Efficient Bayesian parameter inference for COVID-19 transmission models

by

Renny Doig

B.Sc., Dalhousie University, 2019

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

© Renny Doig 2020
SIMON FRASER UNIVERSITY
Fall 2020

Copyright in this work rests with the author. Please ensure that any reproduction
or re-use is done in accordance with the relevant national copyright legislation.

Declaration of Committee

Name:	**Renny Doig**
Degree:	**Master of Science**
Thesis title:	**Efficient Bayesian parameter inference for COVID-19 transmission models**
Committee:	**Chair:** Joan Hu Professor, Statistics and Actuarial Science
	Liangliang Wang Supervisor Associate Professor, Statistics and Actuarial Science
	Lloyd Elliott Committee Member Assistant Professor, Statistics and Actuarial Science
	Jiguo Cao Examiner Professor, Statistics and Actuarial Science

Abstract

Many transmission models have been proposed and adapted to reflect changes in policy for mitigating the spread of COVID-19. Often these models are applied without any formal comparison with previously existing models. In this project, we use an annealed sequential Monte Carlo (ASMC) algorithm to estimate parameters of these transmission models. We also use Bayesian model selection to provide a framework through which the relative performance of transmission models can be compared in a statistically rigorous manner. The ASMC algorithm provides an unbiased estimate of the marginal likelihood which can be computed at no additional computational cost. This offers a significant computational advantage over MCMC methods which require expensive post hoc computation to estimate the marginal likelihood. We find that ASMC can produce results that are comparable to MCMC in a fraction of the time.

Keywords: COVID-19; Bayesian inference; sequential Monte Carlo; transmission models; model comparison

Acknowledgements

First of all, I would like to thank my supervisor Liangliang Wang. Her patience and guidance were invaluable in completing this research project. With her help I have been able to explore many new and interesting statistical methodologies which has made working with her consistently interesting and engaging. Her positive attitude has made completing this Masters program a very enjoyable experience; I look forward to continuing to work with her as I continue my studies.

Thank you to my committee members: Joan Hu, Jiguo Cao, and Lloyd Elliott for the insightful comments and interesting discussion during my thesis defence.

I would also like to thank my professors for providing engaging and challenging courses over the past year: Jinko Graham, Boxin Tang, Ian Bercovitz, Rachel Altman, and Tom Loughin. And thank you, to my fellow graduate students, who provided an entertaining and welcoming community at Simon Fraser. In particular I would like to thank James, Thompson, Zubia Mansoor, and Lisa McQuarrie who have provided great friendship and support. Thank you Lisa for your thorough help reviewing this thesis.

Finally, thank you to my parents, Peter and Karen, for their constant love and support, and to my grandmother Betty, whose daily chats have been an invaluable diversion.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The coronavirus disease 2019 (COVID-19) was first reported in the city of Wuhan, China in December 2019. Since then, this disease has spread throughout the world, which has resulted in over 1.1 million deaths worldwide as of November 1, 2020 [22]. In lieu of a vaccine, governments across the world have been implementing control measures, such as contact tracing or physical distancing, in an effort to slow the spread of COVID-19. Understanding the dynamics of disease transmission and how control measures can affect these dynamics is vital to ensuring that policy-makers implement well-informed decisions. As heavier control measures, such as business closures, can have serious impacts on the livelihood of individuals, it is important to ascertain if these measures will have the desired effect. Through disease transmission models, epidemiologists have been quantifying both physical characteristics of the coronavirus as well as the impacts of control measures on the spread of the disease.

Transmission models are compartmental meta-population models which separate stages of an infectious disease into several compartments. These compartments could represent stages such as an individual being infectious or being placed in quarantine. Often only a small subset of these compartments are observable. These models are an example of time-varying processes, called dynamic systems, which describe the spread of an infectious disease throughout a population. Dynamic systems are a system of differential equations (DEs) used to describe some physical process. These systems are often used in applied sciences because they provide structure and well-defined mechanisms for complex time-varying processes and allow for interaction between components of the system. Furthermore, the parameters characterizing these systems often have clear physical interpretations. For these reasons, dynamic systems are used in many fields, including epidemiology.

Due to their popularity in the applied sciences, estimating the system parameters has been a frequent area of statistical research. The simplest process for estimating these parameters is to find the least squares estimates. When the DE system is linear, the system has an analytic solution and the least squares estimates can easily be obtained. However, often the system does not have an analytic solution and thus an iterative non-linear least squares (NLS) algorithm is required. This iterative algorithm requires numerically solving the DE

system many times and can incur a large computational toll. Additionally, when only a subset of the states are observed or when the observations are subject to large measurement noise the resulting parameter estimates can be very poor [15]. Huang et al. propose a Bayesian model for dynamic systems in which they sample from the posterior through a combination of Gibbs sampling and Hamiltonian Monte Carlo. However, these traditional Monte Carlo techniques are often the victim of poor mixing and long computation times. To that end, we use annealed sequential Monte Carlo (ASMC) to sample from the posterior distribution. ASMC is an efficient Monte Carlo technique that does not suffer from the poor mixing of traditional Monte Carlo methods and has an algorithm which can easily be run in parallel.

Sequential Monte Carlo (SMC) is a class of Monte Carlo techniques which are based on importance sampling and resampling algorithms. SMC was originally designed for inference on sequences of distributions defined on measurable spaces of increasing dimension [10, 19]. The SMC framework was extended for sampling from sequences of distributions with a common measurable space by Del Moral (2006) [6]; it is to this class of SMC samplers that ASMC belongs. Due to the design of the SMC algorithm, it is very parallelizable, which can lead to major improvements in computation time compared to other Monte Carlo methods. ASMC has been implemented by Wang et al. (2019) [26] for estimating the posterior distribution of phylogenetic trees. Another advantage of the SMC family of algorithms is that the marginal likelihood is estimated during the sampling procedure. Traditional Monte Carlo techniques often require some post hoc sampling procedure to estimate the marginal likelihood. In this project, we use an ASMC algorithm to obtain estimates of transmission model parameters. From these parameter estimates we can estimate the underlying number of active cases in the population. We will also use the marginal likelihood estimates to perform Bayesian model comparison as a means of comparing transmission models.

Model selection is a common problem in statistics where multiple models for a process are compared to determine which model provides the best fit to the observed data. This project focuses on the context of transmission models for COVID-19. Due to the urgency and intensity of research on COVID-19, many new complex transmission models have been proposed. However, formal model assessment is seldom performed to quantitatively assess whether the complexities of these models are necessary. We consider transmission models designed to reflect the effects of physical distancing and other control measures. Often these models are proposed and justified by the expertise of the researcher, but rarely are they rigorously compared to simpler models or previously proposed models explaining similar phenomenon. For that reason, one focus of this project is to provide a framework through which epidemiological researchers can compare proposed transmission models.

In Section 2 we will introduce some examples of transmission models which have been used to model the effect of control measures, as well as models used to explain randomness in the data collection process. These two types of models will comprise the likelihood of

our data. In Section 3 we introduce the annealed SMC algorithm in detail. Section 4 introduces the Bayesian model comparison framework as well as two methods for estimating the marginal likelihood. In addition to detailing the ASMC estimate of the marginal likelihood we introduce one way in which the marginal likelihood could be estimated from the results of a Markov chain Monte Carlo. In Section 5 we perform a simulation study in which we demonstrate our proposed method's ability to estimate model parameters and active cases, as well as the performance of the model selection. An analysis of British Columbia case counts is performed in Section 6. Finally, a discussion of our results and potential future directions is given in Section 7.

Chapter 2

Modeling Infection Data

In this section we separate the likelihood of our data into two separate components: the transmission model and the observation model. The transmission model governs the dynamics of the underlying infections in the population; this component is deterministic. The observation model describes how the underlying states are related to the observed data being used to inform our inference; this component is probabilistic.

2.1 Transmission models

In epidemiology, compartmental transmission models are used to understand the spread of a disease at the population level. These models are defined by deterministic dynamics which govern how the disease spreads. The compartments of these models correspond to stages of a disease, such as when an individual is susceptible to the disease, when they can spread the disease, etc. In this section we introduce four examples of transmission models which can be used to model COVID-19.

2.1.1 SEIR model

One of the basic transmission models is the SEIR model which is composed of four compartments: individuals who are susceptible (S) to the disease, those who have been exposed (E) to the disease but are not yet symptomatic or infectious, infected (I) individuals who express symptoms and can spread the disease, and removed (R) individuals who are either recovered with immunity or dead. A delay between contact with infectious individuals and displaying symptoms has been noted for COVID-19 [1, 23, 27], which has made the SEIR model a foundation for many of the transmission models that have been developed in recent months. Note that for this simple model we assume that becoming symptomatic and becoming infectious are equivalent events.

The movement of individuals between these four compartments over time is governed by the system of ODEs in Equation 2.1; a visual representation can be seen in Figure 2.1.

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\frac{\beta SI}{N} \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \frac{\beta SI}{N} - \delta E \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \delta E - \gamma I \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma I.
\end{aligned} \tag{2.1}$$

In the above equations, N is the total population size, β is the transmission rate of the disease, δ is the rate at which individuals develop symptoms after becoming infected, and γ is the removal rate. We can interpret $1/\delta$ as the average length of time between when an individual first comes into contact with the disease and when they begin showing symptoms; this is called the incubation period or latent period. Similarly, we can interpret $1/\gamma$ as the average duration of infection.

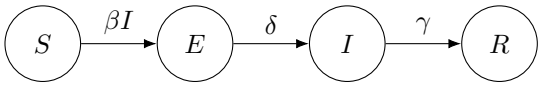

Figure 2.1: Visual representation of the SEIR transmission model.

Another quantity which we can derive from this model's parameters is the basic reproduction number, R_0, which has received much attention in the media. The definition of R_0 is the expected number of secondary infections caused by a single infectious individual in a wholly susceptible population. In the SEIR model, it is defined as $R_0 = \beta/\gamma$. The reason that this parameter has received so much attention is that it can be used as a benchmark for determining whether a disease will cause a pandemic. If $R_0 < 1$, then the system reaches a "disease-free" equilibrium where the disease dies out. On the other hand, if $R_0 > 1$, then such an equilibrium does not occur and the disease will spread rapidly throughout the population unless something changes to affect the transmission dynamics. The rapidity of disease spread will depend on how much greater than 1 R_0 is. This concept is closely related to the *force of infection*, which is defined as the rate at which susceptible individuals become infected and in this model can be written as $\beta\frac{I}{N}$.

There are a few limitations to this model. As is common in meta-population transmission models, we assume that an individual is equally likely to come into contact with any other individual in the population. This is called *homogeneous mixing*. As mentioned earlier, one of the major simplifying assumptions of the SEIR model is that expressing symptoms and being infectious are equivalent. More sophisticated models can allow for asymptomatic or pre-symptomatic transmission of the disease. We also do not incorporate birth or death rates

into the model. This can easily be done [18], but is not necessary for modelling the spread of COVID-19 as the number of infected individuals is low relative to the total population. Since this model does not differentiate between death and recovery, we could not use it directly to infer the infection fatality rate.

2.1.2 Simple physical distancing model

In an effort to mitigate the spread of COVID-19, many governments have introduced physical distancing measures to reduce the amount of person-to-person contact that takes place. To incorporate these measures into the transmission dynamics, an additional parameter is used to slow the rate of infection. The ODE system which defines this model is shown in Equation 2.2. It was originally proposed for modeling the cases in India by Patrikar et al. [23]. The parameter $f \in [0, 1]$ represents a proportional reduction in the force of infection as a result of physical distancing measures. For example, if $f = 0.5$ then the force of infection is half of what it would be with no physical distancing. When there is no distancing, $f = 1$ and we recover the SEIR model. On the other hand, when $f = 0$ there is perfect distancing and the infection cannot spread.

