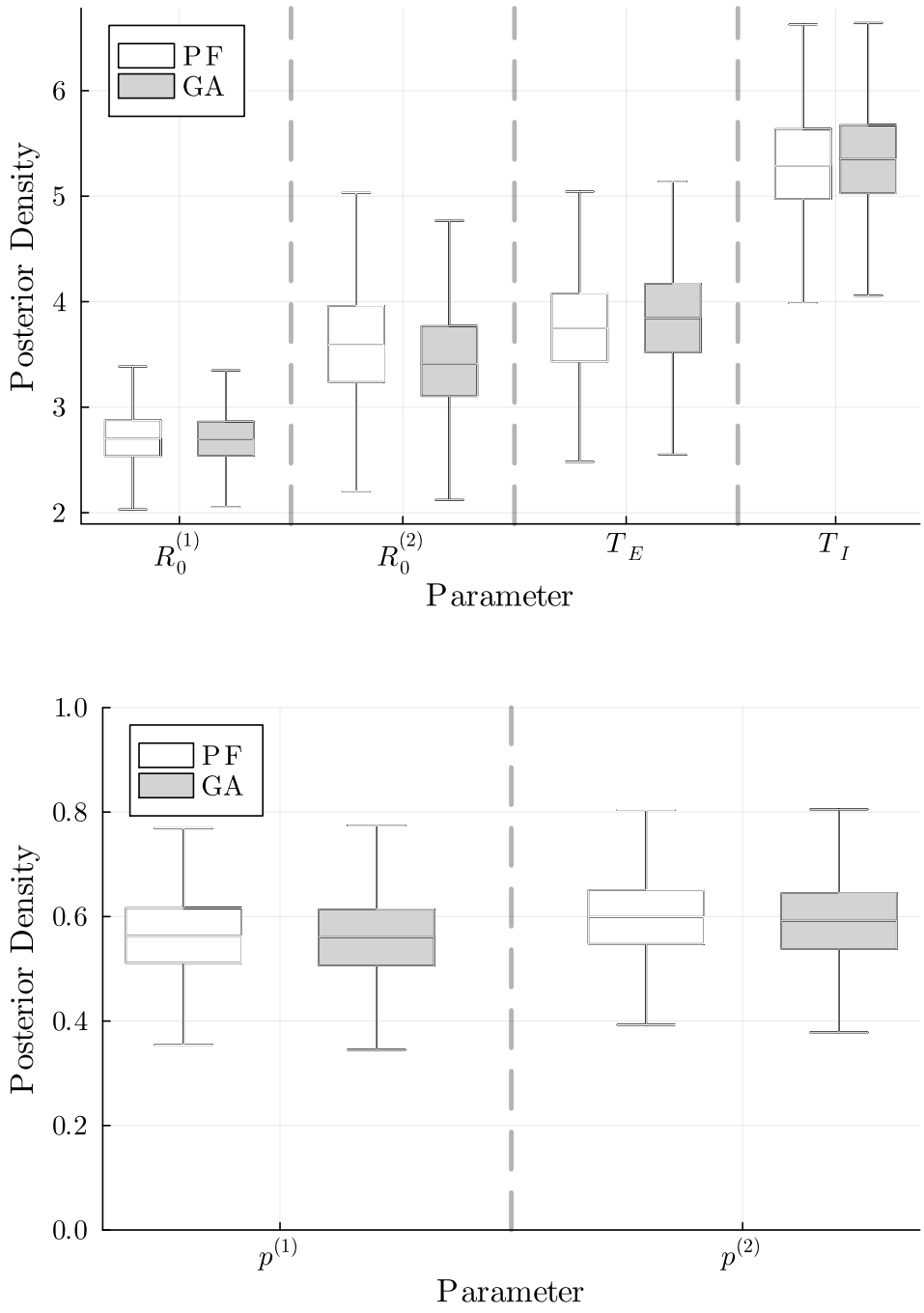


Figure 4.5: Boxplot comparison of the parameter estimates produced using the bootstrap particle filter (PF) and the Gaussian approximation (GA) for the disease (top) and observation (bottom) related parameters.

4.3 Feasibility For Large Model

In this section, our goal is to test the computational feasibility of the GA for a much larger immigration network. In particular, we choose to have 20 subpopulations, a similar size as the model formulated by Shearer et al. [9] for the early phases of the SARS-CoV-2 pandemic. Again following Shearer et al. [9] we structure the immigration network with a single *source* subpopulation (where the disease originates) 18 intermediary subpopulations and one *sink* subpopulation. Immigration occurs from the source to the intermediaries and from the intermediaries to the sink (see Figure 4.6).

The data for this test is simulated using the $\text{SE}^{(2)}\text{I}^{(2)}\text{R}$ branching process approximation model with an added exponentially distributed delay between symptom onset and observation. As before, the SSA is used to simulate a sample path from this model using the parameters given in Table 4.4, and the data is the daily observed cases for each subpopulation. We assume very little knowledge of the observation probabilities for most subpopulations, with only 2 subpopulations with well-known and high observation probabilities. This is similar to the approach taken by Bhatia et al. [11] to estimate unobserved cases. Figure 4.7, shows the simulated daily cases for each subpopulation. We then use the GA as part of a pmMH scheme to estimate all the non-fixed parameters.

Running the pmMH scheme for $\sim$ 10 hours on 3 cores gives an effective sample size of $\sim$ 1500, after removing 50,000 burn-in samples[2]. The convergence of the chain was validated with the $\hat{R}$ diagnostic [57]. Figure 4.8 shows the marginal posterior distributions for R_0 and p. For most of these parameters, the true values lie in reasonably high-density areas of the posterior distributions. However, some parameters lie in the tails of the posterior suggesting some of the generated subpopulation data is more likely to occur with different sets of parameters. These aberrant estimates may be simply due to an anomalous simulated sample path. Alternatively, the large number of parameters that are being simultaneously estimated could be causing identifiability issues. Of particular practical concern is the posterior for $R_0^{(2)}$ which has most of the probability density below 1 with the true value being 1.5. Recall that $R_0 < 1$ leads to an outbreak that will die out while $R_0 > 1$ would likely lead to a full outbreak. This highlights the importance of incorporating as much valid prior information as possible, something we explore in the next section.

Figure 4.9 shows the linear correlations between basic reproduction numbers as well as between basic reproduction numbers and observation probabilities. Correlations between $R_0^{(i)}$ show mostly positive correlations, explained in part by the fact all share the same T_I, which, as we have discussed in previous chapters, is positively correlated with $R_0^{(i)}$. More-

[2] All code was written in Julia v1.8 and the processor used was the Intel(R) Core(TM) i7-8650U CPU.

over, correlations between the $R_0^{(i)}$ appear to be the largest between subpopulations with the larger outbreaks. This effect is likely due to how much more informative the data from these outbreaks are. There is also a general trend of negative correlation between $R_0^{(i)}$ and $p^{(i)}$ which is as expected since an outbreak with high $R_0^{(i)}$ and low $p^{(i)}$ will produce similar data to an outbreak with low $R_0^{(i)}$ and high $p^{(i)}$. We also observe a pattern of positive correlation between $p^{(1)}$ and the $R_0^{(i)}$ for some of the intermediary subpopulations. This effect can be understood by considering that higher $p^{(1)}$ leads to a smaller proportion of unobserved cases and thus a lower immigration rate to intermediary subpopulations, which must then be compensated for with a higher R_0 in the intermediary subpopulations.

Event	State Update	BP Rate
Infection (1)	$(1,0,0,0,0,0)$	$\beta\mathbf{z}_1$
Infection (2)	$(1,0,0,0,0,0)$	$\beta\mathbf{z}_2$
$\text{Exposed}_1 \to \text{Exposed}_2$	$(-1,1,0,0,0,0)$	$2\delta\mathbf{z}_1$
Symptoms onset (observed)	$(0,-1,1,0,1,0)$	$2p\delta\mathbf{z}_1$
Symptoms onset (unobserved)	$(0,-1,1,0,0,0)$	$2(1-p)\delta\mathbf{z}_2$
$\text{Infectious}_1 \to \text{Infectious}_2$	$(0,0-1,1,0,0)$	$2\lambda\mathbf{z}_3$
Recovery	$(0,0,0,-1,0,0)$	$2\lambda\mathbf{z}_4$
Observation recorded	$(0,0,0,0,-1,1)$	$d\mathbf{z}_5$

Table 4.3: Table of events and rates for $\text{SE}^{(2)}\text{I}^{(2)}\text{R}$ branching process approximation model with delay.

Parameter	True value	Prior	Formulation
Reproduction Number j: $R_0^{(j)}$	(See Figure 4.8)	$\mathcal{N}_{[0,20]}(2, 20^2)$	$R_0 = \frac{\beta}{\lambda}$
Mean latency period: T_E	5	$\mathcal{N}_{[0,6]}(3,2)$	$T_E = \frac{1}{\delta}$
Mean infectious period: T_I	14	$\mathcal{N}_{[2,14]}(8,1)$	$T_I = \frac{1}{\lambda}$
Observation probabilities (high): p_2, p_{19}	0.9	Beta$(9,1)$	-
Observation probabilities (low): p_j, $j \neq 2, 19$	0.6	Beta$(2,1)$	-
Delay mean: T_D	4	fixed	$\frac{1}{d}$
Observation variance: σ^2	1	fixed	-

Table 4.4: Parameters and priors used for $\text{SE}^{(2)}\text{I}^{(2)}\text{R}$ simulation study.

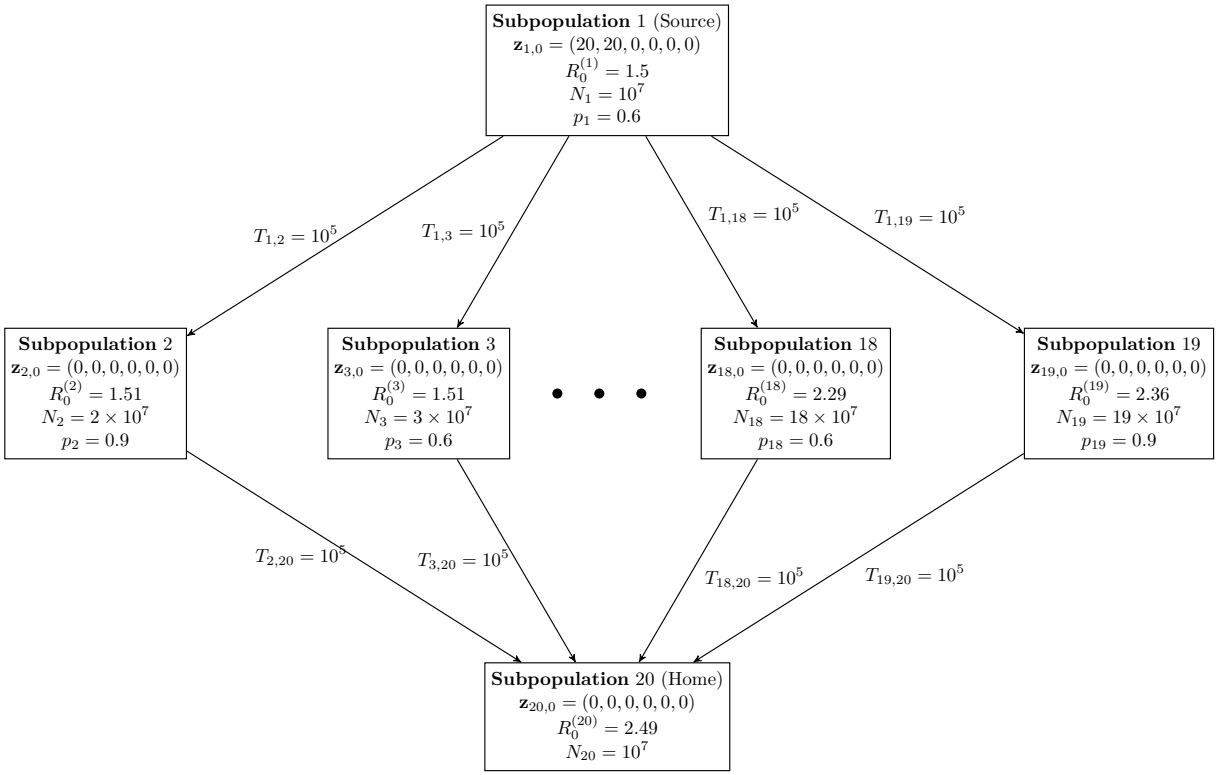


Figure 4.6: Travel network diagram for the 20 subpopulation model. Values of N_i are chosen to give some variety in the immigration rates.

