Bayesian Nonparametric Methods for Individual-Level Stochastic Epidemic Models

Rowland Guy Seymour

A Thesis presented for the degree of

Doctor of Philosophy

School of Mathematical Sciences

University of Nottingham

2^{nd} December 2019

Abstract

Simulating from and making inference for stochastic epidemic models are key strategies for understanding and controlling the spread of infectious diseases. Current methods for modelling infection rate functions are exclusively parametric. This often involves making strict assumptions about the way the disease spreads and choices which may lack any biological or epidemiological justification. To remove the need for making such assumptions, we develop a Bayesian nonparametric framework which allows us to learn how the disease spreads directly from the data.

In this thesis, we consider individual-level models where the infection rate between each pair of individuals depends on characteristics of their relationship. We begin by considering infectious diseases where the infection rate between any two individuals can be modelled by a function of a single characteristic, for example, the distance between them. We model this function nonparametrically by assigning a Gaussian Process prior distribution to it and then develop an efficient data augmentation Markov Chain Monte Carlo algorithm to infer this function, alongside the prior distribution hyperparameters and the times individuals were infected.

We develop this methodology further, first for multi-type outbreaks and then for outbreaks where the infection rate function depends on more than one characteristic. For multi-type outbreaks, where the infection rate between two individuals not

only depends on the characteristics, but also the type of individual being infected, we develop a Multi-Output Gaussian Process method. This method allows us to compare how susceptible each type of individual is to infection. We extend our Gaussian Process method into several dimensions for modelling outbreaks where the infection rate between individuals can be modelled as a function of multiple continuous variables.

Finally, we demonstrate our results on two data sets, giving new insights and analysis. The first is an outbreak of Avian Influenza in the Netherlands in 2003, where over 30 million birds were culled. Using the posterior predictive distribution of our nonparametric model, we simulate outbreaks of Avian Influenza to assess various control measures. Alongside our nonparametric analysis, we are able to investigate which of the pre-emptively culled farms were infected. The second is an outbreak of Foot and Mouth Disease in Cumbria, UK. We are able to analyse the relationship between the infection rate of farms with different kind of livestock, showing that farms with both cattle and sheep were much more susceptible to the virus than farms with a single type of livestock.

Acknowledgements

Firstly, I would like to thank my supervisors, Theo Kypraios and Phil O'Neill, without whom none of this would have been possible. Their guidance and advice has been greatly appreciated. Secondly, I would like to thank my parents who have supported me throughout my whole education. Thirdly, I would like to thank George, who has been a great friend to me during our PhDs.

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/N50970X/1. I am grateful to Thomas Hagenaars and colleagues (Wageningen Bioveterinary Research), The Netherlands Food and Consumer Product Authority and the Dutch Ministry of Agriculture, Nature and Food Quality for sharing anonymised outbreak, culling and denominator data of the Dutch 2003 HPAI epidemic with me. I am grateful for access to the University of Nottingham High Performance Computing Facility.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Epidemic Modelling

Mathematical models for outbreaks of infectious diseases can help us understand how a disease spreads between individuals. Mathematical models have been used to study the spread of a wide range of infectious diseases from the Ebola virus (Lekone and Finkenstädt, 2006) and HIV (Jacquez et al., 1988) to Avian Influenza (Retkute et al., 2018) and Swine Fever (Stegeman et al., 1999). Fitting epidemic models to data from outbreaks of infectious diseases allows us to understand what occurred in the outbreak and which factors played a substantial role in the spread of the disease. For example, in an outbreak of measles, we may be able to identify which areas or age groups were most susceptible to the disease. Simulating future outbreaks from epidemic models allows us to develop control strategies and prevent future outbreaks of the disease. For example, after having identified a particular age group that is susceptible to being infected with the measles virus, we can simulate outbreaks where children in the identified age group have been vaccinated.

Understanding how diseases spread has long been of interest to mathematicians.

Bernoulli was the first to model a disease mathematically in 1760 when he investigated the spread of smallpox and furthered the argument for vaccination. Another key contribution to the area came from Kermack and McKendrick (1927) where the authors developed the idea of compartmental models. This is where individuals are either susceptible to infection, infected with the disease, or removed from the outbreak. Removed individuals are those who have had the disease and are no longer infectious. This may be because they have recovered from the disease and have developed immunity, they have been quarantined, or have died.

In this thesis, we are concerned with stochastic models for outbreaks of infectious diseases. This is in contrast to deterministic models, which are specified through a system of differential equations (Bacaer, 2011). Stochastic models allow us to capture the randomness in how humans and animals behave, which is a key part to modelling disease transmission. Stochastic models can be more difficult to formulate and analyse than their deterministic counterparts. There are several significant texts in the area of stochastic epidemic models including Bailey (1975), Becker (1989) and Andersson and Britton (2000).