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= -f\beta I\frac{S}{N} \\ \frac{\mathrm{d}E}{\mathrm{d}t} &= f\beta I\frac{S}{N} - \delta E \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \delta E - \gamma I \\ \frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma I \end{aligned} \tag{2.2}$$

2.1.3 Full physical distancing model

Having introduced a simple SEIR-type model which incorporates the effects of physical distancing measures, we now introduce a more complex model which accounts for these measures in a similar way. This model was introduced by Anderson et al. [1] to model the spread of infection in British Columbia. The features that differentiate this transmission model from the two previous models are pre-symptomatic infectiousness, quarantine following diagnosis, and partial adherence to physical distancing measures.

To incorporate the first two features, the authors expand the exposed (E) and infectious (I) compartments from the previous model into four compartments, E_1, E_2, I, and Q. E_1 represents the period where an individual has been infected, but cannot infect others, equivalent to E in the previous models. Then an individual progresses into the E_2 compartment where they do not show symptoms, but are able to infect others; individuals move from E_1 to E_2 with rate k_1. Then individuals progress into the symptomatic and infectious compartment, I, with rate k_2. At this point, an infected individual has two paths: they

are diagnosed and enter quarantine, which occurs with rate q, or they remain undetected and move directly into the removed compartment. The difference between I and Q is that individuals in I can infect others, while individuals in Q cannot. Movement from either Q or I into R is done with rate $1/D$, where D is the expected duration of infection. Note that a caveat of the pre-symptomatic infectious (E_2) compartment is that an individual will at some point transition into the I compartment; this model will not account for asymptomatic infectiousness.

The third feature, the partial adherence to physical distancing measures, is admitted by creating a "distanced" counterpart of each of the compartments. The portion of the population that is not practicing physical distancing fall into the non-distanced compartments, X, and the portion that do practice these measures fall into the distanced compartments, X_d. In this way, the model can not only capture the effects of physical distancing, but also the rate at which the population adheres to it. The effect of physical distancing is captured with the same mechanism as the simple physical distancing model, through a parameter $f \in [0, 1]$. In the simple model, f was the proportion by which physical distancing measures reduced the force of infection. In this model, that reduction is only applied to the contribution from distanced infectious individuals to the force of infection. Note that in the first line of Equation 2.3 the effect of distancing, f, is only applied to the I_d and E_{2d} terms. Similarly, susceptible individuals who are practicing physical distancing will experience a two-fold reduction in the force of infection (see the first line of Equation 2.4). The first effect is a result of the reduced contact of distancing susceptible individuals to everyone else. The second effect is the reduced contact of distancing infectious individuals who have also reduced their contact with others. Movement between a compartment and its distanced analog, or vice versa, is permitted in any compartment. Individuals will enter the distanced compartments with rate u_d and leave with rate u_r. Once infected, the evolution of the disease, i.e. the rates at which individuals move between compartments, is the same regardless of whether an individual is practicing physical distancing or not. For example, two infectious individuals, one in I and one in I_d, will both recover at rate $1/D$ and enter quarantine with rate q. In the long run, the system will settle on a fraction $\frac{u_d}{u_d+u_r}$ of the population which are practicing physical

distancing. The dynamics of the non-distancing compartments are defined by

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\beta\left[I + E_2 + f(I_\mathrm{d} + E_{2\mathrm{d}})\right]\frac{S}{N} - u_\mathrm{d}S + u_r S_\mathrm{d} \\
\frac{\mathrm{d}E_1}{\mathrm{d}t} &= \beta\left[I + E_2 + f(I_d + E_{2\mathrm{d}})\right]\frac{S}{N} - k_1 E_1 - u_\mathrm{d}E_1 + u_r E_{1\mathrm{d}} \\
\frac{\mathrm{d}E_2}{\mathrm{d}t} &= k_1 E_1 - k_2 E_2 - u_\mathrm{d}E_2 + u_r E_{2\mathrm{d}} \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= k_2 E_2 - qI - \frac{I}{D} - u_\mathrm{d}I + u_r I_\mathrm{d} \\
\frac{\mathrm{d}Q}{\mathrm{d}t} &= qI - \frac{Q}{D} - u_\mathrm{d}Q + u_r Q_\mathrm{d} \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \frac{I}{D} + \frac{Q}{D} - u_\mathrm{d}R + u_r R_\mathrm{d}
\end{aligned} \tag{2.3}$$

and the dynamics of distancing individuals is given by

$$\begin{aligned}
\frac{\mathrm{d}S_\mathrm{d}}{\mathrm{d}t} &= -f\beta\left[I + E_2 + f(I_\mathrm{d} + E_{2\mathrm{d}})\right]\frac{S_\mathrm{d}}{N} + u_\mathrm{d}S - u_r S_\mathrm{d} \\
\frac{\mathrm{d}E_{1\mathrm{d}}}{\mathrm{d}t} &= \ f\beta\left[I + E_2 + f(I_d + E_{2\mathrm{d}})\right]\frac{S_\mathrm{d}}{N} - k_1 E_{1\mathrm{d}} + u_\mathrm{d}E_1 - u_r E_{1\mathrm{d}} \\
\frac{\mathrm{d}E_{2\mathrm{d}}}{\mathrm{d}t} &= k_1 E_{1\mathrm{d}} - k_2 E_{2\mathrm{d}} + u_\mathrm{d}E_2 - u_r E_{2\mathrm{d}} \\
\frac{\mathrm{d}I_\mathrm{d}}{\mathrm{d}t} &= k_2 E_{2\mathrm{d}} - qI_\mathrm{d} - \frac{I_\mathrm{d}}{D} + u_\mathrm{d}I - u_r I_\mathrm{d} \\
\frac{\mathrm{d}Q_\mathrm{d}}{\mathrm{d}t} &= qI_\mathrm{d} - \frac{Q_\mathrm{d}}{D} + u_\mathrm{d}Q - u_r Q_\mathrm{d} \\
\frac{\mathrm{d}R_\mathrm{d}}{\mathrm{d}t} &= \frac{I_\mathrm{d}}{D} + \frac{Q_\mathrm{d}}{D} + u_\mathrm{d}R - u_r R_\mathrm{d}.
\end{aligned} \tag{2.4}$$

A visualisation of this system can be seen in Figure 2.1.3.

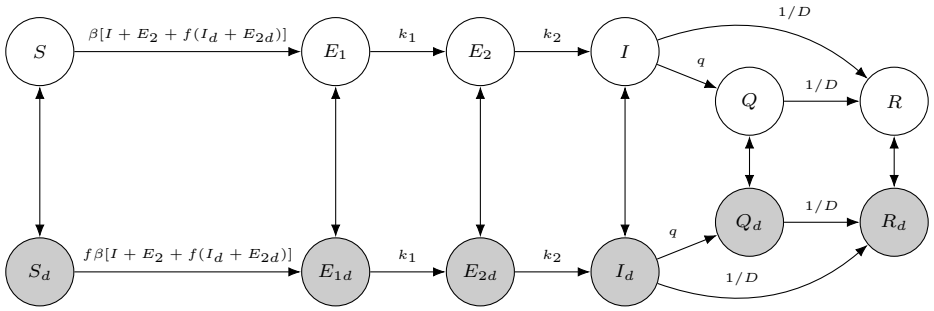

Figure 2.2: Visual representation of the physical distancing transmission model developed by [1].

2.1.4 Contact tracing model

The final transmission model we consider was used to model the effects of physical distancing and contact tracing measures in Ontario, Canada [27]. This model defines the dynamics of government interventions differently than the previous models. The primary differences between this model and previous models is that it accounts for contact tracing and asymptomatic infectiousness.

Starting from the susceptible (S) compartment, ν is the proportion of contacts between infectious and susceptible individuals that result in infection and c is the rate at which these contacts occur. Note that in the previous models the infection rate β was used to control the infectiousness of the disease, whereas in this model it is defined through a proportion. So, ν controls the infectiousness of the disease, where ν close to 1 represents a very infectious disease, and c controls how frequently people interact. An increase in physical distancing measures would be reflected in this model by decreasing c.

Once infected, individuals are in the exposed (E) compartment for a duration of $1/\sigma$. A proportion, ρ, of those individuals become symptomatic infectious (I) and $1-\rho$ become asymptomatic infectious (A). While the full physical distancing model allowed for pre-symptomatic infectiousness, this is the only model we introduce that allows for individuals to go from susceptible to removed without displaying symptoms. Asymptomatic individuals are assumed to recover at rate γ_A. Symptomatic individuals either recover directly from this compartment at rate γ_I, or they are admitted to a hospital (H) at rate δ_I. They also impose that a fraction θ of the asymptomatic individuals will be able to infect others.

Another feature taken into account in this transmission model is the effect of contact tracing. This assumes that a proportion ω of the symptomatic individuals who came into contact with infectious individuals are identified via contact tracing and sent into quarantine. Individuals who were not infected move into a susceptible quarantine compartment (S_q) where they stay for duration $1/\lambda$. Traced individuals who were infected move into an exposed quarantine compartment (E_q), until they show symptoms. Once they begin showing symptoms they are moved to the hospital (H); this occurs at rate δ_q. Individuals in the hospital recover at rate γ_H. The dynamics are defined in Equation 2.5 and a visualization is shown in Figure 2.3. Note that this models allows for different duration of infection ($\delta_\cdot$) depending on whether an individual is symptomatic, asymptomatic, or hospitalized.

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -(\nu c + c\omega(1-\nu))S(I+\theta A)/N + \lambda S_q \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \nu c(1-\omega)S(I+\theta A)/N - \sigma E \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \sigma\rho E - (\delta_I + \alpha + \gamma_I)I \\
\frac{\mathrm{d}A}{\mathrm{d}t} &= \sigma(1-\rho)E - \gamma_A A \\
\frac{\mathrm{d}S_q}{\mathrm{d}t} &= (1-\nu)c\omega S(I+\theta A)/N - \lambda S_q \\
\frac{\mathrm{d}E_q}{\mathrm{d}t} &= \nu c\omega S(I+\theta A)/N - \delta_q E_q \\
\frac{\mathrm{d}H}{\mathrm{d}t} &= \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma_I I + \gamma_A A + \gamma_H H
\end{aligned}
\tag{2.5}
$$

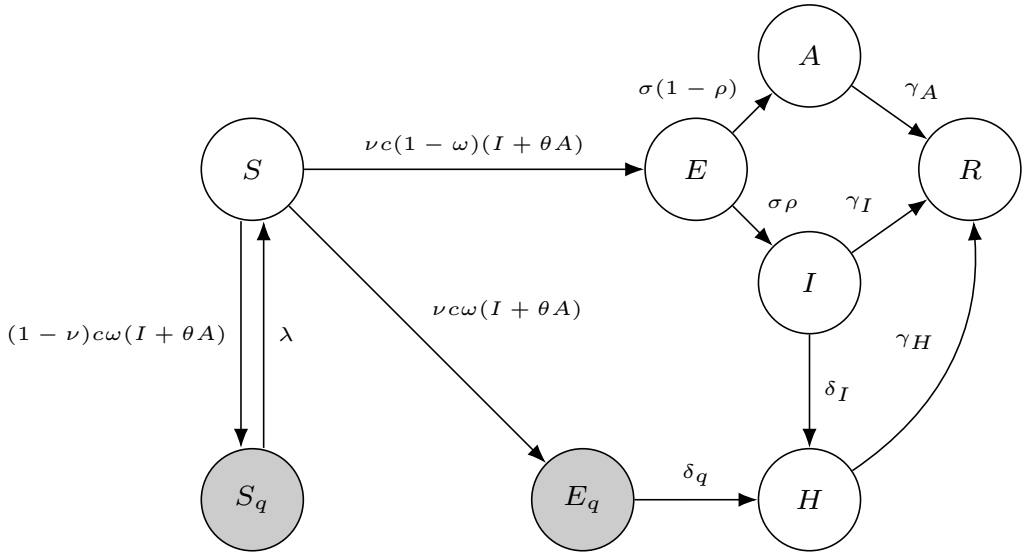

Figure 2.3: Visual representation of the contact tracing model developed by [27].