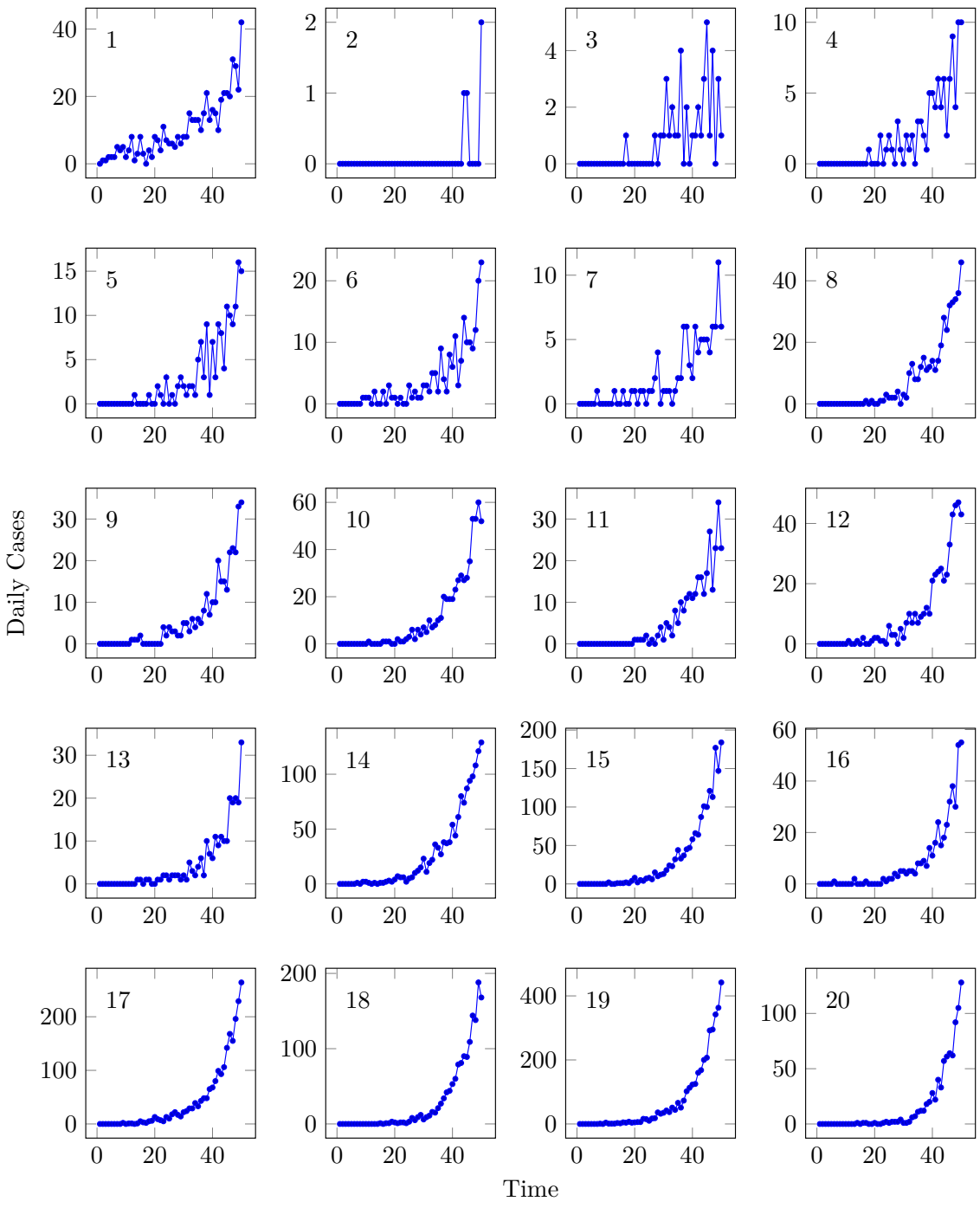


Figure 4.7: Simulated daily case counts for each subpopulation. The top left corner of each plot identifies the subpopulation.

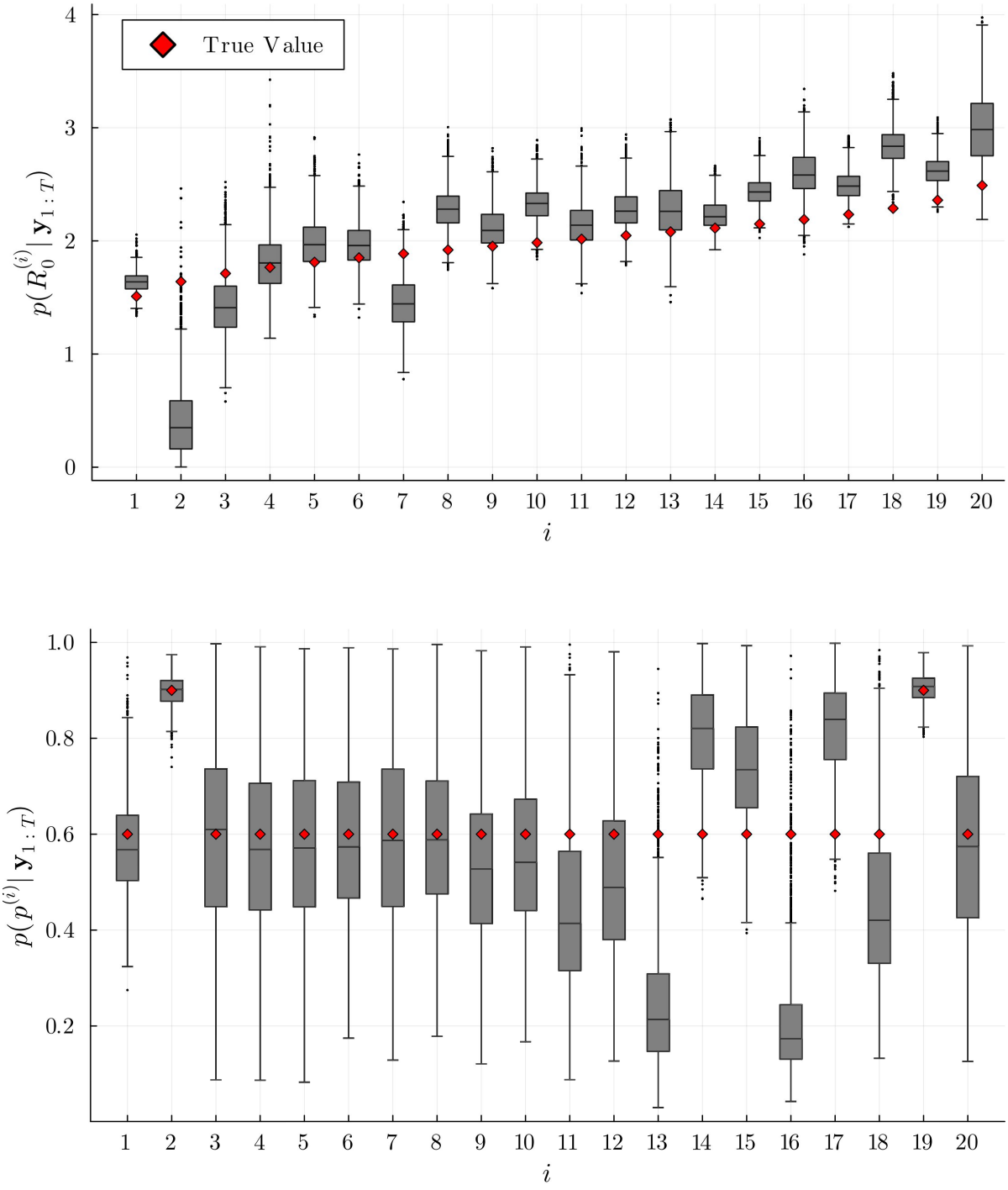


Figure 4.8: Boxplots of marginal posterior distributions for R_0 (top) and p (bottom).

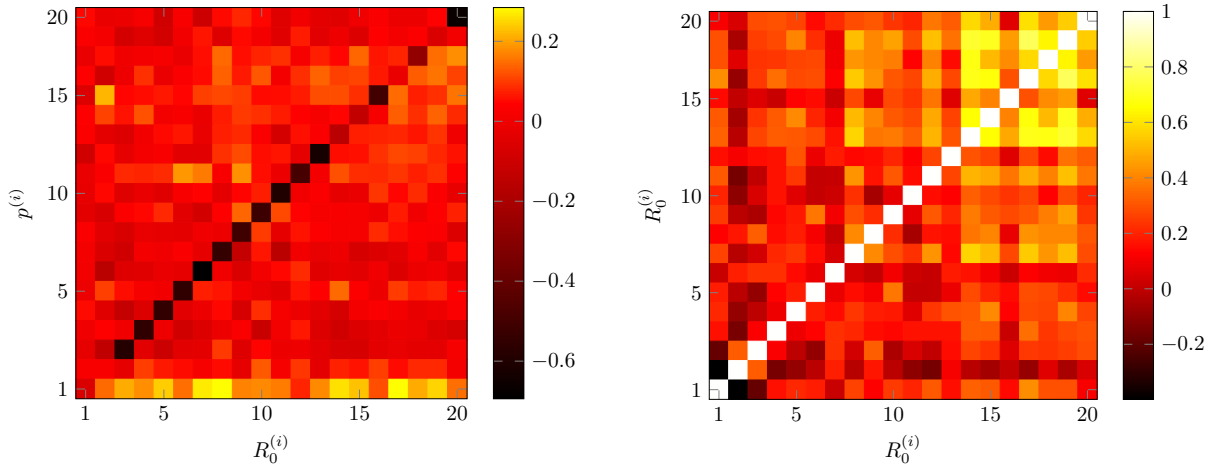


Figure 4.9: Heat map of linear correlations between $\{R_0^{(i)}\}_{i=1}^{20}$ and $\{R_0^{(i)}\}_{i=1}^{20}$ (left) as well as the linear correlations between $\{p_0^{(i)}\}_{i=1}^{20}$ and $\{R_0^{(i)}\}_{i=1}^{20}$ (right).

4.3.1 Hierarchical Model

For global pandemics we expect the disease dynamics to be similar across different subpopulations. Parameters like the mean infectious and latent period that relate specifically to the biology of the disease should be similar across different subpopulations. However, the basic reproduction number will be correlated across subpopulations (since it's the same underlying disease) but still different (due to differences in health care and population density). Ideally, this knowledge should be incorporated into our prior inferences about the parameters. This can be done with a hierarchical model [21]. In particular, we assume the basic reproduction numbers for each subpopulation come from the same distribution

$$\begin{aligned} R_0^{(j)} &\sim \mathcal{N}(\mu_R, \nu_R^2), \\ j &= 1, 2, \ldots, 20. \end{aligned}$$

Here μ_R, ν_R^2 are called *hyperparameters* and they define the prior for $R_0^{(j)}$ (see Figure 4.10). We parameterise $R_0^{(j)}$ in terms of the deviations from μ_R

$$\begin{aligned} R_0^{(j)} &= \mu_R + \nu_R \eta_R^{(j)}, \\ \eta_R^{(j)} &\sim \mathcal{N}(0, 1), \\ j &= 1, 2, \ldots, 20. \end{aligned}$$

This allows us to directly target the differences between populations. Table 4.5 details the parameters and priors we use for the hierarchical model. Using this setup we rerun

the inference routine. We found that the overall performance of the chain was similar to the independent model.