1.1.1 Motivation

Inference for infectious disease models is often complicated by the amount of data observed in an outbreak. Although we observe when an individual suffers from symptoms or is removed from the general population, we do not observe when they are infected. This makes understanding how or when one individual infected another difficult. When modelling an infectious disease, we need to quantify the probability one individual infects another, and with limited data this is challenging. We may also want to quantify various factors. For example, specifying that children are more susceptible to the measles virus than adults.

This thesis deals with individual-level models, where we specify the infection rate from one individual to another. This allows us to include information about the relationship between the individuals; for example, the distance between them, their ages and sex, or their vaccination status. Current methods for modelling the

spread of a disease in an individual-level model typically use parametric forms; that is they propose the exact parametric form for the infection rate from one individual to another. The choice of parametric form for the kernel specifying the infection rate between each pair fo individuals is often arbitrary and lacks biological justification. It is may be difficult to include all biological effects in the parametric form. For example, there is evidence that cattle are more susceptible to Foot and Mouth disease than sheep or pigs when the virus is transmitted by wind (Alexandersen et al., 2003). Including this effect in a model, without it being too narrow, is difficult. Another challenging problem is where there is a spatial element to the spread of a disease. Consider an outbreak of Foot and Mouth Disease on cattle farms, the probability one farm infects another may depend on the distance between the farms. Given the observed data, quantifying this spatial element is difficult and may require a complex parametric form. Parametric forms are strict assumptions which are difficult to justify.

To avoid making such strict assumptions, in this thesis we will develop a Bayesian nonparametric approach. This method will allow us to learn how a disease spreads directly from the data. We will develop methods for including both continuous and categorical variables, allowing us to model the spread of a wide variety of diseases.

1.2 Bayesian Inference and Computation

Throughout this thesis, we will develop methods for performing Bayesian inference. We now outline the main theory and methods for Bayesian inference, however for a detailed introduction see Bernardo and Smith (1994) or Gelman et al. (2013).

1.2.1 Bayes' Rule

Bayesian inference is conducted through the posterior distribution for the parameter θ given the observed data y. We construct this using Bayes' rule, which is given by:

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)}, \tag{1.1}$$

where $\pi(y|\theta)$ is a likelihood function of the observed data and $\pi(\theta)$ is the prior distribution for the model parameter θ. The prior distribution quantifies our beliefs about the parameter θ before observing any data and the posterior distribution defines the probability distribution of θ given the observed data y.

One difficulty when carrying out Bayesian inference is computing the normalising constant $\pi(y)$ as we need to integrate – or compute the sum – over all possible values of θ. Thus, the normalising constant is given by:

$$\pi(y) = \int \pi(y|\theta)\pi(\theta)d\theta.$$

However, as this does not depend on the parameter of interest, we usually omit this and consider the unnormalised posterior distribution

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta).$$

Another difficulty in Bayesian inference is the choice of prior distribution, and this has received much attention (Bernardo and Smith, 1994). In this thesis, we will use two methods for selecting prior distributions. The first is to choose a conjugate prior distribution, that is a distribution that when combined with the likelihood function leads to a posterior distribution that is of the same distribution family. These priors are often chosen for their analytical and computational convenience. The second method is to choose a prior distribution that contains as little information as possible about the parameter of interest. These are often known as non-informative prior distributions and are used when we have little prior information about the parameter.

1.2.2 Markov Chain Monte Carlo Methods

In Bayesian inference, we wish to learn about the posterior distribution and one way to do this is to draw samples. Often the posterior distribution does not have a form from which we can draw samples easily or directly. Markov Chain Monte Carlo methods allow us to generate samples from any probability distribution, even when it is not tractable.

With MCMC methods we generate a sequence of samples that is a Markov chain and the stationary distribution of the Markov chain is the posterior distribution

$\pi(\theta|y)$. We now outline two MCMC methods which are relevant to this thesis: the Metropolis Hastings algorithm and the Gibbs sampler. For a more detailed discussion see Gelman et al. (2013, §10 − §13).