2.2 Observation models

In this section, we introduce two examples of models which relate the solution of the transmission model to the observed daily cases by accounting for measurement uncertainty. Not all epidemiological models include an observation model. For example, in Wu et al. [27], the transmission model parameters are chosen to minimize the squared error loss between the cumulative reported cases and the cumulative number of hospitalizations (H) per their

transmission model. However, in this project we assume an observation model which accounts for randomness in the data collection process. For that purpose, we introduce two models: a standard model which makes observations directly on the underlying transmission model compartments and a more complicated model that reflects certain features of data collection for COVID-19 in British Columbia.

2.2.1 Standard count distributions

Commonly, these transmission models are fit to count data, such as daily or cumulative case counts. As a result, the Poisson and negative binomial distributions, which are often used to model count data, are used to reflect measurement error in these types of infection data. Let us denote the number of observed cases at time t as $y(t)$. Assuming the measurement error is Poisson distributed yields $y(t) \sim \text{Poisson}(\mu_t)$, where μ_t is the expected number of cases. Similarly, assuming a negative binomial distribution over the case counts would give $y(t) \sim \text{NB}(\mu_t, \psi)$, where ψ is the dispersion parameter such that $\mathbb{V}ar(y(t)) = \mu_t + \frac{\mu_t^2}{\psi}$.

The mean number of counts, μ_t, is some function of the solution to the transmission model which relates the underlying infection to the observations. In the case of the SEIR model, we could have $\mu_t = I(t)$. For the full physical distancing model we would be making observations about all symptomatic individuals and therefore $\mu_t = I(t)+I_d(t)$. Furthermore, we could scale these by some fraction p to reflect systematic under-reporting of cases as we have seen with COVID-19 [1, 25]. E.g. for the SEIR model, $\mu_t = p \cdot I(t)$.

2.2.2 Observation delay model

The second observation model we discuss is another innovation of Anderson et al. [1], which accounts for a delay in reporting cases. The authors note that in British Columbia there is a lag between symptom onset in an individual and when they are reported in the confirmed case counts. To account for this they define the mean of their negative binomial observation model, μ_t, to be an average of the lagged *active cases* and assign a Weibull distribution over the lag times. They define active cases as being $I + I_d$, the total number of symptomatic individuals. The formula for the mean is given by

$$\mu_t = p \int_0^M k_2(E_2(t-s) + E_{2d}(t-s))w(s)\mathrm{d}s.$$

As with the standard observation model, $p \in (0, 1]$ is used to reflect that only a fraction of the underlying active cases are available to be observed due to under-reporting. $k_2(E_2(t-s) + E_{2d}(t-s)$ is the number of individuals who begin showing symptoms s days prior to time t. We assume that this lag, s, follows a Weibull distribution with shape and scale parameters w_{shape} and w_{scale}, respectively. The upper bound of the integral, M, denotes some upper bound on the delay, imposed for computational convenience.

Chapter 3

Estimating Model Parameters

In this section we discuss how we estimate the transmission and observation model parameters. To do that we introduce a Bayesian framework for a system of ODEs, which can be directly applied to the transmission models seen in the previous section. Then we discuss both Markov chain Monte Carlo and annealed sequential Monte Carlo, which we will be using to sample from the posterior distributions.

3.1 Bayesian framework of an ODE system

In the general setting of systems of ODEs we are interested in K states $\mathbf{x}(t) = [x_1(t), \ldots, x_K(t)]^T$ which evolve as functions of time, t. We describe the evolution of these states throughout time by the system of equations

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{g}(\mathbf{x}(t)|\boldsymbol{\beta}, \mathbf{x}_0),$$

where $\mathbf{g} = [g_1(\mathbf{x}), \ldots, g_K(\mathbf{x})]^T$ is a vector of functions which describe the dynamics of each state as functions of the whole system. These functions are characterized by a set of parameters, $\boldsymbol{\beta}$, and a vector of initial states, $\mathbf{x}_0$. In many settings both $\boldsymbol{\beta}$ and $\mathbf{x}_0$ are treated as parameters which need to be inferred.

Often, in practice we only observe a subset of states $K_0 \subseteq \{1, \ldots, K\}$. So in this model we are allowed to observed $|K_0| \leq K$ states, where $|\cdot|$ denotes the cardinality of the set. Here we assume that we make observations on those K_0 states at the same n discrete time points, $t_1, \ldots, t_n$. We denote the n observations on the k-th state as $\mathbf{y}_k = [y_k(t_1), \ldots, y_k(t_n)]^T$. We allow the observations to follow a general distribution with density p_{obs}. The j-th observation on the k-th state is assumed to follow

$$y_{kj} \sim p_{obs}(x_k(t_j|\boldsymbol{\beta}, \mathbf{x}_0)|\boldsymbol{\eta}),$$

where $\boldsymbol{\eta}$ are a set of parameters that characterize the likelihood. Note that both $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ can be time-dependent, but we have suppressed the notation here for simplicity. So the

likelihood of the entire ODE system is given by

$$\mathcal{L}(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{x}_0) = \prod_{k \in K_0} \prod_{j=1}^{n} p_{obs}(y_{kj}|x_k(t_j|\boldsymbol{\beta}, \mathbf{x}_0), \boldsymbol{\eta}).$$

For example, consider a SEIR transmission model where we only have observations on I. The observations are assumed to follow a negative binomial observation model with the mean defined as $\mu_t = I(t)$. Here we would have $K = 4$ states in our system, but we only make observations on $I(t)$, so $K_0 = \{3\}$. Our transmission model parameters are $\boldsymbol{\beta} = [\beta, \delta, \gamma]^T$ and our observation model parameters are $\boldsymbol{\eta} = [p, \psi]^T$. For the initial states, we need only to estimate $E(0)$ and $I(0)$. We can assume that there are no recovered individuals at the start of an outbreak and due to the closed population we can compute $S(0) = N - E(0) - I(0)$. We would therefore express the likelihood of a single observation as

$$p_{obs}(y_{kj}|x_k(t_j; \boldsymbol{\beta}, \mathbf{x}_0), \boldsymbol{\eta}) = \frac{\Gamma(\psi + y_{kj})}{y_{kj}!\Gamma(\psi)} \left(\frac{\psi}{\psi + p \cdot x_k(t_j; \boldsymbol{\beta}, \mathbf{x}_0)} \right)^{\psi} \left(\frac{p \cdot x_k(t_j; \boldsymbol{\beta}, \mathbf{x}_0)}{\psi + p \cdot x_k(t_j; \boldsymbol{\beta}, \mathbf{x}_0)} \right)^{y_{kj}}.$$

For convenience, we will compress the estimated quantities into a single vector, $\boldsymbol{\theta} = [\boldsymbol{\beta}^T, \boldsymbol{\eta}^T, \mathbf{x}_0^T]^T$.

It is important to note here that the likelihood depends on the underlying states of the system, or more specifically, the solution to the system conditional on the model parameters and initial states. In a setting where the ODEs are linear, then an analytic solution to the system can be derived and computing the likelihood is trivial. However, often this is not the case. When the ODE system is nonlinear, we must rely on numerical solvers to find the trajectory $\mathbf{x}(t)$ for a given set of parameters. There are many different numerical solvers; here we use an implementation of the fourth-order Runge-Kutta method, which is a popular method for such problems.

The Runge-Kutta method relies on a discretization of the dynamic process. Like the Euler method, the Runge-Kutta method uses Taylor expansions about certain points to obtain an estimate of $x(t + \Delta t)$, for some small time step Δt. In fact, the first-order Runge-Kutta method is the Euler method. The reason why the fourth-order Runge-Kutta method is many software packages' default approach for solving DEs is that it provides solutions that are accurate up to $O(\Delta t^4)$, while the Euler method is only accurate up to $O(\Delta t)$[21].

The equations for solving a single time step using a fourth-order Runge-Kutta method are

$$\begin{aligned} k_1 &= \Delta t \cdot g(x, t) \\ k_2 &= \Delta t \cdot g(x + \frac{1}{2}k_1, t + \frac{1}{2}\Delta t) \\ k_3 &= \Delta t \cdot g(x + \frac{1}{2}k_2, t + \frac{1}{2}\Delta t) \\ k_4 &= \Delta t \cdot g(x + k_3, t + \Delta t) \\ x(t + \Delta t) &= x(t) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4). \end{aligned}$$

In order to fully specify this as a Bayesian model, we also define the prior distribution over the transmission model parameters, observation model parameters, and the initial states, $\pi_0(\boldsymbol{\theta}) = \pi_0(\boldsymbol{\beta})\pi_0(\boldsymbol{\eta})\pi_0(\mathbf{x}_0)$. Then we can write the posterior distribution using Bayes rule:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\gamma(\boldsymbol{\theta})}{Z} = \frac{\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})}{Z},$$

where $\gamma(\boldsymbol{\theta})$ is the unnormalized posterior distribution and $Z = \int_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{y}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})$ is the marginal likelihood of the data. Often this integral is intractable, although it is not necessary in order to sample from $\pi(\boldsymbol{\theta}|\mathbf{y})$. In the upcoming sections we discuss how we can sample from the posterior distribution, as well as how we can estimate Z.

3.2 Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a well-researched and very popular method for sampling from posterior distributions. The algorithm uses a proposal distribution to generate a sequence of samples which construct an ergodic Markov chain that admits the posterior distribution, $\pi(\boldsymbol{\theta}|\mathbf{y})$, as its stationary distribution. To do this we start with some initial sample, $\boldsymbol{\theta}^{(0)}$. Then for the next $i = 1, \ldots, N$ samples we follow this procedure.

1. Propose a sample from the proposal distribution

   $$\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)}),$$

   where $q(\cdot)$ is a probability density with the same domain as $\pi(\boldsymbol{\theta}|\mathbf{y})$.

2. Compute the Metropolis-Hastings (MH) ratio

   $$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{(i-1)}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{(i-1)})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})}\right\} = \min\left\{1, \frac{\gamma(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{(i-1)}|\boldsymbol{\theta}^*)}{\gamma(\boldsymbol{\theta}^{(i-1)})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})}\right\},$$

   which accounts for the discrepancy between the proposal and target distributions.

3. α is the acceptance probability of $\boldsymbol{\theta}^*$. We either accept $\boldsymbol{\theta}^*$ and set $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^*$ with probability α, or reject it and set $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$ with probability $1 - \alpha$.

A common choice for the proposal distribution is $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)}) = \mathcal{N}(\boldsymbol{\theta}^{(i-1)}, \sigma)$. In this case, σ should be treated as a tuning parameter which is selected to yield a reasonable acceptance rate. In order to ensure that the Markov chain has converged to its stationary distribution, a number of samples at the beginning of the chain are discarded.