Figure 4.11 shows the comparison between the posterior distributions for R_0. We see that generally, the hierarchical model tends to give lower variance posteriors, which is expected since the hierarchical prior imparts more information. Additionally, overall the hierarchical model appears to give higher densities for the true values. This suggests that information added in the hierarchical prior has remedied some of the identifiability issues associated with R_0. However, the issues with the observation probabilities are still present, emphasizing the need for more informative priors.

Parameter	True value	Prior
Reproduction number mean: μ_R	2	$\mathcal{N}_{[0,20]}(2, 20^2)$
Reproduction number variance: ν_R^2	0.5	$\mathcal{N}_{[0,\infty]}(0, \frac{1}{9})$
Subpopulation j deviation: $\eta^{(j)}$	(See Figure 4.11)	$\mathcal{N}(0, 1)$
Mean latency period: T_E	5	$\mathcal{N}_{[0,6]}(3, 2)$
Mean infectious period: T_I	14	$\mathcal{N}_{[2,14]}(8, 1)$
Observation probabilities (high): p_2, p_{19}	0.9	Beta$(9, 1)$
Observation probabilities (low): p_j, $j \neq 2, 19$	0.6	Beta$(2, 1)$
Delay mean: T_D	4	fixed
Observation variance: σ^2	1	fixed

Table 4.5: Parameters and priors used for the hierarchical simulation study.

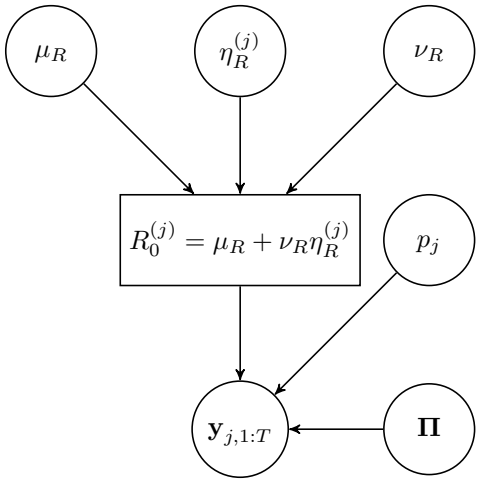


Figure 4.10: Hierarchical parameter diagram. Here $\mathbf{\Pi}$ denotes the parameters that are identical across all subpopulations: T_I, T_E, T_D and σ^2.

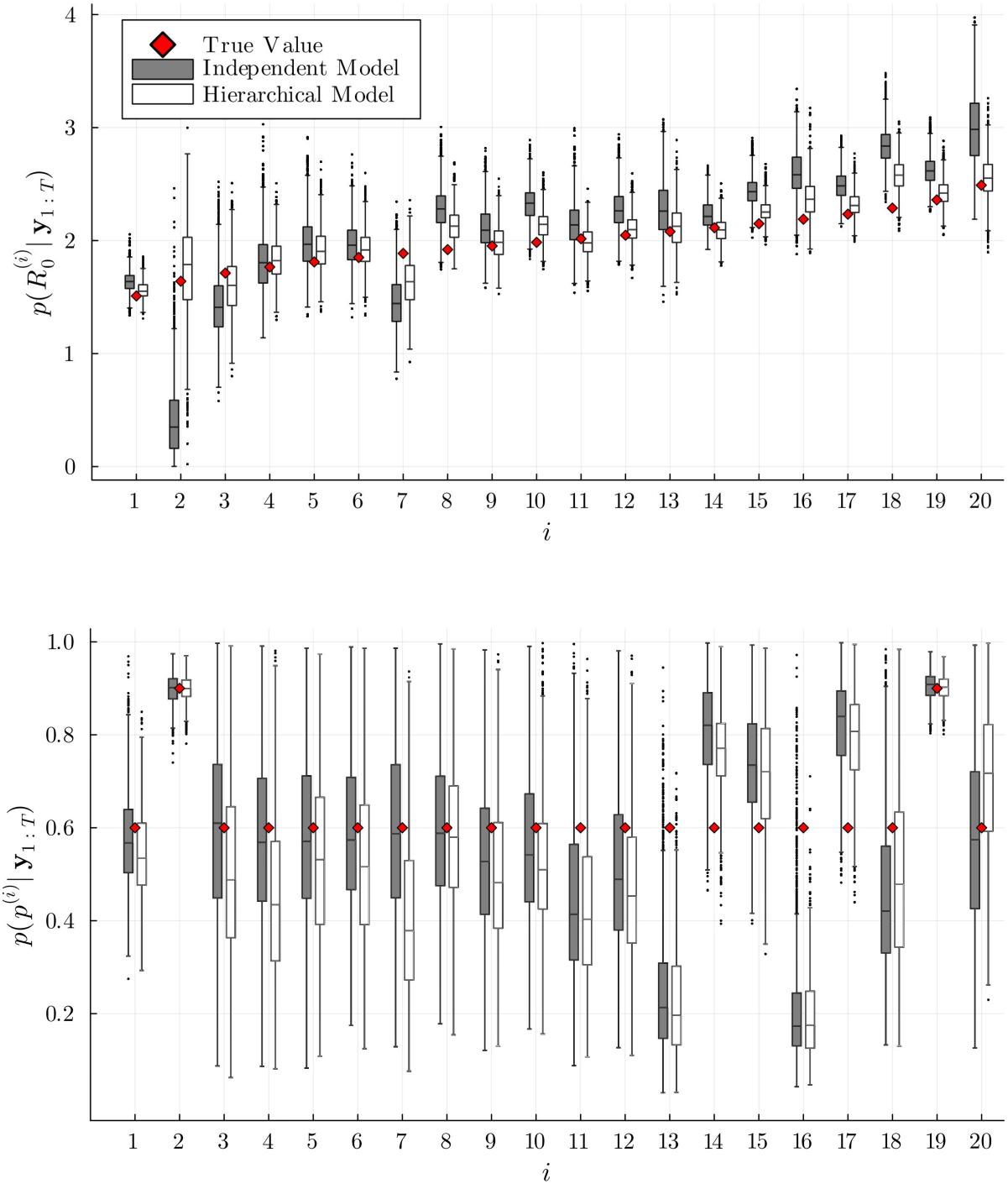


Figure 4.11: Boxplots of marginal posterior distributions for R_0 (top) and p (bottom) for both the independent and hierarchical models.

4.3.2 Effect of Observation Priors

In this section, we explore the effect of changing our prior knowledge of the observation probability on the independent and hierarchical models. We repeat the same procedure but instead, we assume subpopulations with high, low and unknown levels of observation probability (See Table 4.6). Overall this assumes far more knowledge about which subpopulations we expect to have higher or lower observation probabilities. Moreover, we also include some subpopulations which are simply not monitoring the disease. This element of the setup mimics the early phases of a global pandemic where it may be unclear if cirtain contries are conducting/reporting cases. With this setup, we simulate a set of observations (see Figure 4.12).

As before, we conduct MCMC to obtain the parameter posteriors using the Gaussian approximation for both the independent and hierarchical models. Figure 4.13 shows the comparison between the posterior distributions for R_0 and p. We again see that the hierarchical model posteriors align more closely with the true values for R_0. In particular, for subpopulation 17, which has no observation, the independent model performs very poorly. The reason for this poor performance is evident in the corresponding observation probability estimates. Note that the independent model appears to favour higher observation probability estimates, for these subpopulations. This is then compensated for with lower estimates for R_0 to account for the lack of observation. This explanation is evidenced by the high negative correlation (~ -0.4) between $p^{(17)}$ and $R_0^{(17)}$, for the independent model. This issue is not observed in the hierarchical model, likely due to the more restrictive priors. Furthermore, as expected the variance in the posterior distributions of R_0 is less for the hierarchical model in most subpopulations due to the extra knowledge encapsulated by the hierarchical prior.

Comparing Figure 4.13 and 4.11 the variances in R_0 values for the hierarchical model are generally lower. This result is as expected since we are assuming more knowledge of the observation probability. Moreover, this increase in prior knowledge appears to have mitigated some of the issues with estimating p seen in Figure 4.11.

Parameter	True value	Prior
Observation probabilities (high)	0.9	$\text{Beta}(9, 1)$
Observation probabilities (low)	0.6	$\text{Beta}(2, 1)$
Observation probabilities (unknown)	0	$\text{Exp}_{[0,1]}(1)$

Table 4.6: Parameters and priors used for the hierarchical simulation study.

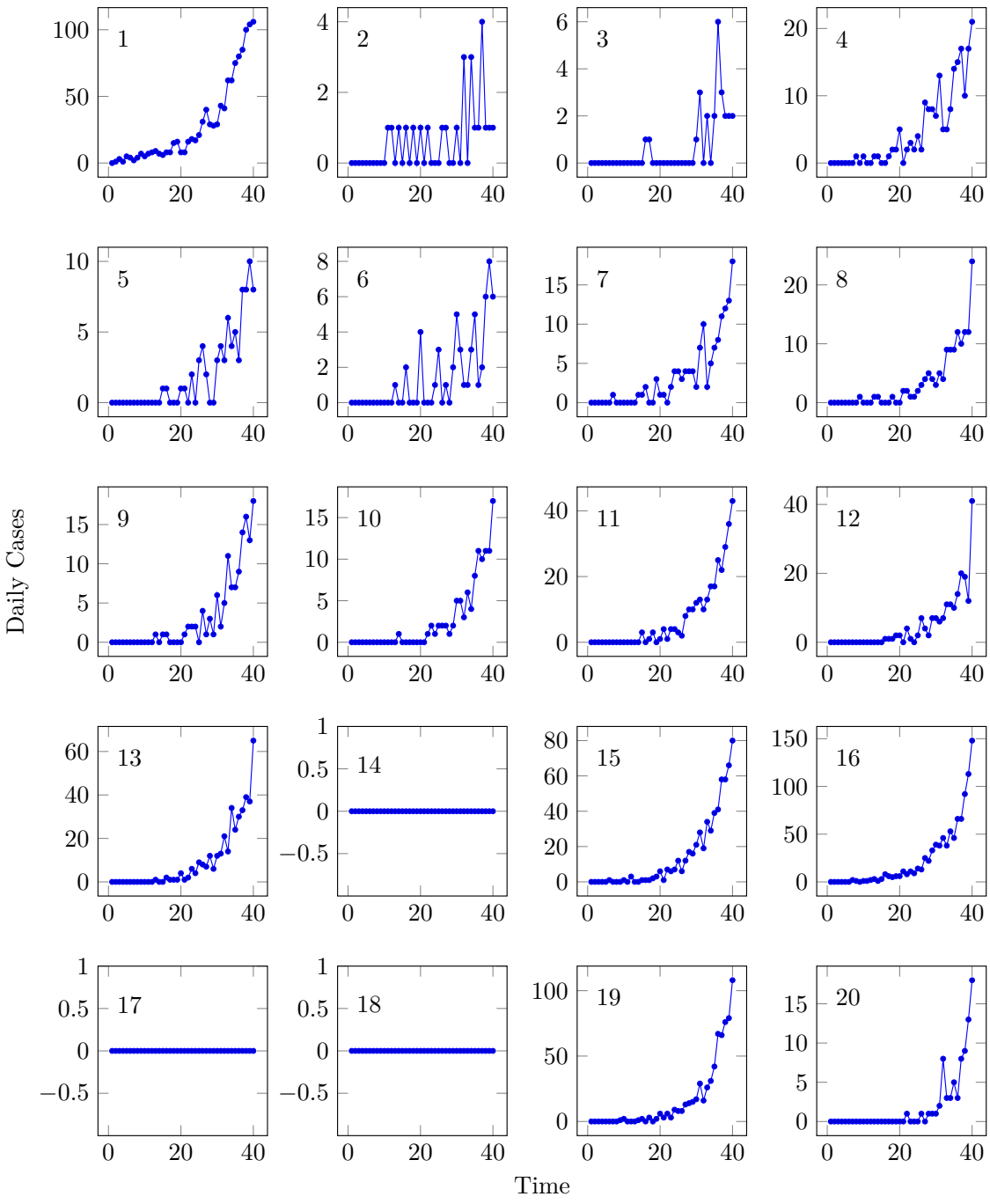


Figure 4.12: Simulated daily case counts for each subpopulation. The top left corner of each plot identifies the subpopulation.