1.2.2.1 Metropolis-Hastings Algorithms

Suppose we wish to draw samples of θ from the posterior distribution $\pi(\theta|y)$. The Metropolis-Hastings algorithm begins by us choosing some initial value for the Markov chain $\theta^{(0)}$. On the i^{th} iteration of the algorithm, we propose a new value of the Markov chain θ' from a proposal distribution $q(\cdot|\theta^{(i-1)})$. We then accept this proposal with probability

$$p_{acc} = \frac{\pi(\theta'|y)}{\pi(\theta|y)} \frac{q(\theta^{(i-1)}|\theta')}{q(\theta'|\theta^{(i-1)})} \wedge 1,$$

where $a \wedge b = \min\{a, b\}$. If we accept the proposal, then we set $\theta^{(i)} = \theta'$, otherwise $\theta^{(i)} = \theta^{(i-1)}$. Rewriting the posterior distribution as that shown in equation (1.1), we see that the ratio of posterior distributions in the acceptance probability allows us to avoid calculating the normalising constant in the posterior distribution:

$$\begin{aligned} p_{acc} &= \frac{\pi(\theta'|y)}{\pi(\theta|y)} \frac{q(\theta^{(i-1)}|\theta')}{q(\theta'|\theta^{(i-1)})} \wedge 1 \\ &= \frac{\pi(y|\theta')}{\pi(y|\theta)} \frac{\pi(\theta')}{\pi(\theta)} \frac{\pi(y)}{\pi(y)} \frac{q(\theta^{(i-1)}|\theta')}{q(\theta'|\theta^{(i-1)})} \wedge 1 \\ &= \frac{\pi(y|\theta')}{\pi(y|\theta)} \frac{\pi(\theta')}{\pi(\theta)} \frac{q(\theta^{(i-1)}|\theta')}{q(\theta'|\theta^{(i-1)})} \wedge 1. \end{aligned}$$

The choice of proposal distribution can considerably effect the convergence rate and mixing of the resulting Markov chain. If the proposal distribution given the current value of the Markov chain is too broad in the space of plausible parameter values, then proposed values will be frequently rejected. Conversely, if the proposal distribution is too narrow, the Markov chain will be slow to converge. Two common proposal distributions are the normal and uniform distributions, which result in a random walk Metropolis-Hasting algorithm. For the normal proposal distribution, we propose values by:

$$\theta' \sim N(\theta^{(i-1)}, \sigma^2),$$

where σ^2 controls the distance between values in the parameter space. Similarly for the uniform proposal distribution, we propose values such that

$$\theta' \sim U[\theta^{(i-1)} - a, \theta^{(i-1)} + a],$$

where a controls the step size.

The Metropolis-Hastings algorithm is given in algorithm 1.

Algorithm 1 The Metropolis Hasting Algorithm

Initialise the chain with value $\theta^{(0)}$
for $i \leftarrow 1$ to m **do**
    Propose $\theta' \sim q(\cdot|\theta^{(i-1)})$
    Compute $p_{acc} = \frac{\pi(\theta'|y)}{\pi(\theta|y)} \frac{q(\theta^{(i-1)}|\theta')}{q(\theta'|\theta^{(i-1)})}$
    Sample $u \sim U[0, 1]$
    if $p_{acc} < u$ **then**
        $\theta^{(i)} = \theta'$
    else
        $\theta^{(i)} = \theta^{(i-1)}$
    end if
end for

1.2.2.2 Gibbs Samplers

The Gibbs sampler is a special case of the Metropolis-Hastings algorithm, where every proposal is accepted. Suppose $\theta = \{\theta_1, \ldots, \theta_n\}$ is a vector consisting of n parameters and let $\theta_{-j} = \{\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots \theta_n\}$ be the vector of parameters excluding the j^{th} parameter. The Gibbs sampler can be implemented if the conditional distribution of θ_j given the remaining parameters is known and has a closed form. To implement the Gibbs sampler, we sample a value of θ_j from the distribution $\pi(\theta_j|\theta_{-j}, y)$ for $j = 1, \ldots, n$. A full algorithm is given in algorithm 2.

Algorithm 2 The Gibbs Sampler

1: Initialise the chain with value $\theta^{(0)} = \{\theta_1^{(0)}, \ldots, \theta_n^{(0)}\}$
2: **for** $i \leftarrow 1$ to m **do**
3: Sample $\theta_1^{(i)} \sim \pi(\theta_1 | \theta_2^{(i-1)}, \ldots, \theta_n^{(i-1)}, y)$
4: Sample $\theta_2^{(i)} \sim \pi(\theta_2 | \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_n^{(i-1)}, y)$
5: Sample $\theta_3^{(i)} \sim \pi(\theta_3 | \theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_n^{(i-1)}, y)$
6: $\vdots$
7: Sample $\theta_n(i) \sim \pi(\theta_n | \theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_{n-1}^{(i)}, y)$
8: **end for**

1.2.2.3 Burn-in and Thinning

After running the MCMC algorithm for a specified number of iterations, we have a chain of samples which should resemble samples from the posterior distribution. Often, the early samples in the sequence are influenced by the choice of the initial value $\theta^{(0)}$. We reduce the impact of this choice by removing these early samples. We refer to these early samples as a burn-in period and the length of this period is problem specific and often chosen by visually checking the resulting Markov chain. To do this, we use trace plots, to inspect how well the Markov chain is mixing.