However, Markov chain Monte Carlo is not well-suited for Bayesian ODE sampling because the posterior surface can often be complicated with local maxima, or colinearity between parameters. This can lead to poor mixing and difficulty assessing the convergence of MCMC chains. In the next section we introduce a more advanced Monte Carlo method which is less affected by multimodality and, as a result, is better-suited for these types of problems.

3.3 Annealed Sequential Monte Carlo

Sequential Monte Carlo (SMC) is a Monte Carlo method which is derived from importance sampling (IS). It was initially developed as an efficient method for sampling from a sequence of distributions which are defined on measurable spaces of increasing dimension. The framework was adapted for sampling from sequences of distributions defined on a common measurable space by Del Moral in 2006 [6]. The particular variation we consider here is annealed SMC (ASMC), following Wang et al. (2019) [26].

The objective of ASMC is to generate a sample of weighted particles which represent the posterior distribution, $\pi(\boldsymbol{\theta})$, known as the target distribution. To do this we construct a sequence of R artificial intermediate distributions, $\{\pi_r(\boldsymbol{\theta})\}_{0 \leq r \leq R}$. The initial distribution is some distribution which can easily be sampled from. In this application we set our initial distribution to be our prior distribution, $\pi_0(\boldsymbol{\theta})$. We define these distributions through a sequence of annealing parameters, $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_R = 1$. We define the r-th intermediate distribution to be $\pi_r(\boldsymbol{\theta}) = \pi_0(\boldsymbol{\theta})[\mathcal{L}(\boldsymbol{\theta})]^{\alpha_r}$, so that $\pi_R(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})$. By constructing the distributions in this way we allow the particles to explore the parameters space easily in the early iterations and gradually incorporate the complexity through the likelihood. It is this gradual incorporation of complexity that allows ASMC to explore multiple maxima simultaneously, whereas MCMC is liable to get stuck in local maxima for many iterations.

The sampling procedure starts with a sample of equally weighted particles drawn from the prior distribution, $\{\boldsymbol{\theta}_0^{(k)}, 1/K\}_{1 \leq k \leq K}$. At each intermediate step r we aim to generate a sample of weighted particles $\{\boldsymbol{\theta}_r^{(k)}, W_r^{(k)}\}_{1 \leq k \leq K}$, where $W_\cdot^{(k)}$ is the normalized weight of the k-th particle such that $\sum_{k=1}^K W_\cdot^{(k)} = 1$. In order to generate this sample we take the sample from the previous iteration, $\{\boldsymbol{\theta}_{r-1}^{(k)}, W_{r-1}^{(k)}\}_{1 \leq k \leq K}$ and propagate them with the Markov kernel $T_r(\boldsymbol{\theta}_{r-1}, \boldsymbol{\theta}_r)$. This kernel should be chosen such that it admits $\pi_r(\boldsymbol{\theta})$ as its stationary distribution. The resulting proposed particle follows the *importance distribution*,

$\eta_r(\boldsymbol{\theta}_r) = \int \eta_0 \prod_{i=1}^{r} T_i(\boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_i)$. In the standard SMC framework the unnormalised importance weights are computed as

$$w_{r,k} = \frac{\eta_r(\boldsymbol{\theta}_{r,k})}{\gamma_r(\boldsymbol{\theta}_{r,k})}. \tag{3.1}$$

However, the integral in η_n is not tractable. To circumvent this problem Del Moral et al. (2006) [6] introduce an artificial backwards kernel, $L_{r-1}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_{r-1})$. Using this artificial backwards kernel, we can define an artificial joint intermediate distribution,

$$\tilde{\gamma}(\boldsymbol{\theta}_{1:r}) = \gamma_r(\boldsymbol{\theta}_r) \prod_{i=1}^{r-1} L_i(\boldsymbol{\theta}_{i+1}, \boldsymbol{\theta}_i).$$

This gives us the following recursive expression for the unnormalized importance weights

$$\begin{aligned} w_r(\boldsymbol{\theta}_{1:r}) &= \tilde{\gamma}_r(\boldsymbol{\theta}_r)/\eta_r(\boldsymbol{\theta}_{1:r}) \\ &= w_{r-1}(\boldsymbol{\theta}_{1:r-1})\tilde{w}_r(\boldsymbol{\theta}_{r-1}, \boldsymbol{\theta}_r) \end{aligned}$$

where $\tilde{w}(\boldsymbol{\theta}_{r-1}, \boldsymbol{\theta}_r)$ is called the unnormalized incremental weight function, and can be written as

$$\tilde{w}_r(\boldsymbol{\theta}_{r-1}, \boldsymbol{\theta}_r) = \frac{\gamma_r(\boldsymbol{\theta}_r) L_{r-1}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_{r-1})}{\gamma_{r-1}(\boldsymbol{\theta}_{r-1}) T_r(\boldsymbol{\theta}_{r-1}, \boldsymbol{\theta}_r)}. \tag{3.2}$$

By introducing an artificial backwards kernel to avoid the intractable integration problem, we introduce variance to our importance weight estimates. This variance can be minimized by selecting an optimal backwards kernel,

$$L_{r-1}^{\text{opt}}(\boldsymbol{\theta}_r^{(k)}, \boldsymbol{\theta}_{r-1}^{(k)}) = \frac{\pi_r(\boldsymbol{\theta}_{r-1}^{(k)}) T_r(\boldsymbol{\theta}_r^{(k)}, \boldsymbol{\theta}_r^{(k)})}{\pi_r(\boldsymbol{\theta}_r^{(k)})}.$$

Substituting this optimal kernel back into the unnormalized incremental importance weight function in Equation 3.2 gives us a convenient reduction of the form

$$\begin{aligned} \tilde{w}_r^{(k)} &= \frac{\gamma_r(\boldsymbol{\theta}_r^{(k)}) L_r^{\text{opt}}(\boldsymbol{\theta}_r^{(k)}, \boldsymbol{\theta}_{r-1}^{(k)})}{\gamma_{r-1}(\boldsymbol{\theta}_{r-1}^{(k)}) T_r(\boldsymbol{\theta}_{r-1}^{(k)}, \boldsymbol{\theta}_r^{(k)})} \\ &= \frac{\gamma_r(\boldsymbol{\theta}_r^{(k)})}{\gamma_{r-1}(\boldsymbol{\theta}_{r-1}^{(k)})} \cdot \frac{\pi_r(\boldsymbol{\theta}_{r-1}^{(k)}) T_r(\boldsymbol{\theta}_{r-1}^{(k)}, \boldsymbol{\theta}_r^{(k)})}{\pi_r(\boldsymbol{\theta}_r^{(k)})} \cdot \frac{1}{T_r(\boldsymbol{\theta}_{r-1}^{(k)}, \boldsymbol{\theta}_r^{(k)})} \\ &= \frac{\gamma_r(\boldsymbol{\theta}_r^{(k)}}{\gamma_{r-1}(\boldsymbol{\theta}_{r-1}^{(k)})} \cdot \frac{\gamma_r(\boldsymbol{\theta}_{r-1}^{(k)})}{Z_r} \cdot \frac{Z_r}{\gamma_r(\boldsymbol{\theta}_r^{(k)})} \\ &= \frac{\gamma_r(\boldsymbol{\theta}_{r-1}^{(k)})}{\gamma_{r-1}(\boldsymbol{\theta}_{r-1}^{(k)})}. \end{aligned}$$

The SMC algorithm is comprised of three steps: *compute weights, propagate, resample.* First, an initial sample of equally weighted particles is generated, $\{\boldsymbol{\theta}_0^{(k)}, 1/K\}_{1 \leq k \leq K}$. We

use $\{\tilde{\boldsymbol{\theta}}_r^{(k)}\}$ to denote particles which have been resampled. We repeat each of the next three steps at each iteration $r = 1, \ldots, R$.

1. *Compute weights.* The normalized importance weights are computed for each of the K particles as

$$W_r^{(k)} \propto w_r^{(k)} = w_{r-1}^{(k)} \cdot \frac{\gamma_r(\tilde{\boldsymbol{\theta}}_{r-1}^{(k)})}{\gamma_{r-1}(\tilde{\boldsymbol{\theta}}_{r-1}^{(k)})} = w_{r-1}^{(k)} \mathcal{L}(\mathbf{y}|\tilde{\boldsymbol{\theta}}_{r-1}^{(k)})^{\alpha_r - \alpha_{r-1}}. \tag{3.3}$$

   Note that the importance weight formula (3.3) does not depend on the sample proposed in this iteration, only the particles from the previous step. This differs from the standard SMC algorithm [10] and is beneficial, as computing the importance weight before proposing new particles reduces the variance of the importance weight estimates.

2. *Propagate.* The next generation of samples, $\{\boldsymbol{\theta}_r^{(k)}\}_{1 \leq k \leq K}$, are propagated via π_r-invariant MCMC moves, $\{\boldsymbol{\theta}_r^{(k)} \sim T_r(\tilde{\boldsymbol{\theta}}_{r-1}^{(k)}, \cdot)\}_{1 \leq k \leq K}$. These propagation moves follow one iteration of the MH algorithm discussed in the previous section. It is important to note here that since the particles are propagated independently of one another this step can be done entirely in parallel, which gives ASMC has a notable computational advantage over MCMC.

3. *Resample.* This is the step which characterizes the SMC algorithm relative to other importance sampling based techniques. In the resampling step we aim to eliminate particles with low relative importance and propagate particles with greater importance. The result is K equally weighted particles $\{\tilde{\boldsymbol{\theta}}_r^{(k)}, 1/K\}_{1 \leq k \leq K}$. The simplest resampling algorithm is multinomial resampling, which performs sampling with replacement from the sample of particles with probabilities given by their unnormalized weights. However, this has been shown to produce weight estimates with higher variance than alternative resampling algorithms such as stratified or residual resampling [8]. In our implementation we use systematic resampling, which has been demonstrated to produce low variance estimates of the importance weights in practice, although no theory guarantees that this will always be the case.

The resampling step allows SMC to mitigate the issue of weight degeneracy found in sequential importance sampling. However, each time we resample we reduce the number of unique particles. This problem is known as *sample degeneracy* and is a well-documented issue with SMC [11]. Methods have been proposed such as resample-move [13] and block SMC [9] to reduce, but not solve, this issue. It is shown in [4] that each time resampling is performed, it increases the variance of our importance weight estimates. Therefore, we should only resample when the degeneracy in our weights reaches some threshold ϵ. One

common threshold is the effective sample size (ESS). The ESS measures the effective number of "perfect samples" from our target distribution, and can be estimated by

$$\widehat{\text{ESS}}_r = \frac{1}{\sum_{k=1}^{K}(W_r^{(k)})^2}.$$

The ESS takes on values between 1 and K. When ESS is 1 exactly one particle has all of the probability mass and ESS $= K$ when each particle has equal weight. A common strategy is to use $\epsilon = 0.5K$ as the resampling threshold.

After completing the R iterations of the ASMC algorithm we have a sample of weighted particles, $\{\boldsymbol{\theta}_R^{(k)}, W_R^{(k)}\}_{1\leq k\leq K}$, that can be used to approximate the posterior density through the formula

$$\hat{\pi}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \delta_{\boldsymbol{\theta}_R^{(k)}}(\boldsymbol{\theta}) W_R^{(k)},$$

where $\delta_{\boldsymbol{\theta}_R^{(k)}}(\boldsymbol{\theta})$ is a Dirac delta mass at $\boldsymbol{\theta}_R^{(k)}$. From this identity we can estimate posterior quantities, such as the posterior mean:

$$\widehat{\mathbb{E}_\pi[\boldsymbol{\theta}]} = \sum_{k=1}^{K} \boldsymbol{\theta}_R^{(k)} W_R^{(k)}.$$

We can also estimate features of the posterior distribution by resampling from the sample of weighted particles and computing estimates as we would with standard Monte Carlo methods. However, since resampling adds variance to our sample, quantities should be estimated using the weighted sample whenever possible.