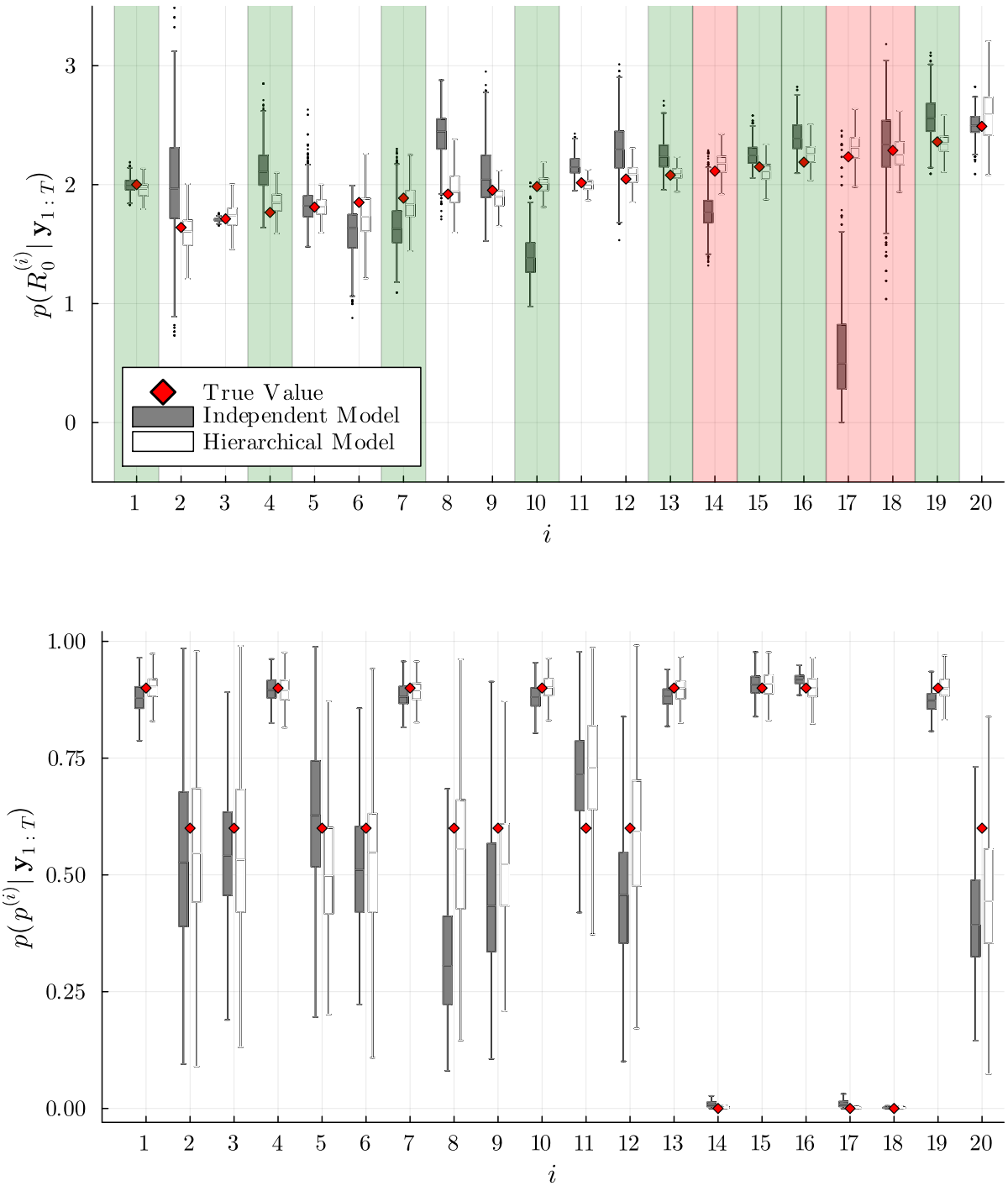


Figure 4.13: Boxplots of marginal posterior distributions for R_0 (top) and p (bottom) for both the independent and hierarchical models. Green and red shading denotes subpopulations with high and low levels of observation respectively.

4.3.3 Importation Risk Estimation

In this section, we briefly illustrate how the estimates produced by the methods we have developed can be used to produce actionable insights relating to pandemic policy. In particular, in Australia, estimating the risk of disease importation was of great interest during the early phase of the SARS-CoV-2 pandemic [9]. Here we show how full posterior distributions can be produced for the expected number of imported outbreaks and the probability of an imported outbreak. Here an imported outbreak refers to an outbreak stated by an imported individual.

For a given infectious or exposed individual we define the event that an outbreak caused by that individual to be simply that the progeny of that individual never becomes extinct. The extinction probabilities $\mathbf{x}$ will solve the system

$$\mathbf{P}_{\boldsymbol{\theta}}(\mathbf{x}) + \mathbf{u}_c = \mathbf{x}, \quad \text{subject to} \quad \mathbf{x} \in [0,1]^r \tag{4.14}$$

where $\mathbf{P}_{\boldsymbol{\theta}}(\mathbf{x})$ is the progeny generating function given $\boldsymbol{\theta}$, and c is the index of the observed state [50, p. 12]. Given our Poisson immigration process, the number of new outbreaks in the sink subpopulation, per day, caused by subpopulation i is a censored Poisson process. Thus, we can derive the probability of an imported outbreak ($\text{IOP}_{i,t}$) from subpopulation i, over the period $[t, t+1)$.

$$\text{IOP}_{i,t} = 1 - \exp((\mathbf{x} - \mathbf{1})\boldsymbol{\alpha}_{i,20,t}^T), \tag{4.15}$$

and the expected number of imported outbreaks per day ($\text{ENO}_{i,t}$), from subpopulation i, over the period $[t, t+1)$.

$$\text{ENO}_{i,t} = (\mathbf{1} - \mathbf{x})\boldsymbol{\alpha}_{i,20,t}^T. \tag{4.16}$$

Calculating $\text{ENO}_{i,t}$ and $\text{IOP}_{i,t}$ using our samples from $p(\boldsymbol{\theta}, \mathbf{z}_t|\mathbf{y}_{1:T})$ gives us the posterior distribution for these quantities. Moreover, we are not restricted to $t \leq T$ since we can simply simulate forward samples from $p(\mathbf{z}_T, \boldsymbol{\theta}|\mathbf{y}_{1:T})$ to obtain $p(\mathbf{z}_{T+\tau}, \boldsymbol{\theta}|\mathbf{y}_{1:T})$ which can be used to derive $p(\text{ENO}_{i,T+\tau}|\mathbf{y}_{1:T})$ and $p(\text{IOP}_{i,T+\tau}|\mathbf{y}_{1:T})$.

Figure 4.14 gives a snapshot of $\text{ENO}_{i,t}$ and $\text{IOP}_{i,t}$ using our hierarchical model and independent model, described in Section 4.3.2. For subpopulation 17 we see that the lower R_0 estimate given by the independent model significantly affects both the ENO and IOP with the hierarchical model adhering much closer to the true values. For subpopulation 18 we instead see the hierarchical model can give a much lower variance posterior distributions, aligning well with the true values. This illustrates that even in subpopulations with zero observations reasonable importation risk estimates can be produced with the hierarchical model.

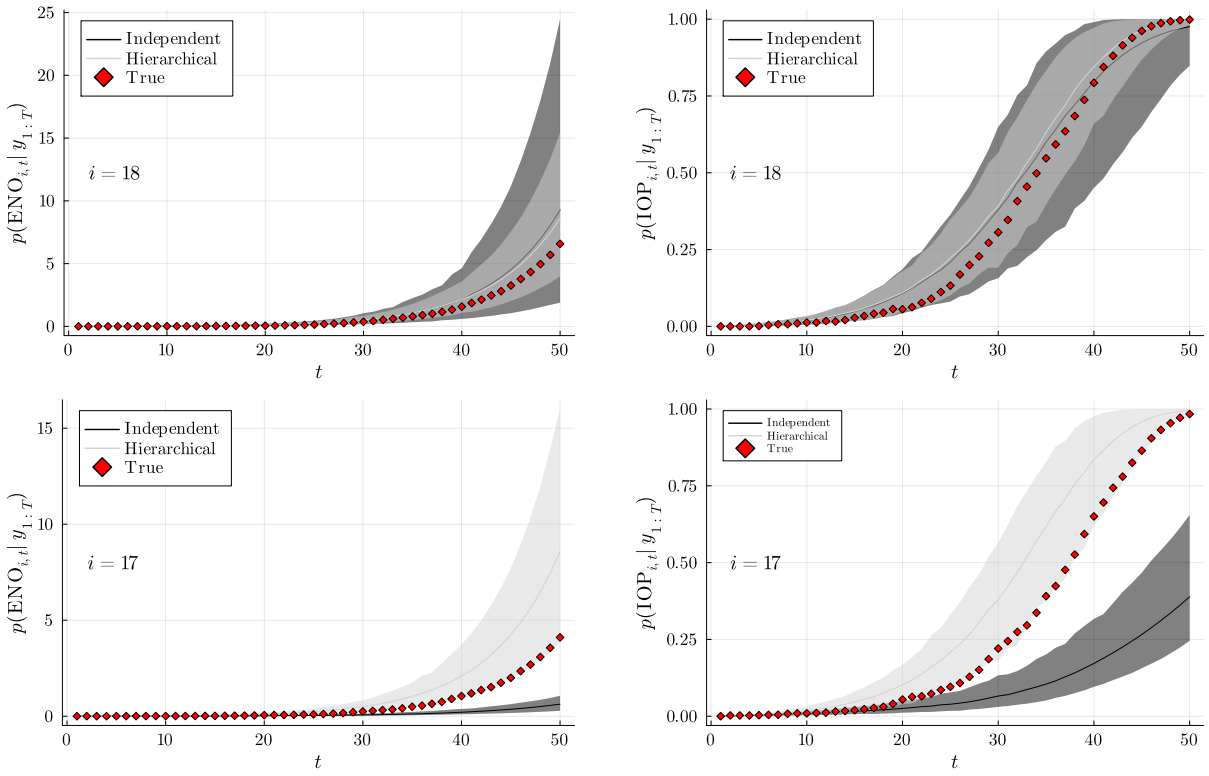


Figure 4.14: Plots of the posterior distributions of $\mathrm{ENO}_{i,t}$ and $\mathrm{IOP}_{i,t}$. Ribbons indicate 90% credible intervals.

Chapter 5

Discussion

In this thesis, our primary goal has been to develop a modelling framework for estimating the risk of disease importation during the early phase of a pandemic. To achieve this goal we developed an efficient Gaussian approximation of the bootstrap particle filter that can be feasibly used as part of the pseudo-marginal Metropolis-Hastings algorithm scheme for continuous-time branching process models (Chapter 3). We then showed how this Gaussian approximation makes it possible to jointly estimate the states and parameters for an epidemic model with a metapopulation structure (Chapter 4).

5.1 Inference for Partially Observed Continuous Time Branching Processes

In Chapter 3 we have developed an algorithm for approximate state and likelihood estimation that can run significantly faster than a standard PF, making it more feasible to be used as part of a Metropolis-Hastings scheme for parameter estimation [29]. Additionally, the likelihood estimates this algorithm produces have less variance, reducing the number of bootstrap samples required, further reducing computational time. Moreover, we have shown through simulation, given sufficiently high agent counts, that the Gaussian approximation can be reasonably accurate for both the SIR and $SE^{(2)}I^{(2)}R$ branching process models.

Continuous time branching processes pose an interesting inferential challenge since, in general, it is infeasible to obtain exact transition probabilities [50]. Thus, any inference method that is used must overcome this efficiently. The most obvious approach is through simulation of the process, this being how the bootstrap particle filter works [62, p. 129–165]. Despite this, as previously discussed, this can be far too computationally expensive, due to exponentially growing agent counts. One alternative avenue that has been explored in the literature is to approximate the transition probabilities numerically.