Once we have removed the early samples, the chain of samples the should resemble a sequence of samples drawn from the posterior distribution. However, we can choose to store every k^{th} iteration as this not only reduces the required storage space, but also reduces autocorrelation between the samples. This is known as thinning.

1.2.3 Bayesian Nonparametric Methods

In Bayesian inference for parametric models, we treat the parameters as random variables and use a prior distribution to describe the knowledge or uncertainty about the parameter. A Bayesian nonparametric model places a prior distribution on an infinite parameter space and invokes a finite set of parameters (Orbanz and Teh, 2017). We usually consider the infinite parameter space to be the set of all possible solutions to the problem. An example of this would be the set of all continuous,

differentiable functions when carrying out a regression problem.

There are a wide variety of techniques for carrying out Bayesian nonparametric methods. When modelling densities, a typical Bayesian nonparametric method employed is a Dirichlet Process. (Gelman et al., 2013). For regression problems, one widely used method is Gaussian Processes (GPs), which allow us to model functions without specifying an exact parametric form (Rasmussen and Williams, 2006) – this is the method we will be using in this thesis.

Bayesian nonparametric models grow in dimension as the number of data points increase, and as such can describe a greater range of variation in the data. They also reduce the number of assumptions we need to make about the underlying generating process compared to parametric models. This is because the Bayesian nonparametric framework is a lot more flexible than a parametric framework. Using Bayesian nonparametric methods allow us to design models which can better describe the data and reduce the assumptions modellers need to make.

However, this ability to capture a large range of effects and flexibility results in challenges we do not face when implementing parametric models. Having a model with a large dimension greatly increases the difficulty in fitting Bayesian nonparametric models compared to parametric models. This difficulty comes in several forms, including technical ability to fit models, computational complexity in performing inference for the models, and interpreting the models. Having a more flexible framework also makes it more challenging to compare models. When comparing parametric models, we often penalise each model based on how many parameters it has, however as Bayesian nonparametric models have an infinite number of parameters, this is not possible.

Parametric and Bayesian nonparametric models often see large differences in the uncertainty in the results. The uncertainty in results for parametric models is often considerably smaller than that for Bayesian nonparametric models. Naïvely, this makes parametric methods more attractive. However, Bayesian nonparametric models allow us to better quantify the uncertainty. As parametric models require very strict assumptions, the best-fitting such model may not not accurately represent

the underlying process and will not estimate the correct level of uncertainty in the results. This may result in incorrect predictions being made from the model. Bayesian nonparametric methods are more flexible and result in larger uncertainty around estimates, but the models are better posed.

1.3 Stochastic Epidemic Modelling

We now introduce stochastic epidemic models, which will be used in this thesis for modelling the spread of infectious diseases.

1.3.1 Compartmental Models

The building blocks of stochastic epidemic models are the compartments. Suppose there is an outbreak of an infectious disease in a population. We put each individual in the population in a compartment depending on their current status in the outbreak (Andersson and Britton, 2000). The most common compartments are the Susceptible, Infected and Removed compartments. A model with these three compartments is referred to as an SIR model; at any point in the outbreak an individual may be susceptible, infected or removed. An individual is susceptible if they have not been infected with the disease and are susceptible to contracting it. Infected individuals are individuals who are infected with the disease and we assume that all infected individuals are also infectious and able to pass the disease on to susceptible individuals. Removed individuals are those who have had the disease and are no longer able to pass the disease on. This may be because they no longer have the disease, have been put into quarantine, or otherwise have been removed from the population. Once an individual enters the removed class, they cannot be reinfected. Typically, an individual will start in the susceptible compartment and, if infected, then move to the infected compartment, before entering the removed compartment when they are no longer infected. An example of this model is given in figure 1.1.

Compartments allow us to build flexible models as we are not limited by the type or number of compartments. Another possible model is an SIS model, where instead

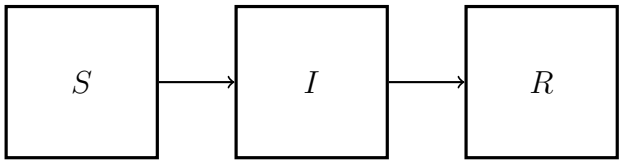


Figure 1.1: A graphical representation of an SIR stochastic epidemic model.

of being removed after being infected, individuals return to being susceptible. This is used for modelling diseases such as the common cold. For some diseases an individual may not be infectious upon becoming infected and to model this we can introduce an exposed compartment between the susceptible and infectious compartments. This is often referred to as an SEIR model. In this thesis, we consider outbreaks among populations of a fixed size as we assume the outbreaks to occur over a short period. However it is possible to include demographic changes in compartmental models. Including demographic changes is particularly useful for modelling endemic diseases where we are considering the number of cases of a disease over a long time span or diseases such as HIV, where there is chance a pregnant mother with HIV will pass on the disease to her baby. Given these compartmental models, we now discuss stochastic models for epidemics.