3.4 Selecting the annealing parameters

Here we discuss how to select the annealing parameters that define our intermediate distributions. Consider the difference between subsequent annealing parameters, $\phi_r = \alpha_r - \alpha_{r-1}$. When ϕ_r is very large, π_r and π_{r-1} will be very different and as a result the importance weights will have high variance. On the other hand, if ϕ_r is very small then the sequence of intermediate distributions will be very long which will increase the computational cost. In this section we introduce an adaptive annealing parameter scheme which produces α_r such that the corresponding ϕ_r are small enough that the proposals are reasonable and not so small that it incurs unnecessary computational burden.

The adaptive annealing scheme we use here is based on the adaptive approximate Bayesian computation scheme developed by Del Moral et al. (2012) [7] and was used in the context of ASMC by [26]. It uses the notion of conditional effective sample size (CESS), which was introduced by [28]. The CESS quantifies how well a sample from π_{r-1} can be used to estimate the distribution π_r in a similar way to how ESS describes the quality of a sample of particles to estimate the target distribution. A simple modification of CESS was

used by [26], called the relative CESS (rCESS), which is CESS scaled to lie in $(1/K, 1)$:

$$\text{rCESS}_r(W_{r-1}^{(\cdot)}, \tilde{w}_r^{(\cdot)}) = \frac{(\sum_{k=1}^K W_{r-1}^{(k)} \tilde{w}_r^{(k)})^2}{\sum_{k=1}^K W_{r-1}^{(k)} (\tilde{w}_r^{(k)})^2}.$$

Recall from Equation 3.3 that $w_r^{(k)} = w_{r-1}^{(k)} \left[\mathcal{L}(\mathbf{y}|\tilde{\boldsymbol{\theta}}_{r-1}^{(k)})\right]^{\phi_r}$. There are two important comments to make about the form of the weight update. The first is that the weight update depends on r only through ϕ_r. This means that we can compute the importance weights at iteration prior to the sampling step, which will reduce the variance of our importance weight estimates. The second is that $w_r^{(k)}$ is a monotonically decreasing function of ϕ_r, which means that we can control ϕ_r, and consequently α_r, through the equation

$$f(\alpha) = \text{rCESS}(W_{r-1}^{(\cdot)}, \mathcal{L}(\mathbf{y}|\tilde{\boldsymbol{\theta}}_{r-1}^{(\cdot)})^{\alpha - \alpha_{r-1}}) = \phi, \tag{3.4}$$

where $\phi \in (0, 1)$ is a tuning parameter which controls how similar subsequent intermediate distributions are to each other. While in general we don't have a closed form solution to this problem, due to the monotonicity of $f(\alpha)$ we can use a bisection algorithm to find the α^* which satisfies Equation 3.4.

Chapter 4

Bayesian Model Comparison

4.1 Bayes factor

The problem of comparing a discrete set of Bayesian models is often approached with the Bayes factor [16]. Recall from Section 3.1 that we write the posterior distribution as

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{y}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}},$$

where $\int_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{y}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta}$ is called the marginal likelihood. The marginal likelihood can be interpreted as the probability of observing the data under likelihood $\mathcal{L}$. Consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$ with different likelihoods and corresponding sets of parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. In the context of this project, $\mathcal{M}_1$ and $\mathcal{M}_2$ would correspond to two competing transmission models. The Bayes factor comparing $\mathcal{M}_1$ to $\mathcal{M}_2$ is defined as the ratio of the respective marginal likelihoods,

$$B_{1,2} = \frac{p(\mathbf{y}|\mathcal{M}_1)}{p(\mathbf{y}|\mathcal{M}_2)} = \frac{\int \mathcal{L}(\mathbf{y}|\boldsymbol{\theta}_1, \mathcal{M}_1)\pi_0(\boldsymbol{\theta}_1|\mathcal{M}_1)}{\int \mathcal{L}(\mathbf{y}|\boldsymbol{\theta}_2, \mathcal{M}_2)\pi_0(\boldsymbol{\theta}_2|\mathcal{M}_2)}.$$

Values of $B_{1,2}$ greater than 1 suggest evidence in favour of model 1 over model 2. Unlike standard frequentist hypothesis testing procedures, which can only provide evidence against a null/in favour of an alternative, the Bayes factor can provide evidence in favour of either model. For the purpose of model selection with Bayes factors there is no theoretical threshold for determining when one model is better than another. However, Kass and Raftery [17] suggest a rule of thumb for gauging the evidence in favour of model 1 provided by $B_{1,2}$ which is given in Table 4.1. It should be noted that in order for the Bayes factor to be valid for model comparison the prior distributions in each model must be proper distributions.

As the Bayes factor often requires the computation of an intractable integral, it is not a trivial task to perform. We present next how the marginal likelihoods can be computed from ASMC as well as from MCMC.

$B_{1,2}$	Evidence in favour of $\mathcal{M}_1$
1 to 3	Weak
3 to 20	Positive
20 to 150	Strong
> 150	Very strong

Table 4.1: Guidelines for model selection from a Bayes factor suggested by Kass and Raftery [17].

4.2 Marginal likelihood estimate from ASMC

In ASMC, as with other importance sampling procedures, the marginal likelihood of the data, Z_R, can be estimated using the importance weights. In order to arrive at an estimate of Z_R, we first rely on estimates of the sequential ratios of normalizing constants,

$$\frac{Z_r}{Z_{r-1}} = \frac{\int \gamma_r(\boldsymbol{\theta}_r) d\boldsymbol{\theta}_r}{\int \gamma_{r-1}(\boldsymbol{\theta}_{r-1}) d\boldsymbol{\theta}_{r-1}}.$$

Del Moral et al. (2006) [6] show that this ratio can be estimated by

$$\widehat{\frac{Z_r}{Z_{r-1}}} = \sum_{k=1}^{K} W_{k,r-1} \tilde{w}_{r,k}.$$

When resampling is performed at each step we can estimate Z_R by the product of the average unnormalized weight at each iteration: $\prod_{r=1}^{R} \frac{1}{K} \sum_{k=1}^{K} \tilde{w}_{k,r}$. However, in general resampling is not performed at every step; in this case we can estimate the marginal likelihood as

$$\widehat{Z_R} = \prod_{r=1}^{R} \sum_{k=1}^{K} W_{r-1}^{(k)} \tilde{w}_r^{(k)}.$$

It has been shown that this estimator is unbiased, $\mathbb{E}[\widehat{Z_R}] = Z$ [5]. Note that this result is for the case where the sequence of annealing parameters is fixed. The other important feature of this estimator is that it is estimated during the sampling process. This is a notable computational advantage over standard Monte Carlo methods which require additional sampling procedures to estimate the marginal likelihood.

4.3 Bridge Sampling for estimating marginal likelihood

When traditional Monte Carlo procedures are used, such as MCMC, an additional sampling procedure must be implemented in order to estimate the marginal likelihood. In this work we only consider bridge sampling, which is one such estimation procedure that produces a lower-variance estimate than the naive Monte Carlo estimate. The bridge sampling method [20], uses samples from the posterior distribution as well as samples from a proposal distribution.

The estimator is given by

$$p(\mathbf{y}) = \frac{\mathbb{E}_{g(\boldsymbol{\theta})}[h(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})]}{\mathbb{E}_{p(\boldsymbol{\theta}|\mathbf{y})}[h(\boldsymbol{\theta})g(\boldsymbol{\theta})]} \approx \frac{\frac{1}{n_2}\sum_{j=1}^{n_2} h(\tilde{\boldsymbol{\theta}}_j)p(\mathbf{y}|\tilde{\boldsymbol{\theta}}_j)p(\tilde{\boldsymbol{\theta}}_j)}{\frac{1}{n_1}\sum_{i=1}^{n_1} h(\boldsymbol{\theta}_i^*)g(\boldsymbol{\theta}_i^*)},$$

where $h(\boldsymbol{\theta})$ is the *bridge function*, $g(\boldsymbol{\theta})$ is a proposal distribution, $\{\boldsymbol{\theta}_i^*\}$ are n_1 samples from the posterior distribution, and $\{\tilde{\boldsymbol{\theta}}_j\}$ are n_2 samples from the proposal distribution.

The implementation of bridge sampling that we use is the `bridgesampling` package in R [14]. This package uses the optimal bridge function h that minimizes the relative mean squared error of the estimator. That choice of h requires an iterative procedure to obtain the estimate $\hat{p}(\mathbf{y})$. Starting from an initial guess $\hat{p}(\mathbf{y})^{(0)}$, the $t+1$-th estimate is given by

$$\hat{p}(\mathbf{y})^{(t+1)} = \frac{\frac{1}{n_2}\sum_{j=1}^{n_2} \frac{\ell_{2,j}}{s_1\ell_{2,j}+s_2\hat{p}(\mathbf{y})^{(t)}}}{\frac{1}{n_1}\sum_{i=1}^{n_1} \frac{1}{s_1\ell_{1,i}+s_2\hat{p}(\mathbf{y})^{(t)}}}$$

where $\ell_{1,i} = \frac{p(\mathbf{y}|\boldsymbol{\theta}_i^*)p(\boldsymbol{\theta}_i^*)}{g(\boldsymbol{\theta}_i^*)}$ and $\ell_{2,j} = \frac{p(\mathbf{y}|\tilde{\boldsymbol{\theta}}_j)p(\tilde{\boldsymbol{\theta}}_j)}{g(\tilde{\boldsymbol{\theta}}_j)}$. For a proposal distribution we use the default proposal provided by the `bridgesampling` package. This is a multivariate normal proposal distribution where the mean vector and covariance matrix are set to match those of the posterior sample. Any parameters whose parameter space is bounded must be transformed to the real line; this is done within the package.

Chapter 5

Simulation Study

In this section we carry out a simulation study to determine the accuracy of both the model selection and the parameter estimates provided by ASMC. We will compare these results to MCMC. The data will be generated in order to reflect COVID-19 epidemic data.

5.1 Data simulation procedure

In order to perform a simulation that well emulates the proposed method's performance on real data we simulated data from the estimated transmission model in [1]. In this article Anderson et al. assume the full physical distancing transmission model. To reflect control measures in BC the authors define a piece-wise constant physical distancing forcing parameter, $f(t)$. We replace the observation delay likelihood in [1] with a negative binomial likelihood to reduce computation time. We fix two sampling fractions to reflect changes in testing procedures; the first, $p_1 = 0.1$ before $t = 41$ and the second $p_2 = 0.3$ for the remainder. These values were chosen to reflect values estimated from B.C. data [1]. A table of parameter values can be found in Table 6.1 and a more detailed discussion of the model in the context of BC COVID-19 data can be found in Section 6. We generate thirty datasets of sample size one hundred at equally space time points between $t = 0$ and $t = 130$.

5.2 Model selection

To evaluate the model selection performance of ASMC relative to MCMC, we fit three transmission models: the SEIR model, simple physical distancing model, and full physical distancing model. Using the ASMC marginal likelihood estimates and bridge sampling with MCMC, we compute the three pairwise Bayes factors between these three transmission model and obtain model comparison results for each method on all thirty simulated datasets. In the SEIR and simple physical distancing models we fix $\delta = 1/3$ and $\gamma = 1/5$, following [23]. Additionally for these models, we estimate $R_0 = \beta/\gamma$ instead of estimating β directly. This is sometimes done in epidemiology studies as R_0 is of often of greater interest than β. For the full physical distancing model we fix all parameters except for R_0, q, u_d, f_1, and f_2.