In particular, by truncating the state space and taking the matrix exponential of the corresponding infinitesimal generator the transition probabilities can be approximated. However, this approach can also become prohibitively computationally expensive as the agent population gets large. Xu and Minin [36] and Xu et al. [41] have improved upon this method, making clever use of the progeny-generating function. Despite this, when agent counts are high, even with these improvements, computational time can still become problematic, particularly for use as part of a pmMH scheme that requires repeated evaluation of the PF. The reason for this is that the transition density of a CTBP is discrete and non-finite, hence, the approximations derived in these works have to be truncated to a finite subset of the reasonably likely states. Then differential equations for each of the states need to be numerically solved. As a result, this class of methods is susceptible to increases in the agent population as well as the dimension of the state space. Consequently, inference for CTBPs is usually focused on deriving point estimates, using expectation maximisation type algorithms, rather than full posterior distributions [79, 80, 41]. Our approach is instead focused on modelling over a shorter time series, with less informative data, where the uncertainty in our estimates is crucial to quantify. This is also what precludes us from using taking a deterministic approach, which would not capture the variance of the process sufficiently [30, 66, 31]. Our method differs from these approaches by applying a continuous Gaussian approximation of the transition density for large populations, meaning we do not have to perform a calculation for each state, only the mean and variance. As a result, the computational complexity of the method we develop is not dependent on the agent population only the dimension of the state space.

In our results, one clear theme observed across our experiments is that the filtering distributions produced by the Gaussian approximation appear to get closer to the correct filtering distributions generated by the PF as the agent counts increase. This effect is even seen when using the Gaussian approximation throughout the entire time series, meaning the errors from inaccurate early filtering distributions do not compound over time. We suspect this is in part caused by the new observations preventing the mean of the posterior estimates from deviating significantly. Also, the Gaussian approximation will be most accurate for the highly informative parts of the time series, since a high agent count tends to lead to high observation numbers. This effect can help override the more inaccurate state estimates produced by the Gaussian approximation at low agent counts, explaining the surprising accuracy of the $s = 0$ switching threshold in Sections 3.3.2 and 3.3.1. Additionally, the mean and variance of $\mathbf{z}_{t+1} \mid \mathbf{z}_t$ is exact, as is $\mathbf{z}_{t+1} \mid \mathbf{y}_t$, even if $\mathbf{z}_t \mid \mathbf{y}_t$ is not Gaussian. This brings us to the limitation of extending this work to continuous time Markov chains (CTMCs) in general. Specifically, for a general CTMC the mean and variance of the transition density are not tractable. Consequently, linearisation would need to be utilised to approximate the mean and variance of the transition kernel. These approximations could lead to errors that may compound throughout the time series [49].

The linearization approach has been applied by Fintzi et al. [81], however, simulation is still utilized to derive predictive distributions. Using linearization to calculate means and variances and assuming Gaussian predictive distributions (see Section 3.2.2) would allow this method to be adapted to a CTMC. The exploration of this possibility remains an endeavour for future research.

One aspect of the Gaussian approximation that we did not explore deeply in this work is the choice of switching threshold, in particular concerning time-constrained practical applications such as a disease outbreak where new state and parameter estimates need to be produced daily. In situations where new state and parameter estimates are required at regular intervals, the switching threshold could be made as high as is feasible. Specifically, the time taken for the likelihood to be calculated with the required variance could be tested at different thresholds. Combining this with the total number of Metropolis-Hastings samples required would enable a researcher to estimate the threshold that gives the most accurate results for a given amount of time.

Given the general formulation of the Gaussian approximation it could theoretically be used for any continuous time branching process. The main bottleneck to increasing the complexity of the model is the dimension of the state space. In particular, the addition of more intermediary states will mean that the number of agents will be spread out across many agent types. This means to attain an accurate approximation, the PF will need to be used for longer and in cases where the overall number of events is higher. Additionally, calculation of the mean and variance requires computations involving matrices of size $r^3 \times r^3$, where r is the dimension of the state space [50], which may be problematic for models with very high dimensional state spaces. Moreover, models including agent types that tend to be significantly smaller (in terms of population) than other agent types will diminish the benefit provided by this method. This limitation exists because the transition density for all agent types is simulated/calculated the same way, meaning if $[\mathbf{z}_t]_1$ is very large (well suited for the Gaussian approximation) and $[\mathbf{z}_t]_2$ is very small (poorly suited for the Gaussian approximation) problems will arise. Specifically, using the PF will incur a significant computational effort as $[\mathbf{z}_t]_1$ is very large. Alternatively, using the Gaussian approximation may be inaccurate due to the low value of $[\mathbf{z}_t]_2$. The difficulty with computing the different agent types using different methods is that there is often a significant correlation between $[\mathbf{z}_t]_1$ and $[\mathbf{z}_t]_2$. We suspect that maintaining this correlation while using separate methods would be challenging.

Section 3.2.2 brings to light some issues with the broader applicability of the Gaussian approximation. In particular, the fact that the mean of the filtering distributions must dominate the variance suggests that how informative the data is will impact the accuracy of the Gaussian approximation. Recall that the accuracy of the predictive distribution

approximation is dependent on the elementwise standard deviation of the filtering distribution being small in comparison to the elementwise mean. Further, the elementwise standard deviation will be larger for less informative data. However, this is not something we found to be significant in our results, suggesting much less informative data would be required to observe the effect. This may also pose an issue for longer-term forecasting since the variance of the state estimates will naturally grow larger as data is no longer being observed. In this case, it would make sense to draw samples from the filtering at the latest time when data is available and simulate forward with Gillespie's algorithm to obtain forecasts.

5.2 Meta-Population Pandemic Modelling Framework

In Chapter 4 we have shown how the Gaussian approximation of the PF can be applied to an epidemic model with a metapopulation structure. More approximations needed to be made to achieve this goal. Through simulation, we found that these approximations do not significantly affect the results, of the cases tested. In particular, the results of the bootstrap particle filter and Gaussian approximation align very closely, even under conditions expected to exacerbate errors. Specifically, as discussed in Section 4.1.3, high immigration rates will cause the filtering distributions to have systematically lower variance. In our simulation study, we assumed that 1% of the population in each subpopulation travels to every other subpopulation every day, exceeding any immigration rate we would observe in practice. Hence, the observed correspondence between the methods, in this case, indicates that applying the Gaussian approximation to a metapopulation model will give a reasonable approximation of posterior distributions. However, it is important to note that, although we do not expect this the be an issue for epidemic models, this result may not apply to situations with significantly higher immigration rates. Despite this, given the excellent correspondence between methods, we suspect the Gaussian approximation may still be accurate for higher immigration rates than we tested. For these cases, it would be important to validate the Gaussian approximation via simulation.

Another limitation to consider is that the accuracy issues present in the single population case will carry over to the metapopulation case. In particular, the accuracy of the predictive distribution approximation is dependent on the elementwise standard deviation of the filtering distribution being small in comparison to the elementwise mean. In practice, this could be problematic for subpopulations with very small case counts, causing lower estimates of infected individuals and R_0. This maybe be, in part, the reason for the low estimate of $R_0^{(2)}$ in Figure 4.8. Fortunately, populations with low case counts are not very informative meaning the use of accurate priors is quite effective in mitigating this issue. This effect is evident in Figure 4.11 where we observe that the hierarchical model

assigns greater probability density to the true value. Note that the hierarchical model is not imposing a prior directly on the value of $R_0^{(2)}$ forcing it to increase, rather it is using the information obtained from other larger outbreaks, which are both more informative and more accurately estimated by the Gaussian approximation. In this way, the hierarchical structure of the parameters helps to reduce the impact of errors associated with the Gaussian approximation. Despite this, it must be emphasised that this improvement requires the hierarchical structure to be accurate to the situation, which is assumed in our testing. The hierarchical model does make sense in the context of a global pandemic since the same underlying disease is the cause. Moreover, we cannot state for certain that the hierarchical model is *improving* the estimate for $R_0^{(2)}$ simply because the posterior assigns more density to the real value. It is possible that the *correct* posterior assigns low density to the true value, it is just less likely. The only way to truly test the accuracy of the posteriors is to compare the result to a bootstrap particle filter, which would be computationally infeasible.

While the hierarchical model brought posteriors for R_0 closer to the true values it did not have this effect for the observation probabilities (see Section 4.3.1). This issue was somewhat mitigated by using stricter priors for the observation probabilities of more subpopulations (see Section 4.3.2). However, in cases where very little is known about the observation probabilities of most subpopulations, this issue is important. While we cannot be certain that the deviation of the posterior distributions from the true values implies an inaccuracy it is suggestive of a problem with the framework. It is unclear whether the inaccuracy of the Gaussian approximation is causing the issue or if there is no issue and the posteriors are approximately correct. In the cases where we can calculate the true posterior distributions, we have found good adherence to the ones produced by the Gaussian approximation. However, these tests were performed on models with far fewer subpopulations due to computational limitations. While it is unclear how increasing the number of subpopulations would hinder the performance of the Gaussian approximation it certainly cannot be ruled out. This narrow validation of the Gaussian approximation for the metapopulation is a weakness of this work and represents an opportunity for future work.

In Section 4.3.3 we showed how the joint state and parameter estimates can be used to derive posterior distributions for the probability of an imported outbreak and the expected number of imported outbreaks—the primary goal of this thesis. However, further work can explore how other actionable insights can be derived by extending the framework. For example, the effect of airport screening individuals from certain countries on the probability of an imported outbreak could be analysed similarly by censoring the immigration rates and simulating from the posterior predictive distribution. This inference could be used to determine an optimal allocation of screening resources or to determine

if a proposed policy is effective [82].

The framework developed in Chapter 4 builds on the method developed by Shearer et al. [9]. The most important extension we make is including the parameter estimation within the risk estimation framework rather than relying on early estimates [16]. Estimating parameters in this way makes use of the latest data, rather than having to wait for estimates to be published based on older datasets. This is of great practical importance for the early phases of a novel disease outbreak when data is scarce. Additionally, the parameter estimates in the literature may not perfectly carry over to the underlying dynamic model being used. Alternatively, including parameter estimation in the framework ensures perfect correspondence between parameter estimates and the model of the epidemic dynamics.

Shearer et al. [9] makes use of the immigration model from De Salazar et al. [10], where imported cases I are a Poisson process with rates dependent on immigration rates from China

$$I \sim \text{Pois}(\beta C), \tag{5.1}$$

where β is fit to the data. This statistical approach differs from the mechanistic approach we take. Since we estimate the number of infected individuals in each country we can determine the expected number of imports based on the travel volume between subpopulations. The mechanistic approach removes the need to estimate any immigration-related parameters, which will make the data more informative for the other parameters. However, a weakness of our approach is that it relies on the assumption that travelling individuals are sampled independently and uniformly from the population. Clearly, this assumption is not completely valid but the effect of it on predictive capability is an open question.

Bhatia et al. [11] made use of a model similar to De Salazar et al. [10], again using air travel with China to estimate the proportion of unobserved cases in various countries. To help in estimating this proportion, Singapore is assumed to have perfect case observation. This gives a reference point for how the rate of immigration from China relates to the total number of observed and unobserved cases. In Section 4.3 we show how this idea can be expanded upon and applied to our framework. Specifically, subpopulations expected to have high levels of monitoring can have priors tight for the observation probability p. This gives a more accurate picture of our knowledge since we cannot know that every single case is being observed or the exact observation probability. Additionally, this approach, in tandem with the hierarchical structure, can help resolve the issues associated with estimating R_0 and the observation probability together. Without tight priors on the observation probability, it can be difficult to determine if data was generated with a high R_0 and low observation probability or a low R_0 and high observation probability. Having

some subpopulations with tight priors on p fixes that issue for those subpopulations, and then the hierarchical structure helps to mitigate those issues for other subpopulations (See Section 4.3.2).