1.3.2 The General and Standard Stochastic Epidemic Models

We begin by defining the so-called *general* Susceptible-Infective-Removed stochastic epidemic model (see e.g. Bailey (1975)). In this model, we consider a population consisting of N individuals and for any time $t \geq 0$, an individual may be *susceptible* to the disease, *infected* with the disease, or *removed* from the infection process. We assume that the population consists entirely of susceptible individuals and at time $t = 0$, one individual becomes infected. Once infected, an individual remains infected for a time period drawn from an Exponential distribution with rate parameter γ and density function:

$$h(x|\gamma) = \gamma \exp\{-\gamma x\}, \quad \gamma > 0,\, x > 0.$$

When this time has elapsed they enter the removed class. We assume the infectious time period for each individual is independent of that of other individuals. While an individual is infected, we assume it makes infectious contact with a given susceptible individual at the time points given by a homogenous Poisson process with rate β, and infectious contact results in the susceptible individual becoming infected immediately. We assume the infectious contact process for each pair of individuals is independent of any other pair. The outbreak is declared over once there are no infected individuals remaining in the population. At this time, individuals are either susceptible, having avoid infectious contact, or removed, as their infectious period has elapsed. During the outbreak, we assume no new individuals enter the population and no individuals leave the population. For short time periods, this is a reasonable assumption to make.

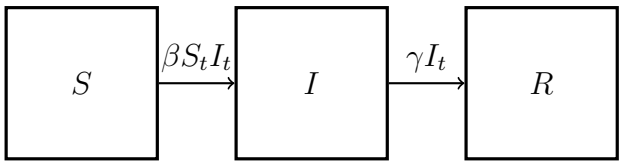


Figure 1.2: A graphical representation of the general stochastic epidemic model.

We define S_t, I_t and R_t to be the number of susceptible, infected and removed individuals at time t respectively. As the homogenous Poisson processes are independent of each other, the process of new infections at time t is an inhomogeneous Poisson process with rate $\beta S_t I_t$. We refer to β as the infection rate and $\beta S_t I_t$ as the infection force. Modelling the infectious period distribution by an exponential distribution makes the stochastic process $\{(S_t, I_t);\, t \geq 0\}$ a Markov process. In Andersson and Britton (2000), the authors also outlined the *standard* epidemic model. This model differs from the general model by allowing the length of the infectious period to be drawn from any arbitrary and non-negative distribution $\mathcal{D}$. This allows for a more realistic choice of infectious period distribution, but results in the stochastic process $\{S_t, I_t; t \geq 0\}$ no longer being Markovian whenever $\mathcal{D}$ is not exponential.

1.3.3 The Individual-level Stochastic Epidemic Model

Often we have information about the individuals in the population which can help us analyse the spread of a disease. Stockdale et al. (2018) describes the analysis of an outbreak of smallpox in a town in Nigeria, where the authors took into account in which compound each individual lived, whether they were a member of a particular church and if an individual had been vaccinated. Another example is presented by Jewell et al. (2009), which analyses an outbreak of Foot and Mouth Disease in the UK, where the authors considered the locations of the farms, as well as the type and number of animals on each farm.

We consider an individual-level stochastic epidemic model, which is a type of heterogeneously mixing model. In this model, we specify the infection rate between each pair of individuals, and we denote the pair-wise infection rate from individual i to individual j by $\beta_{i,j}$. The only difference between the standard stochastic epidemic model and the individual-level model is that we assume an infected individual i infects a susceptible individual j at time points from a Poisson process with rate $\beta_{i,j}$. As in the standard model, if j becomes infected they remain so for a time period drawn from an arbitrary and non-negative distribution $\mathcal{D}$. If we fix $\beta_{i,j} = \beta_0 > 0$ for all pairs of individuals, the model simplifies to the standard epidemic model.

Instead of specifying a distinct value of $\beta_{i,j}$ for each pair of individuals, we define a function that takes the characteristics of the individuals as inputs and outputs the pair-wise infection rate. This is almost always done by specifying an exact parametric form. For example in the aforementioned outbreak of Foot and Mouth Disease, one form for the infection rate from individual i to j is given by:

$$\beta_{i,j} = (\beta_1 (n_i^c)^{\beta_2} + (n_i^s)^{\beta_2}) \cdot (\beta_3 (n_j^c)^{\beta_2} + (n_j^s)^{\beta_2}) \cdot \frac{\beta_4}{d_{i,j}^2 + \beta_5},$$

where n_i^c and n_i^s and the number of cattle and sheep on farm i, $d_{i,j}$ is the Euclidean distance between farm i and j, and $\beta_1, \ldots, \beta_5$ are parameters controlling the infection rate (Jewell et al., 2009). Having proposed the infection rate function, we may perform inference for the function parameters.