We assume a negative binomial observation model with sampling fractions set to the true values so that ψ is the only likelihood parameter that needs to be estimated. Since f_1 and f_2 lie in $(0, 1)$, a beta prior distribution was the natural choice. For the remaining transmission model parameters we assigned lognormal priors, since they all take values along the positive real line. Finally, we assigned a gamma prior distribution over the dispersion parameter ψ. The primary considerations when selecting the prior distributions were 1) the domain of the distribution, and 2) the ability to control the mean and variance; we did not take higher moments into consideration when selecting prior distributions or their hyperparameters. The mean of each prior was chosen to be similar to the true value to reflect an educated guess made by an experienced epidemiological researcher. The variances were chosen to be loosely informative because while it is possible for parameters to take on a wide range of values, only a small range represent a physically feasible scenario.

For each model on each of the thirty datasets, ASMC was run with $K = 1000$ particles, a target rCESS of $\phi = 0.95$, and a resampling threshold of $\epsilon = 0.5K$. MCMC was run so that both methods performed the same number of MCMC moves. This was done by first running ASMC, then running MCMC for $R \times K$ iterations. The result was $\sim 18,000$, $\sim 24,000$, and $\sim 55,000$ MCMC samples for the SEIR, simple, and full physical distancing models, respectively. The reason that the more complicated models require more moves is that they require a tighter sequence of intermediate distributions. We ran ASMC in parallel on 10 cores, so the computation time for ASMC was approximately 1/10 that of MCMC. For MCMC we use the last 1000 samples to perform bridge sampling so that ASMC and MCMC are using the same number of samples for the comparison.

In Table 5.1 we see the model selection performance of ASMC and MCMC. We summarise the performance by considering the different strengths of evidence, as shown in Table 4.1. For each method we compute the proportion of the thirty fits which produced each level of evidence in favour of the correct model. The bottom four rows show that both ASMC and MCMC can successfully identify that some form of physical distancing is required to model the data; the dynamics in the SEIR model are not complex enough to reflect changes in the data. However, when it comes to selecting one physical distancing model over another, ASMC provides at least strong evidence in favour of the correct model in more than 60% of the datasets, whereas MCMC will only show strong evidence in 30% of the datasets. The table also shows that in half of the datasets MCMC was unable to select the correct model at all, whereas ASMC only failed in 10% of the datasets.

5.3 Parameter estimation

In this component of the simulation study we explore how well ASMC recovers the model parameters and estimates their uncertainties. We compare ASMC to MCMC as in the previous section (5.2. In Figure 5.1 we estimate coverage probabilities of the $10, 20, \ldots, 90, 95, 99\%$

		Against	**Weak**	**Positive**	**Strong**	**Very Strong**
Full vs Simple	ASMC	0.1	0.1	0.13	0.27	0.4
	MCMC	0.5	0	0.2	0.07	0.23
Full vs SEIR	ASMC	0	0	0	0	1
	MCMC	0	0	0	0	1
Simple vs SEIR	ASMC	0	0	0	0	1
	MCMC	0	0	0	0	1

Table 5.1: Proportion of fits which yielded the corresponding amount of evidence in favour of the true transmission model.

credible intervals for each of the estimated parameters under both ASMC and MCMC. The estimates shown are the proportion of the thirty credible intervals which contained the true parameter values. Both ASMC and MCMC do provide reasonable coverage for ψ, however, for the transmission model parameters, the samples from MCMC produce very poor credible intervals. Looking at the coverage probabilities from ASMC, we can see that f_1, f_2, and R_0 provide intervals with reasonable coverage. For q and u_d the results are also reasonable, but the intervals tend to be too conservative for higher credible levels, especially for u_d.

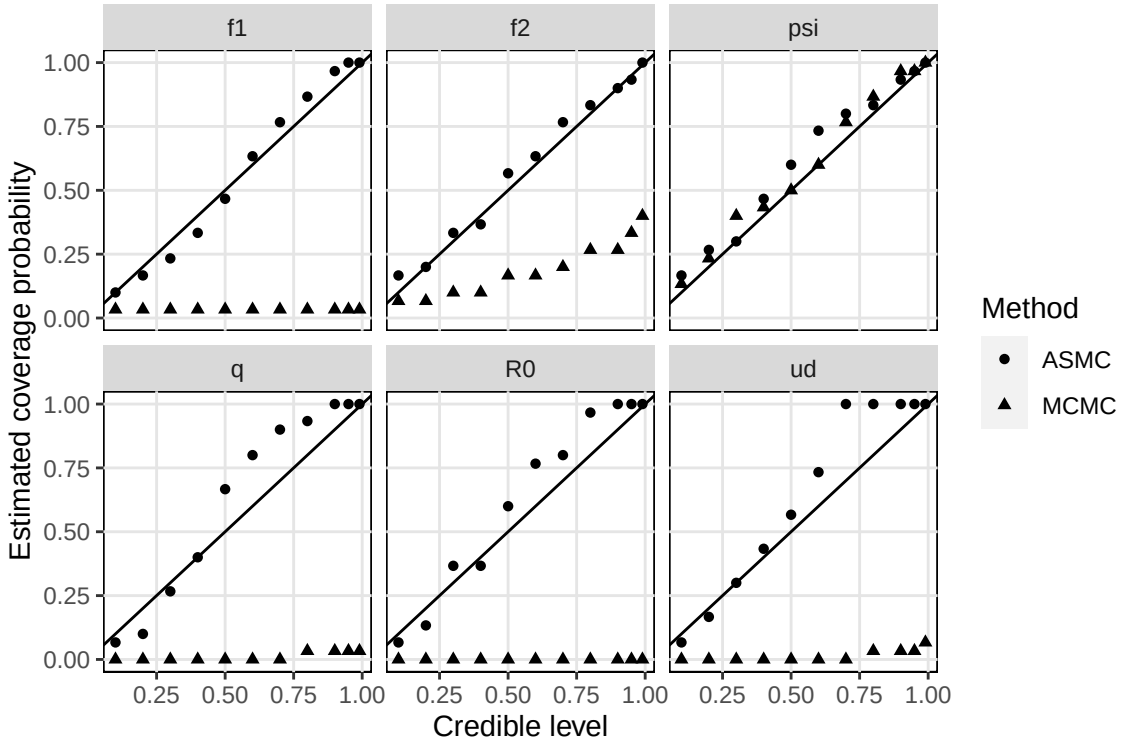

Figure 5.1: Estimated coverage probability of $10, 20, \ldots, 90, 95, 99\%$ credible intervals for each of the estimated parameters from both ASMC and MCMC.

In Figure 5.3 we focus on the ASMC results. We summarize the posterior estimates of each of the parameters from the first ten datasets by their posterior mean, shown by the

black points, and the 95% credible interval, as the horizontal lines. The vertical red lines indicate the true parameter values. For q and u_d we can see that we tend to over-estimate the parameters. This is not an unsurprising result; we only have direct information about the I and I_d compartments, so these posterior will be more strongly influenced by the prior distributions for these parameters. For the other parameter estimates the posterior means seem to provide reasonable estimates.

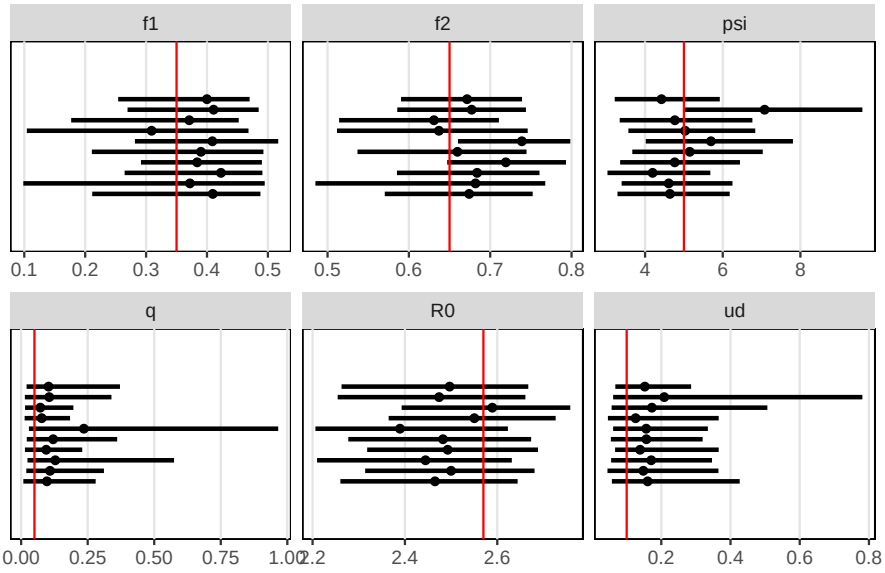

Figure 5.2: Summaries of posterior estimates for only the first ten simulated datasets, to avoid crowding the plot. Black points represent posterior means and horizontal lines are the 95% credible intervals. The vertical red line indicates the true value.

From the posterior sample of transmission model parameters we can construct K corresponding trajectories. In Figure 5.3 we demonstrate the posterior estimates of the number of active cases in the population. Recall that the number of active cases is $I + I_d$. We represent the true underlying number of active cases in red. The posterior mean of the trajectories is shown by the black lines and the shaded regions are the 95% credible band about the trajectory of active cases. We can see from the plot that the estimates tend to be quite reasonable, although we tend to overestimate. Overall the posterior distribution of the active cases is quite reasonable.

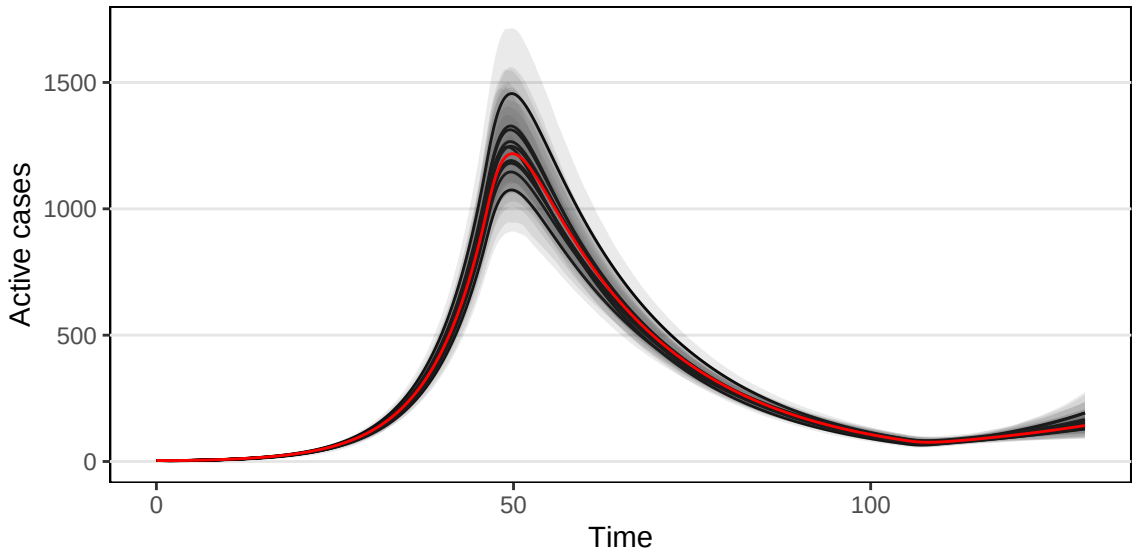

Figure 5.3: Estimated trajectory of number of active cases in the population. Posterior means over the first ten datasets are shown by the black lines; the shaded regions represent the 95% credible bands about these trajectories. The red trajectory is the true active cases used to generate the data.