In Section 4.3 we explored the computational feasibility of applying the Gaussian approximation to a model with 20 subpopulations, finding that an effective sample size of $\sim$ 1500 could be obtained in $\sim$ 10 hours of computation on a laptop. In practice, this means new parameter and state estimates could be produced each day as new case data arrives. We did not test the bootstrap particle filter (PF) against the Gaussian approximation in this problem as a single PF likelihood evaluation was found to take over 10 minutes (note that $\sim$ 50000 likelihood evaluations were required to burn in the Metropolis-Hastings chain). This highlights the difficulty of applying sequential Monte Carol methods to complex parameter estimation problems of this size. The Gaussian approximation alleviates these issues in 2 ways: removing all variance in the likelihood estimation and reducing the time taken to produce the result. Moreover, the bulk of the computation time is in calculating the mean and variance for the CTBP which only needs to be done once per likelihood evaluation regardless of the length of the time series. As a result, it is unlikely that the length of the time series will become a limiting factor.

However, the method we have developed is limited in the number of parameters that can be simultaneously estimated. For our hierarchical model, we estimated 45 jointly, this being reasonably close to the maximum of what is feasible within a day of computation. In practice, we may want to have many subpopulations for each country all with their own parameters, which would certainly not be feasible within our framework. Fundamentally, this issue cannot be fixed by improving the speed of the likelihood estimator (the focus of this work) as it is a problem with the Metropolis-Hastings algorithm itself [83]. One possible avenue for mitigating this effect is Hamiltonian Monte Carlo (HMC) [84]. This method reduces the correlation between the sampled parameters while still maintaining reasonable acceptance probabilities, reducing the number of samples required to sufficiently explore the parameter space. HMC requires the calculation of the gradient of the likelihood function

$$\nabla \ell(\boldsymbol{\theta}) = \left(\frac{\partial p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})}{\partial \theta_1}, \frac{\partial p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})}{\partial \theta_2}, \ldots, \frac{\partial p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})}{\partial \theta_n} \right).$$

For the Gaussian approximation, this can be achieved with automatic differentiation [85]. However, it is unclear whether the extra computation time required to calculate the gradient would offset the advantage provided by HMC and thus would need to be explored in future works.

The hierarchical model we test in this thesis is relatively simple, assuming nothing about the individual subpopulations. In reality, our scientific understanding of epidemi-

ology can give insight into which subpopulations we expect to have higher or lower R_0. A simple example of this is population density. The R_0 of a subpopulation will generally be positively correlated with the population density of that subpopulation [18]. Furthermore, this inference was strongly validated by data from the COVID-19 outbreak [19, 20]. Thus, it makes sense to include this type of inference in our priors. This could be achieved by altering the prior for the subpopulation deviation $\eta^{(j)}$ to make it dependent on the population density PD_i;

$$\eta^{(j)} \sim \mathcal{N}(f(\mathrm{PD}_i), 1), \tag{5.2}$$

where f is some increasing function. Note that, this maintains the hierarchical structure as $R_0^{(j)} = \mu_R + \nu_R \eta^{(j)}$ but also makes use of our knowledge about population density. Incorporating as much prior information as possible into these types of models is vital due to the scarcity of data, thus, representing an important direction for further work.

A strength of this framework is its flexibility concerning the dynamic model, in particular, the Gaussian approximation can be applied to any continuous-time branching process. This allows for a wide variety of disease and observation dynamics to be modelled. For example, outbreak dynamics with high heterogeneity have the possibility of leading to super spreading events (SSEs), where one infected individual is the cause of many secondary infections [86]. For both severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), SSEs were a significant factor in their spread [87, 88]. Additionally, there is strong evidence of COVID-19 exhibiting similar properties [89, 90, 91]. Figure 5.1 shows how super-spreading individuals can be easily incorporated into a CTBP by adding an extra agent type with a greater infection rate. Moreover, super-spreading events can be easily incorporated by adding extra events that create more infectious individuals. Another extension that can be easily incorporated into the framework is the effect of social isolation when infected. Social isolation can similarly be incorporated by adding an extra agent type corresponding to an isolated individual with a lower associated infection rate. However, adding extra agent types increases the dimension of the state space which is a severe bottleneck for the Gaussian approximation. Thus, care needs to be taken to ensure that the increased accuracy provided by these more complex models is worth the extra computational cost. This is particularly important in the early phases of an outbreak, as the scarcity and noise of data are likely to render some of the more complex models redundant. Additionally, during the early phases of a novel pandemic, it is unlikely that our knowledge of the disease dynamics will be sufficient to formulate an *accuracte* complex model. In these situations, simple models would likely be the best choice.

Demining which factors are the most important to capture and how best to incorporate them into these types of models in low-data situations represents an important avenue for further research. Furthermore, the important factors will likely differ depending on

the specific disease, emphasising the importance of a flexible inference framework.

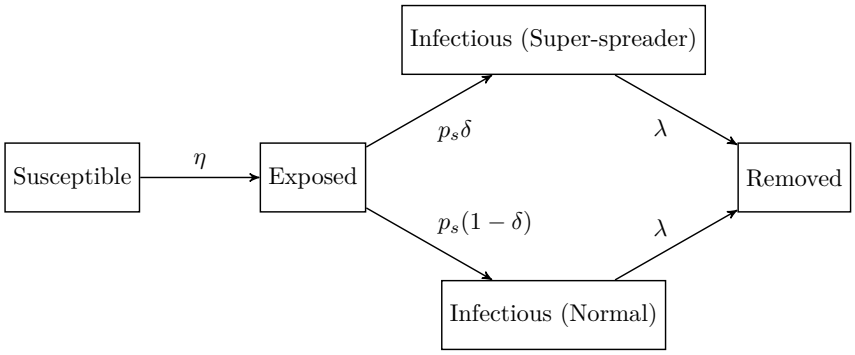


Figure 5.1: CTBP incorporating super spreading individuals. Here η is the effective infection rate, $\frac{1}{\delta}$ is the mean incubation time, λ is the recovery rate and q is the probability of an individual being a super spreader.

While our work has shown the computational feasibility of the framework, we have not deeply explored the model formulation issues. In particular, the choice of the specific branching process model and priors is something that will differ for each outbreak. To answer these questions it will be important to apply the metapopulation framework to real-world datasets. Investigation of these questions represents the most important step in developing this framework into a practically useful tool.

In summary, we have developed a computationally inexpensive approximate inference method for a broad class of partially observed continuous time-branching processes. This method makes strategic use of particle filters and a Gaussian approximation to minimise both computation time and bias. Moreover, we have shown how this method can be adapted to a meta-population model as part of a framework for estimating the importation risk of a novel disease.

Chapter 6

Appendix

6.1 Kalman Filter

For a hidden Markov model with states $\{\mathbf{z}_t\}_{t=0}^T$ and data $\{\mathbf{y}_t\}_{t=1}^T$ defined as

$$\begin{aligned}\mathbf{z}_0 &\sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0),\\ \mathbf{z}_t \mid \mathbf{z}_{t-1} &\sim \mathcal{N}(\mathbf{F}_t\mathbf{z}_{t-1} + \mathbf{B}_t\mathbf{u}_{t-1}, \mathbf{Q}_t),\\ \mathbf{y}_t \mid \mathbf{z}_t &\sim \mathcal{N}(\mathbf{H}_t\mathbf{z}_t, \mathbf{R}_t).\end{aligned}$$

for $t \in \{1, 2, \ldots, T\}$, where $\{\mathbf{z}_t\}_{t=0}^T$, $\{\mathbf{u}_t\}_{t=0}^T$ and $\boldsymbol{\mu}_0$ are r element column vectors, $\{\mathbf{y}_t\}_{t=1}^T$ are d element column vectors, $\boldsymbol{\Sigma}_0$, $\{\mathbf{F}_t\}_{t=1}^T$, $\{\mathbf{B}_t\}_{t=1}^T$ and $\{\mathbf{Q}_t\}_{t=1}^T$ are $r \times r$ matrices, $\{\mathbf{H}_t\}_{t=1}^T$ are $d \times r$ matrices and $\{\mathbf{R}_t\}_{t=1}^T$ are $d \times d$ matrices. In this case, all filtering distributions are Gaussian. The mean $\boldsymbol{\mu}_{t|t}$ and variance $\boldsymbol{\Sigma}_{t|t}$ of the filtering distributions, as well as the likelihood $p(\mathbf{y}_t|\mathbf{z}_t)$ can be derived with Algorithm 4 [29, 56 – 58]. The smoothed distributions are also Gaussian and the means $\left\{\boldsymbol{\mu}_{t|T}\right\}_{t=1}^T$ and variances $\left\{\boldsymbol{\Sigma}_{t|T}\right\}_{t=1}^T$ can be derived with Algorithm 5 [29, 136 – 137].

6.2 Bivariante Posterior Distribution Comparison

Figure 6.1 shows comparisons of the estimated bivariate posterior distributions using the PF and GF, based on the experiment discribed in Section 4.2.2. As expected, the bivariante distributions produced by GF align reasonablly closely the samples produced by the distributionally accurate PF.

Algorithm 4
Kalman Filter

Input: Data $\{\mathbf{y}_t\}_{t=1}^T$, inital mean $\boldsymbol{\mu}_0$ and variance $\boldsymbol{\Sigma}_0$, model parameters $\{\mathbf{u}_t\}_{t=0}^T$, $\{\mathbf{F}_t\}_{t=1}^T$, $\{\mathbf{H}_t\}_{t=1}^T$ and $\{\mathbf{B}_t\}_{t=1}^T$.
Initialise $t \leftarrow 0$, $\ell \leftarrow 0$, $\boldsymbol{\Sigma}_{0|0} \leftarrow \boldsymbol{\Sigma}_0$ and $\boldsymbol{\mu}_{0|0} \leftarrow \boldsymbol{\mu}_0$
for $t \in \{1, 2, \ldots, T\}$ **do**
$\boldsymbol{\mu}_{t|t-1} \leftarrow \mathbf{F}_t \boldsymbol{\mu}_{t-1|t-1} + \mathbf{B}_t \mathbf{u}_{t-1}$.
$\boldsymbol{\Sigma}_{t|t-1} \leftarrow \mathbf{F}_t \boldsymbol{\Sigma}_{t-1|t-1} \mathbf{F}_t^T + \mathbf{Q}_t$.
$\mathbf{K} \leftarrow \boldsymbol{\Sigma}_{t+1|t} \mathbf{H}_t (\mathbf{H}_t \boldsymbol{\Sigma}_{t+1|t} \mathbf{H}_t^T + \mathbf{R}_t)^{-1}$.
$\boldsymbol{\mu}_{t+1|t+1} \leftarrow \boldsymbol{\mu}_{t+1|t} + \mathbf{K}(\mathbf{y}_{t+1} - \mathbf{H}_t \boldsymbol{\mu}_{t+1|t})$.
$\boldsymbol{\Sigma}_{t+1|t+1} \leftarrow (\mathbf{I} - \mathbf{K}\mathbf{H}_t)\boldsymbol{\Sigma}_{t+1|t}$.
$\ell \leftarrow \ell + \log(\phi(\mathbf{y}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t}))$.
end for
Output: Log likelihood ℓ, means $\left\{\boldsymbol{\mu}_{t|t}\right\}_{t=1}^T$ and variances $\left\{\boldsymbol{\Sigma}_{t|t}\right\}_{t=1}^T$.