However, we argue that proposing parametric forms for the infection rate function

requires us to make assumptions an in many cases there is no clear biological or epidemiological basis for using a particular parametric model. We instead develop a Bayesian nonparametric method, which allows for a much more flexible approach avoiding such assumptions. To understand the context around our new method, we now outline current Bayesian and nonparametric methods for inferring model parameters in stochastic epidemic models.

1.4 Bayesian Inference for Epidemic Models

Bayesian methods have been widely applied to infectious disease data; for an introduction see e.g. Held et al. (2019). The first Bayesian inference methods for stochastic epidemic models were developed in Gibson and Renshaw (1998) and O'Neill and Roberts (1999). In Gibson and Renshaw (1998), the authors inferred the parameters of the general stochastic epidemic model and allowed the population to change in size due to birth and death. The authors developed a reversible-jump MCMC method to allow for the population size, N, to change. This allowed for births and deaths to occur whilst the outbreak was continuing. In O'Neill and Roberts (1999), the authors developed a different approach to infer the model parameters in the general stochastic epidemic model. The authors placed a prior distribution on the infection and removal rates and then used a Gibbs sampler in an MCMC framework to sample from the corresponding conditional distributions. They also assumed the times individuals were infected are unobserved and augmented the likelihood function to treat these times as parameters. They developed a Metropolis-Hastings algorithm to infer the infection times.

These models have been extended to allow for individual-level infection rates. For example, in Boender et al. (2007), the authors fitted an individual-level stochastic epidemic model to an outbreak of Avian Influenza in the Netherlands. To model the infection rate from farm i to j, the authors proposed several infection rate functions which depended on the distance between farms i and j. The authors used maximum likelihood methods to estimate the model parameters. Finally, they chose the best

fitting of the models using information criteria. In Jewell et al. (2009), the authors also fitted an individual-level model for an outbreak of Foot and Mouth Disease, and adapted the data augmentation method proposed in O'Neill and Roberts (1999) to infer the infection times of farms which were culled without being tested for the disease, as well as improving efficiency of the method.

1.4.1 Bayesian Nonparametric Methods for Stochastic Epidemic Models

Parametric models, like those proposed by Boender et al. (2007) and Jewell et al. (2009), can be restrictive, as we need to propose exact parametric forms for the infection rate. Given the limited data we observe about the outbreak, accurately specifying plausible functions is challenging and we aim to alleviate these problems and relax assumptions by estimating the infection rate from the data. Bayesian nonparametric methods allows us to fully infer the infection rate without having to impose strict assumptions about the form of the infection rate.

More recently, inference for Bayesian nonparametric epidemic models has been developed. With regards to stochastic epidemic models, step functions and B-Splines have been used in a nonparametric Bayesian framework to estimate a time-dependent infection rate Knock and Kypraios (2014). The authors first modelled the infection rate as a step function and partitioned the time period of the epidemic into k segments, $i_0 = s_1 < s_2 < \ldots < s_k = r_n$, where i_0 is the first infection time and r_n the final removal time. They define the infection rate at time of partition j to be β_j, which allows them to write the infection rate as:

$$\beta(t) = \sum_{j=1}^{k} \beta_j 1_{s_j < t < s_{j+1}}.$$

They placed prior distributions $\beta_1, \ldots, \beta_k$ and used a Gibbs sampler to infer the values of these parameters. Using dependent prior distributions, where $\mathbb{E}\left[\beta_{j+1}|\beta_j\right]$, allowed them to control the smoothness of the function. They then implemented a Metropolis Hastings algorithm to infer the number and locations of the change points

$s_2, \dots s_{k-1}$. They also proposed a similar B-spline method, where the value of the B-spline coefficients and number of knots are inferred using a MCMC framework.

In Xu (2015) and Xu et al. (2016), the authors used GPs to estimate time-dependent infection rates for the general epidemic model. They assume infections occur at time points given by an inhomogeneous Poisson process with rate $\beta(t)$, and develop a nonparametric inference method based on Adams et al. (2009). Xu (2015) modelled the infection rate using a Sigmoidal Gaussian Cox Process, where the infection rate is given by

$$\beta(t) = \beta_0 \sigma(g(t)),$$

where $\sigma(x) = \{1 + \exp\{-x\}\}^{-1}$ and g is a function with a GP prior distribution on it. In Xu et al. (2016), the authors developed an MCMC algorithm to infer the maximum infection rate, β_0, which is equivalent to the rate parameter of the homogeneous Poisson process λ^*, as well as the time-dependent function. This method requires data augmentation as the number of thinned points, m, and their locations, $\mathbf{t} = \{t_1, \dots, t_m\}$ are inferred, as well as β_0 and the function g. In Xu et al. (2016), the authors also developed a similar nonparametric method inferring infection rates of the form $\beta(t)S_tI_t$. Modelling the infection rate by a GP allows for the smoothness of the infection rate function to be inferred as well. This avoids the need for sequentially dependent prior distributions.