Chapter 6

Data Analysis: British Columbia

For our data analysis we will use the confirmed case count data from the British Columbia Centre for Disease Control COVID-19 Dashboard [12]. These data consist of the date of each reported case along with the individual's sex, age group, and the health authority in which it was reported. For this data analysis we will only be using the total number of case counts on each day between February 5, 2020 and June 30, 2020. The data analysis will be comprised of two components: a comparison between three transmission models designed to reflect physical distancing measures and inference of model parameters for the selected model as well as an estimate of the infection trajectory.

6.1 Models considered

The three models we fit to the BC data were the simple physical distancing model, the full physical distancing model, and the contact tracing model. We did not consider the SEIR model for this analysis because it cannot account for effects of physical distancing.

For the simple physical distancing model we fix $\delta = 1/3$ and $\gamma = 1/5$, following the original authors [23]. We assume a piecewise constant structure for $f(t)$, shown in Equation 6.1,

$$f(t) = \begin{cases} 1, \text{ for } t < \tau_1 \\ f_1, \text{ for } \tau_1 \leq t < \tau_2 \\ f_2, \text{ for } t \geq \tau_2. \end{cases} \tag{6.1}$$

This form for the physical distancing parameter was used for BC cases by Stockdale et al. [25]. We choose τ_1 to correspond to March 18, 2020 when physical distancing measures were first implemented in BC Similarly, τ_2 is chosen to match a change point of May 17, 2020, when the physical distancing restrictions were first relaxed. We reparameterize the problem slightly to be expressed in terms of $R_0 = \beta/\gamma$. This leaves the parameters which are being estimated as R_0, f_1, and f_2.

Since the full physical distancing model was originally applied to BC data, we will set up the model in the same way as previous authors [1, 25]. As with the simple physical distancing model we assume a piece-wise constant forcing parameter with the same changepoints. We assume lognormal priors for q and u_d, centered at $\log(0.05)$ and $\log(0.1)$ respectively. For both the physical distancing models we assume a lognormal prior distribution on R_0 centered at $\log(2.57)$ and for the f_1 and f_2 we assume beta distributions centered at 0.35 and 0.65 respectively. These values are taken from the previous estimates which also assume a piece-wise constant forcing parameter [25].

In this analysis we do not estimate the initial number of cases. Instead we again follow the work of Anderson et al. who assume 8 cases in British Columbia at the beginning of February. To distribute these initial cases across the E_1, E_2, I, and their distanced counterparts, they assume that a fraction 0.83 of the population is practicing physical distancing. The authors also impose that 40% of the individuals are non-infectious (E_1), 10% are pre-symptomatic infectious (E_2), and 50% are symptomatic infectious. We use this approach in setting the our initial cases.

Parameter	**Definition**	**Value**
N	Population size	$5,100,000$
u_d	Rate of entering physical distancing state	0.1
u_r	Rate of leaving physical distancing state	0.02
$1/k_1$	Length of noninfectious exposure period (E_1 to E_2)	0.2 days
$1/k_2$	Length of pre-symptomatic infectious period (E_2 to I)	1 day
D	Mean infectious period duration	5 days
q	Quarantine rate	0.05
p_t	Fraction of cases on day t that are tested and reported	0.35 (pre-March 14th)
		0.68 (post-March 14th)
w_{shape}	Weibull parameter in delay-to-reporting distribution	1.73
w_{scale}	Weibull parameter in delay-to-reporting distribution	9.85

Table 6.1: Transmission model and observation model parameters for British Columbia estimated in [25].

For the contact tracing model we again followed the original authors [27] in terms of which model parameters were fixed and which were to be estimated. The model parameters and their values are given in Table 6.1. I've modified the original model to reflect the piecewise constant nature of the physical distancing that we've assumed in the previous models. Since this model does not have an explicit forcing parameter like the previous two, here I estimate three values of c, which correspond to the three segments of $f(t)$ above. We should expect to see $c_2 < c_3 < c_1$ to reflect the physical distancing schedule outlined above. The prior means of these transmission model parameters are taken as the estimated means found in the original paper, which was performed on data from Ontario, Canada. For c_2 and c_3, which were not in the original model I chose a prior designed to reflect the relative magnitudes of the physical distancing effects in the previous models. For β, q, ρ, and θ I

assigned a beta prior distribution and for δ_I, δ_q, γ_I, c_1, c_2, and c_3 I assigned lognormal prior distributions.

Parameter	Value	Fixed/Estimated
β	0.145	Estimated
q	0.1	Estimated
σ	1/5	Fixed
λ	1/14	Fixed
ρ	0.6	Estimated
δ_I	0.1	Estimated
δ_q	0.1	Estimated
γ_I	0.1	Estimated
γ_A	0.139	Fixed
γ_H	0.1	Fixed
α	0.008	Fixed
θ	0.05	Estimated
c_1	11	Estimated
c_2	5	Estimated
c_3	8	Estimated

Table 6.2: Parameters and their corresponding values for the B.C. data analysis. Fixed parameter values were taken from [27] and estimated parameter values listed are the prior means.

In order to keep the comparison between models a strict comparison of the transmission models I assume the observation delay model for all three. In the simple physical distancing model I define the active cases to be $I(t)$ and in the contact tracing model $H(t)$. The mean of the likelihood is constructed using the compartments which feed into these active compartments. I run ASMC with 1000 particles and $\phi = 0.95$ on each of the models.

6.2 Results

The three pairwise Bayes factors for the models we are considering in the problem are shown in Table 6.2. Referring to Table 4.1 we can see that there is weak evidence in favour of the simple physical distancing model against the full physical distancing model. Very strong evidence against the contact tracing model is also observed for both physical distancing models. From these results we can clearly see that the control measure dynamics defined by the contact tracing model are not suitable for British Columbia. The low Bayes factor between the two physical distancing models suggests that they both do a similar job of modelling the outbreak. This suggests that the despite the additional complexity of the full physical distancing model, it does not provide a better fit to the observed data than the simpler model.

In Figure 6.2, we show the posterior estimates of the active cases, as defined earlier. We can see that all three models are able to capture the major distancing events in mid-March

Models	Bayes Factor
Full vs. Simple	0.542
Full vs. Contact	10,100
Simple vs. Contact	18,500

Table 6.3: Bayes factors for the pairwise comparisons between each of the proposed transmission models.

and mid-May. The simple and full trajectories have very similar shape; the main difference is that the full physical distancing model estimates a lower number of active cases. This is to be expected, since this model will split some of the symptomatic individuals off into the quarantine groups, Q and Q_d, which are not reflected in the active case counts. In contrast the contact tracing model, which was heavily out of favour, has a slightly different shape in its trajectory. We can see a sharper peak at the first physical distancing changepoint. We can also see that after the second changepoint, the cases with the contact tracing model level off while the other two showed gradual increases.

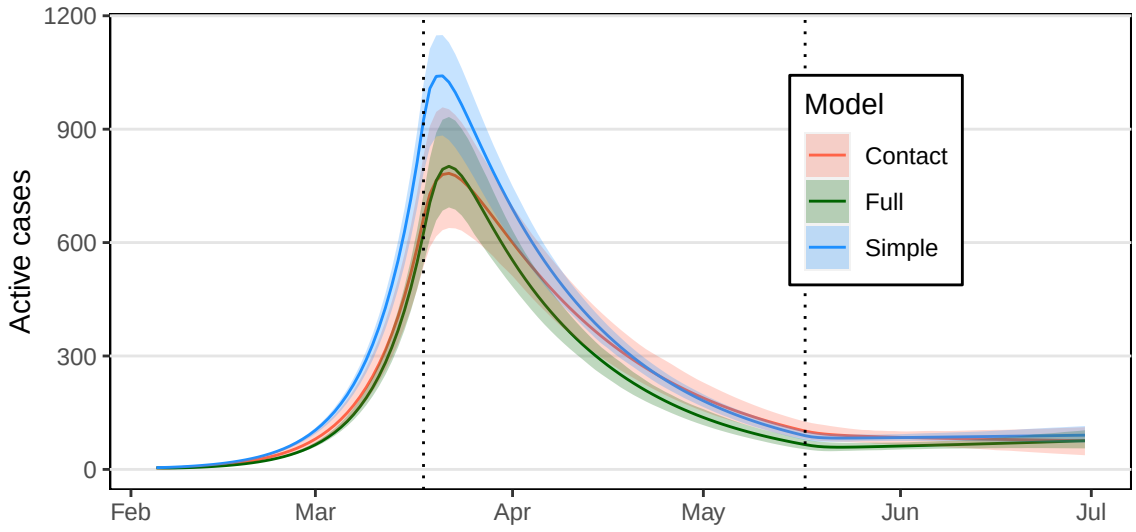

Figure 6.1: Estimated active case trajectories for the three transmission models. The vertical dashed lines correspond to the initial implementation of physical distancing measures (March 18) and relaxation of physical distancing measures (May 17).

The remainder of this section looks in greater detail at the simple physical distancing model since it had the greatest marginal likelihood. The posterior means and 95% credible intervals for the unknown parameters are given in Table 6.2. We compare these results to those found in [25], since they perform their analysis on the same range of data in B.C.. Both the R_0 and the forcing parameters are lower in our analysis than the previous study. In Figure 6.2 we show our posterior estimates of the total number of active cases in the population. The most prominent feature of the active cases is the spike in mid-late March.

This sudden change in the trend from rapidly increasing to rapidly decreasing is a result of physical distancing measures being put in place around this time. Though less prominent we can see in mid May that cases start to increase again; this corresponds to the relaxation of physical distancing regulations on May 18th. Harder to see is the lag between the active cases and the case counts, but it can be seen best at the peak in mid-late March. We can also note that even though there appears to be a second spike in case counts in late April, there is no second spike in the estimated trajectory; this is because the observation model is accounting for an increase in sampling at this point and so the peak is an artefact of the sampling procedure and not the transmission dynamics.

Parameter	**Mean**	**95% Credible Interval**
R_0	2.278	(2.2311,2.3048)
f_1	0.296	(0.2853,0.314)
f_2	0.448	(0.4218,0.4766)
ψ	6.346	(4.5537,8.3261)

Table 6.4: Posterior mean and 95% credible interval for unknown parameters of simple physical distancing model and observation delay.