Algorithm 5
Rauch-Tung-Striebel Smoother

Input: Data $\{\mathbf{y}_t\}_{t=1}^T$, inital mean $\boldsymbol{\mu}_0$ and variance $\boldsymbol{\Sigma}_0$, model parameters $\{\mathbf{u}_t\}_{t=0}^T$, $\{\mathbf{F}_t\}_{t=1}^T$, $\{\mathbf{H}_t\}_{t=1}^T$, $\{\mathbf{B}_t\}_{t=1}^T$ and filtering distribution means $\left\{\boldsymbol{\mu}_{t|t}\right\}_{t=1}$ and variances $\left\{\boldsymbol{\Sigma}_{t|t}\right\}_{t=1}^T$.
Initialise $t \leftarrow T$.
for $t \in \{T-1, \ldots, 2, 1\}$ **do**
$\boldsymbol{\mu}_{t+1|t} \leftarrow \mathbf{F}_{t+1} \boldsymbol{\mu}_{t|t} + \mathbf{B}_{t+1} \mathbf{u}_t$.
$\boldsymbol{\Sigma}_{t+1|t} \leftarrow \mathbf{F}_{t+1} \boldsymbol{\Sigma}_{t|t} \mathbf{F}_{t+1}^T + \mathbf{Q}_t$.
$\mathbf{G} \leftarrow \boldsymbol{\Sigma}_{t|t} \mathbf{F}_{t+1}^T \boldsymbol{\Sigma}_{t+1|t}^{-1}$.
$\boldsymbol{\mu}_{t|T} \leftarrow \boldsymbol{\mu}_{t|t} + \mathbf{K}(\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{t+1|t})$.
$\boldsymbol{\Sigma}_{t|T} \leftarrow \boldsymbol{\Sigma}_{t|t} + \mathbf{G}(\boldsymbol{\Sigma}_{t+1|T} - \boldsymbol{\Sigma}_{t+1|t})\mathbf{G}^T$.
end for
Output: Smoothed distribution means $\left\{\boldsymbol{\mu}_{t|T}\right\}_{t=1}^T$ and variances $\left\{\boldsymbol{\Sigma}_{t|T}\right\}_{t=1}^T$.

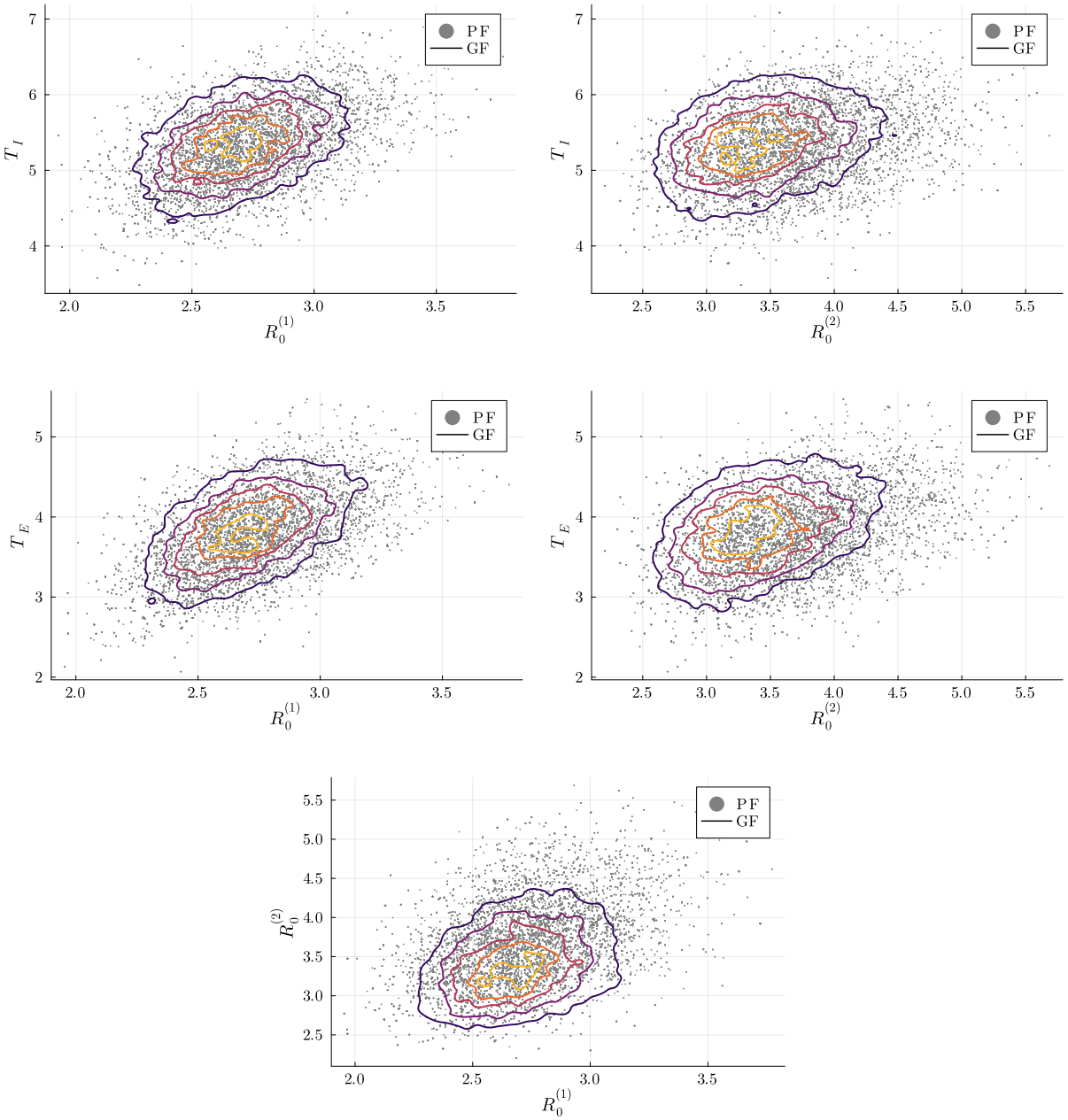


Figure 6.1: Plots showing samples from bivariate posterior distributions. Samples produced using the PF method are shown as points. The samples produced using the GF method are shown as kernal density estiamtes.

Bibliography

[1] Jocelyne Piret and Guy Boivin. Pandemics Throughout History. *Frontiers in Microbiology*, 11, 2021.

[2] Authia Gray and Fablina Sharara. Global and regional sepsis and infectious syndrome mortality in 2019: A systematic analysis. *The Lancet Global Health*, 10:S2, 2022.

[3] Jan M. Brauner, Sören Mindermann, Mrinank Sharma, David Johnston, John Salvatier, Tomáš Gavenčiak, Anna B. Stephenson, Gavin Leech, George Altman, Vladimir Mikulik, Alexander John Norman, Joshua Teperowski Monrad, Tamay Besiroglu, Hong Ge, Meghan A. Hartwick, Yee Whye Teh, Leonid Chindelevitch, Yarin Gal, and Jan Kulveit. Inferring the effectiveness of government interventions against COVID-19. *Science*, 371(6531), 2021.

[4] Matteo Chinazzi, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, Cécile Viboud, Xinyue Xiong, Hongjie Yu, M. Elizabeth Halloran, Ira M. Longini, and Alessandro Vespignani. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 2020.

[5] Fred Brauer, Pauline Van Den Driessche, Jianhong Wu, J. M. Morel, F. Takens, and B. Teissier, editors. *Mathematical Epidemiology*. Lecture Notes in Mathematics. Springer, Berlin, Heidelberg, 2008.

[6] Fred Brauer. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling*, 2(2):113–127, 2017.

[7] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy S.M. Leung, Eric H.Y. Lau, Jessica Y. Wong, Xuesen Xing, Nijuan Xiang, Yang Wu, Chao Li, Qi Chen, Dan Li, Tian Liu, Jing Zhao, Man Liu, Wenxiao Tu, Chuding Chen, Lianmei Jin, Rui Yang, Qi Wang, Suhua Zhou, Rui Wang, Hui Liu, Yinbo Luo, Yuan Liu, Ge Shao, Huan Li, Zhongfa Tao, Yang Yang, Zhiqiang Deng, Boxi Liu, Zhitao Ma, Yanping Zhang, Guoqing Shi, Tommy T.Y. Lam, Joseph T. Wu, George F. Gao, Benjamin J. Cowling, Bo Yang, Gabriel M. Leung, and Zijian

Feng. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *The New England Journal of Medicine*, 382(13):1199–1207, 2020.

[8] Australian Government Department of Health. 2019-nCoV acute respiratory disease, Australia: Epidemiology Report 1 (Reporting week 26 January - 1 February 2020). *Communicable Diseases Intelligence (2018)*, 44, 2020.

[9] Freya M. Shearer, James Walker, Nefel Tellioglu, James M. McCaw, Jodie McVernon, Andrew Black, and Nicholas Geard. Rapid assessment of the risk of SARS-CoV-2 importation: Case study and lessons learned. *Epidemics*, 38, 2022.

[10] Pablo Martinez De Salazar, René Niehus, Aimee Taylor, Caroline O'Flaherty Buckee, and Marc Lipsitch. Identifying Locations with Possible Undetected Imported Severe Acute Respiratory Syndrome Coronavirus 2 Cases by Using Importation Predictions. *Emerging Infectious Diseases*, 26(7):1465–1469, 2020.

[11] Sangeeta Bhatia, Natsuko Imai, Gina Cuomo-Dannenburg, Marc Baguelin, Adhiratha Boonyasiri, Anne Cori, Zulma Cucunubá, Ilaria Dorigatti, Rich FitzJohn, Han Fu, Katy Gaythorpe, Azra Ghani, Arran Hamlet, Wes Hinsley, Daniel Laydon, Gemma Nedjati-Gilani, Lucy Okell, Steven Riley, Hayley Thompson, Sabine van Elsland, Erik Volz, Haowei Wang, Yuanrong Wang, Charles Whittaker, Xiaoyue Xi, Christl A. Donnelly, and Neil M. Ferguson. Estimating the number of undetected COVID-19 cases among travellers from mainland China. *Wellcome Open Research*, 5:143, 2020.

[12] Steven G. Krantz and Arni S. R. Srinivasa Rao. Level of underreporting including underdiagnosis before the first peak of COVID-19 in various countries: Preliminary retrospective results based on wavelets and deterministic modeling. *Infection Control & Hospital Epidemiology*, 41(7):857–859, 2020.

[13] Rupam Bhattacharyya, Ritoban Kundu, Ritwik Bhaduri, Debashree Ray, Lauren J. Beesley, Maxwell Salvatore, and Bhramar Mukherjee. Incorporating false negative tests in epidemiological models for SARS-CoV-2 transmission and reconciling with seroprevalence estimates. *Scientific Reports*, 11(1):9748, 2021.

[14] Danju Zhou, Michelle Pender, Weixi Jiang, Wenhui Mao, and Shenglan Tang. Underreporting of TB cases and associated factors: A case study in China. *BMC Public Health*, 19(1):1664, 2019.

[15] Ichiro Nakamoto and Jilin Zhang. Modeling the underestimation of COVID-19 infection. *Results in Physics*, 25:104271, 2021.

[16] Joseph T Wu, Kathy Leung, and Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet*, 395(10225):689–697, 2020.

[17] Rene Niehus, Pablo M. De Salazar, Aimee R. Taylor, and Marc Lipsitch. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. *medRxiv*, 2020.

[18] Robin A. Weiss and Anthony J. McMichael. Social and environmental risk factors in the emergence of infectious diseases. *Nature Medicine*, 10(12):70–76, 2004.

[19] Karla Therese L. Sy, Laura F. White, and Brooke E. Nichols. Population density and basic reproductive number of COVID-19 across United States counties. *PLOS ONE*, 16(4), 2021.

[20] Paulo R. Martins-Filho. Relationship between population density and COVID-19 incidence and mortality estimates: A county-level analysis. *Journal of Infection and Public Health*, 14(8):1087–1088, 2021.

[21] Peter D. Congdon. *Bayesian Hierarchical Models: With Applications Using R, Second Edition*. CRC Press, first edition, 2020.

[22] David T. Frazier, Christian P. Robert, and Judith Rousseau. Model Misspecification in Approximate Bayesian Computation: Consequences and Diagnostics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(2):421–444, 2020.

[23] Grant D. Brown, Aaron T. Porter, Jacob J. Oleson, and Jessica A. Hinman. Approximate Bayesian computation for spatial SEIR(S) epidemic models. *Spatial and Spatio-temporal Epidemiology*, 24:27–37, 2018.

[24] Joel Hellewell, Sam Abbott, Amy Gimma, Nikos I. Bosse, Christopher I. Jarvis, Timothy W. Russell, James D. Munday, Adam J. Kucharski, W. John Edmunds, Fiona Sun, Stefan Flasche, Billy J. Quilty, Nicholas Davies, Yang Liu, Samuel Clifford, Petra Klepac, Mark Jit, Charlie Diamond, Hamish Gibbs, Kevin van Zandvoort, Sebastian Funk, and Rosalind M. Eggo. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 8(4):488–496, 2020.