The majority of research in nonparametric estimation for epidemic models has looked at estimation for time-dependent infection rates. Our aim is to develop a method for estimating infection rates where they depend on other factors, such as a spatial covariate or type of individual. Our new methods should allow us to learn the infection rate from observed removal times, and we should only need to make very basic assumptions about the infection rate *a priori*, such as the rate being a positive continuous function.

Although this work is related to that described in Xu (2015) and Xu et al. (2016), namely the development of Bayesian nonparametric methods for stochastic epidemic models and the use of Gaussian Process prior distributions, the methods are materially different and address a distinct kind of problem. In Xu et al. (2016), the authors

developed methods for modelling time-dependent infection rates nonparametrically, and the work was done on a population-level. This work contributes to the field by developing Bayesian nonparametric methods for models where the infection rate differs between individuals.

1.5 Gaussian Processes

The nonparametric method we will be using is the GP prior distribution. Intuitively, a GP prior distribution is a prior distribution over a space of functions. We build our assumptions about the function we are modelling through the prior distribution hyperparameters. We will be considering GP prior distributions for modelling one-dimensional functions. We now give a full definition, outline possible covariance functions and give an example in a regression context.

1.5.1 Definition

Given an input space χ, typically a subset of $\mathbb{R}^n$, a GP is a probability distribution over χ, such that the joint distribution over any finite subset $\mathbf{x} = \{x_1, \ldots, x_n\} \in \chi$ follows a multivariate normal distribution. We use the definition from Rasmussen and Williams (2006)[§2, p. 13] and define a GP as follows:

Definition: *A GP is a collection of random variables any finite number of which have a joint Gaussian distribution.*

A GP prior distribution on a function f can be completely defined through its mean function and covariance matrix, which are given by:

$$\mu(x_i) = \mathbb{E}[f(x_i)], \quad i = 1, \ldots, n,$$
$$\Sigma_{i,j} = \mathbb{E}[(f(x_i) - \mu(x_i))(f(x_j) - \mu(x_j))], \quad i, j = 1, \ldots, n.$$

The mean function evaluated at x, $\mu(x)$, specifies the expected value of $f(x)$ *a priori*; throughout this thesis we will use a zero-mean GP prior distribution, where $\mu(x) = 0$

for all $x \in \chi$. Using a zero-mean prior distribution is not restrictive as long as we ensure the distribution covers a large space of plausible functions. The $(i, j)^{th}$ element of the covariance matrix Σ specifies the relationship between $f(x_i)$ and $f(x_j)$ *a priori*. We denote the prior distribution by:

$$f \sim \mathcal{GP}\big(\mu,\, \Sigma\big).$$

The advantage of using a GP over another nonparametric method is that a GP prior distribution allows us to assume a level of smoothness for the infection rate instead of approximating the biological processes that causes the disease to spread. Other methods such as step functions or B-splines do allow smoothness to be considered, but indirectly through more complicated prior constructions. They also require change points, the location and number of which need to be estimated.

1.5.2 Covariance Functions

The covariance matrix is how we allow for the majority of our assumptions about the function we are modelling to be made. Although these assumptions are less strict than in parametric methods, we must specify if the function is continuous or the differentiability class. We also need to consider how quickly moving the function over the input space is or whether it is periodic. Using a covariance function means the covariance between any two points, $f(x_i)$ and $f(x_j)$, is a function of the inputs, x_i and x_j. The covariance is therefore given by:

$$\Sigma_{ij} = k(x_i, x_j).$$

There is a wide variety of choices for the covariance function, for a full overview see Rasmussen and Williams (2006, §4). We now recall several relevant functions.

The squared exponential covariance function is a widely used choice when modelling continuous functions. The squared exponential covariance function is given by:

$$k(x_i, x_j;\, \alpha, l) = \alpha^2 \exp\left\{-\frac{(x_i - x_j)^2}{l^2}\right\} \tag{1.2}$$

and has two hyperparameters: the signal variance parameter α and the length scale parameter l. The signal variance parameter controls the overall variance in the prior distribution and the length scale parameter describes how quickly changing we assume the function to be. The value of the length scale parameter ensures how similar the output of the function is for two nearby input values.