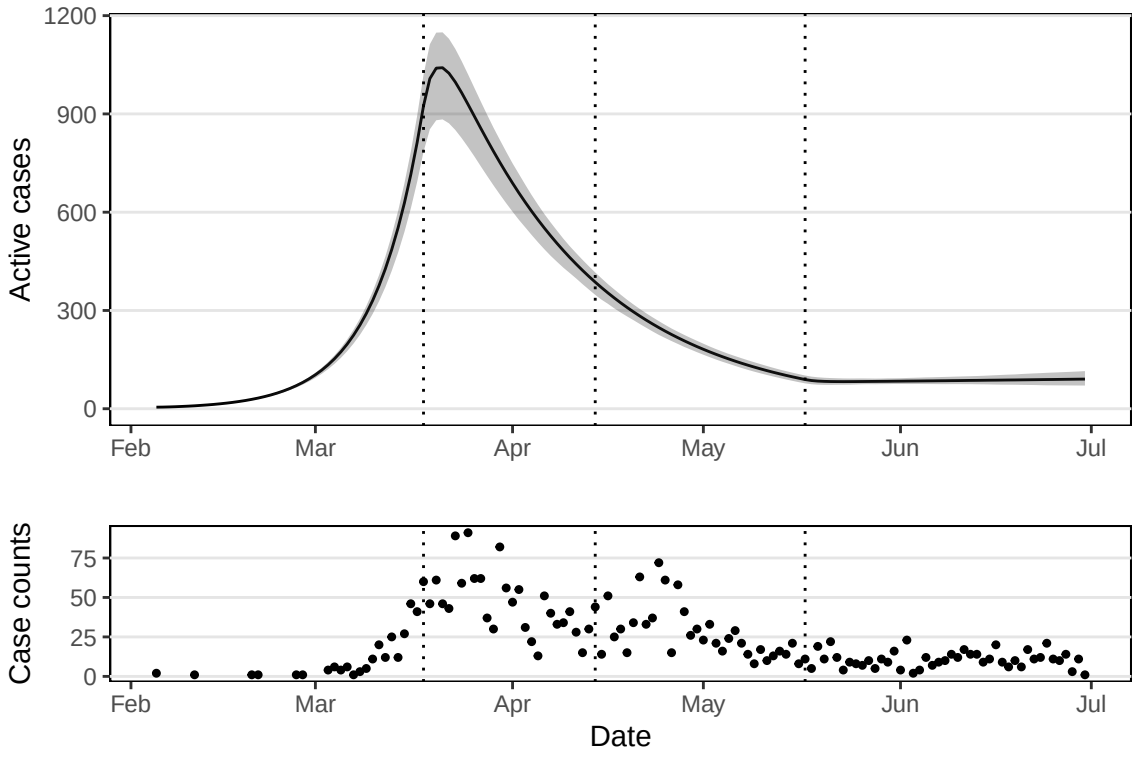

Figure 6.2: Summary of the posterior distribution of the active cases juxtaposed with the observed daily case counts. (Top) The black line is the posterior mean number of active cases (I) and the shaded region is the 95% credible band about the trajectory. (Bottom) Case counts in BC as reported by BCCDC. The vertical dashed lines, in order from left to right, correspond to implementation of physical distancing (March 18), change in testing procedure (April 14), and relaxation of physical distancing (May 17).

Chapter 7

Discussion

In this project we use an annealed SMC algorithm to provide a framework through which COVID-19 transmission models can be efficiently estimated and compared. We provide examples of how epidemiologists model the effects of physical distancing measures on the spread of COVID-19. A Bayesian model is proposed for inferring the parameters of the transmission models used to describe the effects of these distancing measures. An annealed SMC algorithm is used to sample from the posterior distribution of the model parameters. In this algorithm, a sequence of annealed intermediate distributions is constructed to produce a sample which well represents the posterior surface and is not prone to poor mixing as can happen with MCMC. The sequence of annealing parameters is adaptively determined to maximize the efficiency of importance weight estimates without imposing too great a computational cost. In addition to sampling from the posterior distribution, we estimate the marginal likelihood of the proposed transmission model. This estimate can be obtained directly from the importance weights and as a result we obtain the posterior samples and marginal likelihood estimate simultaneously. This offers major computational advantages over standard Monte Carlo algorithms. By running the ASMC algorithm on several transmission models, we can perform Bayesian model selection to determine which model(s) are best suited for explaining the observed daily case counts.

In our simulation study we found that ASMC was much more successful at selecting the true model than MCMC. This is likely due to the fact that many combinations of parameter values can provide similar solutions to the transmission model. This can create posterior surfaces with many local maxima which can result in poor mixing in MCMC; this is potentially the cause of MCMC's poor performance. This is corroborated by the illustration of the point estimates and coverage probability for MCMC, which were often quite dissimilar to the truth and varied between datasets. The reason that both methods experienced difficulty deciding between the full and simple physical distancing models, is that we are only making observations on the compartments corresponding to the active cases. So, when parameters controlling other compartments, such as the quarantine compartment, are varied it can have very little impact on the estimated trajectory of the active compartments. Therefore, the

fewer states in the system we actually observe, the more noise will affect our ability to select the correct model and provide reasonable parameter estimates.

A demonstration of ASMC was performed on confirmed daily case counts from British Columbia. We compared three models: a simple physical distancing model, a more complex physical distancing model which has been used to model the infection in BC, and a contact tracing model which was used in Ontario. We found that the observed data was more likely under the first two models than the contact tracing model. There are two potential reasons for this. The first is that the physical distancing dynamic in the contact tracing model was not very well-suited for the BC data. By comparing the estimated trajectories in Figure 6.2 we can see that the contact tracing model does not capture an increase in active cases after the physical distancing measures were relaxed, merely a leveling out of cases, whereas the other two physical distancing methods do show an increase in active cases. The second potential reason for the preference against the contact tracing model could be a result of the higher dimension of the posterior distribution, relative to the other two. Of the three models, the contact tracing model estimated the most parameters. The Bayes factor implicitly penalizes models of higher dimension, so the data would need to be more likely under the contact tracing model to overcome this. We also found that there was no practical difference between the two physical distancing models in terms of how well they explained the data. This is most likely due to the fact that we are only making observations on the active compartments, so the additional flexibility in modeling pre-symptomatic infectiousness and quarantining has very little impact on the model's ability to explain the observed data.

7.1 Future work

One area for future research is to improve the Markov kernel used to move particles between intermediate distributions. There is a huge amount of literature on Monte Carlo methods which we can draw upon to improve this key step of the ASMC algorithm. An obvious candidate would be to use a Hamiltonian Monte Carlo (HMC) sampler instead; this is commonly used in Bayesian inference and is efficient when dealing with correlated variables. One drawback to using rejection-based samplers such as Metropolis-Hastings and HMC is that it exacerbates the sample degeneracy issue found in SMC samplers. A well-known problem of SMC methods is that the resampling reduces the number of unique particles in the system; this issue is made worse when not all of the particle values are changed during the proposal step of the algorithm due to rejection in the Markov kernel. One possibility is to run a longer MCMC chain when propagating particles; this would increase the probability that the final proposed particle is different from the original, but the computation time would scale linearly with the length of the chain, so this is unappealing. A more interesting

avenue could be exploring the use of rejection-free samplers, such as the bouncy particle sampler [3] or Zig-Zag sampler [2], to move particles between intermediate distributions.

Another avenue for future work would be to provide a more thorough treatment of ASMC for model comparison. This would involve exploring the effect of K and ϕ, the tuning parameters of the ASMC algorithm, on the accuracy of model selection. In this project we use seemingly reasonable values, however it is lacking in any sort of optimization of the tuning parameters in terms of finding a trade-off between accuracy and computation time. Additionally, while we do explore to some extent the effect of variability in the data on our results, we do not explore the effect of variability within the algorithm itself.

The final future direction suggested is to consider the incorporation of basis expansions into this model comparison framework. Basis expansions have been used [15, 24] to estimate parameters of dynamic systems and have demonstrated improvements, especially in the presence of noisy data or incomplete observation of the system. However, the Bayesian model in [15] uses an improper prior distribution to regulate the basis expansions; this would preclude any meaningful interpretation of the marginal likelihood. It could be very beneficial to research possible models which can take advantage of the benefits of basis expansions without sacrificing the capacity for model comparison.

Bibliography

[1] Sean C Anderson, Andrew M Edwards, Madi Yerlanov, Nicola Mulberry, Jessica E Stockdale, Sarafa A Iyaniwura, Rebeca C Falcao, Michael C Otterstatter, Michael A Irvine, Naveed Z Janjua, Daniel Coombs, and Caroline Colijn. Estimating the impact of COVID-19 control measures using a bayesian model of physical distancing. *medRxiv*, 2020.

[2] Joris Bierkens, Paul Fearnhead, and Gareth Roberts. The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. *Annals of Statistics*, 47(3):1288–1320, 2019.

[3] Alexandre Bouchard-Côté, Sebastian J Vollmer, and Arnaud Doucet. The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *Journal of the American Statistical Association*, 113, 2018.

[4] Nicolas Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *The Annals of Statistics*, 32(6):2385–2411, 2004.

[5] Pierre Del Moral. Feynman-kac formulae. In *Feynman-Kac Formulae*, pages 47–93. Springer, 2004.

[6] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[7] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*, 22(5):1009–1020, 2012.

[8] Randal Douc and Olivier Cappé. Comparison of resampling schemes for particle filtering. In *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, pages 64–69. IEEE, 2005.

[9] Arnaud Doucet, Mark Briers, and Stéphane Sénécal. Efficient block sampling strategies for sequential monte carlo. *Journal of Computational and Graphical Statistics*, 15(3):693–711, 2006.

[10] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

[11] Arnaud Doucet and Adam Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, 12, 2009.

[12] British Columbia Centre for Disease Control. BC COVID-19 dashboard. Available at `http://www.bccdc.ca/health-info/diseases-conditions/covid-19/data` (2020/11/14).

[13] Walter R. Gilks and Carlo Berzuini. Following a moving target - monte carlo inference for dynamic bayesian models. *Journal of the Royal Statistical Society B*, 63:127–146, 2001.

[14] Quentin F Gronau, Henrik Singmann, and Eric-Jan Wagenmakers. bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, 92(10):1–29, 2020.

[15] Hanwen Huang, Andreas Handel, and Xiao Song. A Bayesian approach to estimate parameters of ordinary differential equation. *Computational Statistics*, 35:1481–1499, 2020.

[16] H Jeffreys. Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, 31:203–222, 1935.

[17] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[18] William O Kermack and A G McKendrick. Contributions to the mathematical theory of epidemics, part ii. *Proceedings of the Royal Society of London*, 1(38):55–83, 1932.

[19] Jun S Liu and Rong Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, 1998.

[20] Xiao-Li Meng and Wing Hung Wong. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 6(4):831–860, 1996.

[21] Mark Newman. *Computational Physics*. Createspace, revised and expanded edition, 2013.

[22] World Health Organization. WHO coronavirus disease (COVID-19) dashboard. Available at `https://covid19.who.int/` (2020/11/01).

[23] Seema Patrikar, Deepti Poojary, D.R. Basannar, D.S. Faujdar, and Kunte Renuka. Projections for novel coronavirus (COVID-19) and evaluation of epidemic response strategies for India. *Medical Journal Armed Forces India*, 76, 2020.

[24] Jim O Ramsay, Giles Hooker, David Campbell, and Jiguo Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):741–796, 2007.

[25] Jessica E Stockdale, Renny Doig, Joosung Min, Nicola Mulberry, Liangliang Wang, Lloyd T Elliott, and Caroline Colijn. Long time frames to detect the impact of changing COVID-19 control measures. *Submitted to: Eurosurveillance*, 2020.

[26] Liangliang Wang, Shijia Wang, and Alexandre Bouchard-Côté. An annealed sequential Monte Carlo method for Bayesian phylogenetics. *Systematic Biology*, 2019.

[27] Jianhong Wu, Biao Tang, Nicola L Bragazzi, Kyeongah Nah, and Zachary McCarthy. Quantifying the role of social distancing, personal protection and case detection in mitigating COVID-19 outbreak in Ontario, Canada. *Journal of Mathematics in Industry*, 10(15), 2020.

[28] Yan Zhou, Adam M Johansen, and John AD Aston. Toward automatic model comparison: an adaptive sequential Monte Carlo approach. *Journal of Computational and Graphical Statistics*, 25(3):701–726, 2016.