[25] Samir Bhatt, Neil Ferguson, Seth Flaxman, Axel Gandy, Swapnil Mishra, and James A Scott. Semi-Mechanistic Bayesian modeling of COVID-19 with Renewal Processes. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 2023.

[26] Jérôme Levesque, David W. Maybury, and R.H.A. David Shaw. A model of COVID-19 propagation based on a gamma subordinated negative binomial branching process. *Journal of Theoretical Biology*, 512, 2021.

[27] Zebin Zhao, Xin Li, Feng Liu, Gaofeng Zhu, Chunfeng Ma, and Liangxu Wang. Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. *Science of The Total Environment*, 729, 2020.

[28] Longxiang Su, Na Hong, Xiang Zhou, Jie He, Yingying Ma, Huizhen Jiang, Lin Han, Fengxiang Chang, Guangliang Shan, Weiguo Zhu, and Yun Long. Evaluation of the Secondary Transmission Pattern and Epidemic Prediction of COVID-19 in the Four Metropolitan Areas of China. *Frontiers in Medicine*, 7, 2020.

[29] Simo Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, first edition, 2013.

[30] Silvio Romano, Annalisa Fierro, and Antonella Liccardo. Beyond the peak: A deterministic compartment model for exploring the COVID-19 evolution in Italy. *PLOS ONE*, 15(11), 2020.

[31] José M. Carcione, Juan E. Santos, Claudio Bagaini, and Jing Ba. A Simulation of a COVID-19 Epidemic Based on a Deterministic SEIR Model. *Frontiers in Public Health*, 2020.

[32] Andrea Kvitkovičová and Victor M. Panaretos. Asymptotic inference for partially observed branching processes. *Advances in Applied Probability*, 43(4):1166–1190, 2011.

[33] Victor M. Panaretos. Partially observed branching processes for stochastic epidemics. *Journal of Mathematical Biology*, 54(5):645–668, 2007.

[34] Peter Jagers. *Branching Processes with Biological Applications*. Wiley, London ; New York, 1975.

[35] Marek Kimmel and David Springer. Branching Processes in Biology. 19, 2002.

[36] Jason Xu and Vladimir N. Minin. Efficient Transition Probability Computation for Continuous-Time Branching Processes via Compressed Sensing. *Uncertainty in artificial intelligence*, 2015:952–961, 2015.

[37] Koen Frenken and Ron A. Boschma. A theoretical framework for evolutionary economic geography: Industrial dynamics and urban growth as a branching process. *Journal of Economic Geography*, 7(5):635–649, 2007.

[38] Robert J. Elliott, Lakhdar Aggoun, and John B. Moore. *Hidden Markov Models: Estimation and Control*. Springer Science & Business Media, 2008.

[39] Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

[40] Andrew Golightly and Colin S. Gillespie. Simulation of Stochastic Kinetic Models. In Maria Victoria Schneider, editor, *In Silico Systems Biology*, Methods in Molecular Biology, pages 169–187. Humana Press, Totowa, NJ, 2013.

[41] Jason Xu, Peter Guttorp, Midori Kato-Maeda, and Vladimir N. Minin. Likelihood-Based Inference for Discretely Observed Birth-Death-Shift Processes, with Applications to Evolution of Mobile Genetic Elements. *Biometrics*, 71(4):1009–1021, 2015.

[42] Richard Durrett. Continuous Time Markov Chains. In Richard Durrett, editor, *Essentials of Stochastic Processes*, Springer Texts in Statistics, pages 147–200. Springer International Publishing, 2016.

[43] Tom Britton and Etienne Pardoux. Chapter 1 Single Population Epidemics. In Tom Britton and Etienne Pardoux, editors, *Stochastic Epidemic Models with Inference*, Lecture Notes in Mathematics, pages 125–141. Springer International Publishing, Cham, 2019.

[44] Linda J. S. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2):128–142, 2017.

[45] J. L. Doob. Markoff Chains–Denumerable Case. *Transactions of the American Mathematical Society*, 58(3):455–473, 1945.

[46] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4): 403–434, 1976.

[47] Rajesh Ramaswamy and Ivo F. Sbalzarini. A partial-propensity variant of the composition-rejection stochastic simulation algorithm for chemical reaction networks. *The Journal of Chemical Physics*, 132(4):044102, 2010.

[48] Michael Gibson and Jehoshua Bruck. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Physical Chemistry A*, 104, 2000.

[49] Sheldon M. Ross. Continuous-Time Markov Chains. In Sheldon M. Ross, editor, *Introduction to Probability Models (Tenth Edition)*, pages 371–419. Academic Press, Boston, 2010.

[50] Karin S. Dorman, Janet S. Sinsheimer, and Kenneth Lange. In the garden of branching processes. *SIAM Review*, 2004.

[51] Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, 1990.

[52] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[53] G. A. Young and R. L. Smith. *Essentials of Statistical Inference*. Cambridge University Press, Cambridge, UK, first edition, 2010.

[54] George E. P. Box and George C. Tiao. *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, 2011.

[55] Bruno De Finetti and Alberto Mura. *Philosophical Lectures on Probability*. Number v. 340 in Synthese Library : Studies in Epistemology, Logic, Methodology, and Philosophy of Science. Springer, Dordrecht, 2008.

[56] Dani Gamerman and Hedibert F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition*. CRC Press, 2006.

[57] Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao Li Meng. *Handbook of Markov Chain Monte Carlo*. 2011.

[58] Brian D. Ripley. *Stochastic Simulation*. Wiley, New York, first edition, 1987.

[59] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[60] Christophe Andrieu and Gareth O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725, 2009.

[61] Michael K. Pitt, Ralph dos Santos Silva, Paolo Giordani, and Robert Kohn. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134–151, 2012.

[62] Nicolas Chopin and Omiros Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer Series in Statistics. Springer International Publishing, New York City, 2020.

[63] Linda J. S. Allen. Applications of Multi-Type Branching Processes. In Linda J. S. Allen, editor, *Stochastic Population and Epidemic Models: Persistence and Extinction*, Mathematical Biosciences Institute Lecture Series, pages 21–27. Springer International Publishing, Cham, 2015.

[64] Darren J. Wilkinson. *Stochastic Modelling for Systems Biology, Third Edition*. CRC Press LLC, 2018.

[65] Anderson Gray McKendrick and William Ogilvy Kermack. 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(772):700–721, 1927.

[66] Linda J. S. Allen and Amy M. Burgin. Comparison of deterministic and stochastic SIS and SIR models in discrete time. *Mathematical Biosciences*, 163(1):1–33, 2000.

[67] Richard Hooper. Epidemic Modelling: An Introduction. *American Journal of Epidemiology*, 151(8):835–836, 2000.

[68] Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[69] Thomas House, Ashley Ford, Shiwei Lan, Samuel Bilson, Elizabeth Buckingham-Jeffery, and Mark Girolami. Bayesian uncertainty quantification for transmissibility of influenza, norovirus and Ebola using information geometry. *Journal of The Royal Society Interface*, 13(121), 2016.

[70] Marc Baguelin, Albert Jan Van Hoek, Mark Jit, Stefan Flasche, Peter J. White, and W. John Edmunds. Vaccination against pandemic influenza A/H1N1v in England: A real-time economic evaluation. *Vaccine*, 28(12):2370–2384, 2010.

[71] George P. Yanev. Exponential and Hypoexponential Distributions: Some Characterizations. *Mathematics*, 8(12):2207, 2020.

[72] Frank Ball and Peter Donnelly. Strong approximations for epidemic models. *Stochastic Processes and their Applications*, 55(1):1–21, 1995.

[73] Patsy Haccou, Patricia Haccou, Peter Jagers, Vladimir A. Vatutin, and Vladimir Vatutin. *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press, 2005.

[74] A. Doucet, M. K. Pitt, G. Deligiannidis, and R. Kohn. Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika*, 102(2):295–313, 2015.

[75] Victoria Department of Health. Victorian COVID-19 data | Coronavirus Victoria. http://www.coronavirus.vic.gov.au/victorian-coronavirus-covid-19-data, 2022.

[76] Pierre Nouvellet, Anne Cori, Tini Garske, Isobel M. Blake, Ilaria Dorigatti, Wes Hinsley, Thibaut Jombart, Harriet L. Mills, Gemma Nedjati-Gilani, Maria D. Van Kerkhove, Christophe Fraser, Christl A. Donnelly, Neil M. Ferguson, and Steven Riley. A simple approach to measure transmissibility and forecast incidence. *Epidemics*, 22:29–35, 2018.

[77] Hiroshi Nishiura and Gerardo Chowell. The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends. In *Mathematical and Statistical Estimation Approaches in Epidemiology*, pages 103–121. 2009.

[78] Lin Zhang, Haochen Wang, Zhongyang Liu, Xiao Fan Liu, Xin Feng, and Ye Wu. A Heterogeneous Branching Process with Immigration Modeling for COVID-19 Spreading in Local Communities in China. *Complexity*, 2021.

[79] Rui Chen and Ollivier Hyrien. Quasi- and pseudo-maximum likelihood estimators for discretely observed continuous-time Markov branching processes. *Journal of Statistical Planning and Inference*, 141(7):2209–2227, 2011.

[80] Charles Doss, Marc Suchard, Ian Holmes, Midori Kato-maeda, and Vladimir Minin. Fitting Birth-Death Processes to Panel Data with Applications to Bacterial DNA Fingerprinting. *The Annals of Applied Statistics*, 7, 2010.

[81] Jonathan Fintzi, Jon Wakefield, and Vladimir N. Minin. A linear noise approximation for stochastic epidemic models fit to partially observed incidence counts. *Biometrics*, 78(4):1530–1541, 2022.

[82] Timothy W. Russell, Joseph T. Wu, Sam Clifford, W. John Edmunds, Adam J. Kucharski, and Mark Jit. Effect of internationally imported cases on internal spread of COVID-19: A mathematical modelling study. *The Lancet Public Health*, 6(1): 12–20, 2021.

[83] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *The Annals of Applied Probability*, 7(1): 110–120, 1997.

[84] Radford M. Neal. Monte Carlo Implementation. In Radford M. Neal, editor, *Bayesian Learning for Neural Networks*, Lecture Notes in Statistics, pages 55–98. Springer, New York, NY, 1996.

[85] Louis B. Rall, G. Goos, J. Hartmanis, W. Brauer, P. Brinch Hansen, D. Gries, C. Moler, G. Seegmüller, J. Stoer, and N. Wirth, editors. *Automatic Differentiation: Techniques and Applications*, volume 120 of *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 1981.

[86] Alison P. Galvani and Robert M. May. Dimensions of superspreading. *Nature*, 438 (7066):293–295, 2005.

[87] Zhuang Shen, Fang Ning, Weigong Zhou, Xiong He, Changying Lin, Daniel P. Chin, Zonghan Zhu, and Anne Schuchat. Superspreading SARS Events, Beijing, 2003. *Emerging Infectious Diseases*, 2004.

[88] Kyoungkon Kim, Erick T, Jae Choi, Monica Moon, and Min Soo Kim. Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, 2015: Epidemiology, characteristics and public health implications. *Journal of Hospital Infection*, 95, 2016.

[89] Xiao-Ke Xu, Xiao Fan Liu, Ye Wu, Sheikh Taslim Ali, Zhanwei Du, Paolo Bosetti, Eric H Y Lau, Benjamin J Cowling, and Lin Wang. Reconstruction of Transmission Pairs for Novel Coronavirus Disease 2019 (COVID-19) in Mainland China: Estimation of Superspreading Events, Serial Interval, and Hazard of Infection. *Clinical Infectious Diseases*, 2020.

[90] Yang Liu, Rosalind M. Eggo, and Adam J. Kucharski. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet*, 395(10227), 2020.

[91] Akira Endo, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Sam Abbott, Adam J. Kucharski, and Sebastian Funk. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, 5:67, 2020.