The Mátern covariance function can be used to model functions which are continuous but not differentiable and is related to the squared exponential function. The covariance function is given by:

$$k(x_i, x_j; \nu, l) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{(x_i - x_j)\sqrt{2\nu}}{l} \right)^{\nu} K_\nu \left(\frac{(x_i - x_j)\sqrt{2\nu}}{l} \right),$$

where K_ν is a modified Besel function of the ν^{th} kind. As in the squared exponential, l is a length scale parameter. The parameter ν is a differentiability parameter and this covariance function models functions that are $\lfloor \nu - 1 \rfloor$ mean square differentiable. As $\nu \longrightarrow \infty$, the covariance function becomes the squared exponential covariance function.

The rational quadratic covariance function is given by:

$$k(x_i, x_j; \alpha, \gamma, l) = \alpha^2 \left(1 + \frac{(x_i - x_j)^2}{2\gamma l^2} \right)^{-\gamma}$$

and is analogous to the summation of many squared exponential covariance functions with different length scales. The scaling parameter γ defines the weight of the smaller length scale variation to longer length scale variation. As with the Mátern covariance function, as $\gamma \longrightarrow \infty$, the covariance function tends to the squared exponential function.

We can also model functions that are periodic by using the periodic covariance function. This is given by:

$$k(x_i, x_j; \alpha, \rho, l) = \alpha \exp \left\{ -2 \frac{\sin^2(\pi |x_i - x_j| / \rho)}{l^2} \right\}.$$

The parameter ρ defines the periodicity, the length scale is given by l, and α is the signal variance parameter.

Once we have chosen a covariance function, we can then generate samples from the GP. Consider the function f evaluated at the points in the input set $\mathbf{x} = \{x_1, \ldots, x_n\} \subset \chi$. We place the following GP prior distribution on f:

$$f \sim \mathcal{GP}(0,\, \Sigma), \quad \Sigma_{ij} = k(x_i, x_j),$$

where k is a suitably chosen covariance function. To generate samples from this distribution we use a Cholesky based method (algorithm 3). This method transforms samples drawn from a standard normal distribution by using the Cholesky decomposition of the covariance matrix.

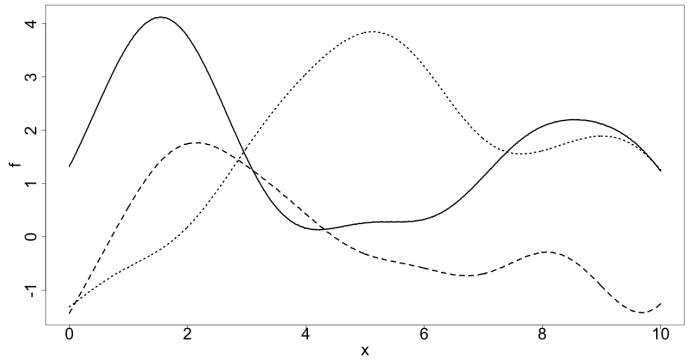
Algorithm 3 Draw a sample from a GP Prior Distribution

1: Construct Σ
2: Compute L, where $LL^T = \Sigma$
3: Sample $\mathbf{u} \sim N(0, \mathbf{I})$
4: Compute $f = L\mathbf{u}$

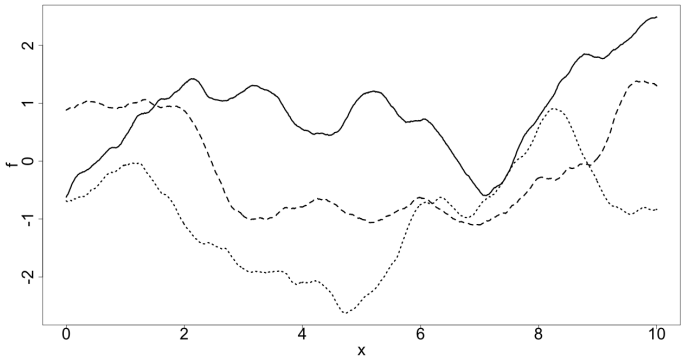
We now demonstrate these four covariance functions in figure 1.3. We place a zero-mean GP prior distribution on a function f and use $\mathbf{x} = \{0, 0.01, 0.02, \ldots, 10\}$ as the input set. We then draw three samples from the prior distribution for each of the covariance functions. The samples from the prior distribution with the squared exponential function are smooth, whereas due to our choice of hyperparameters, the samples from the prior distribution with the Mátern covariance function are only once mean square differentiable. We can see the effect of the scaling parameter in the prior distribution with rational quadratic covariance function, which is equivalent to summing samples from prior distributions with a squared exponential covariance function and different length scales. The periodic function generates functions which are periodic over the input space, and we have chosen the period to be 2.

1.5.3 Gaussian Process Regression

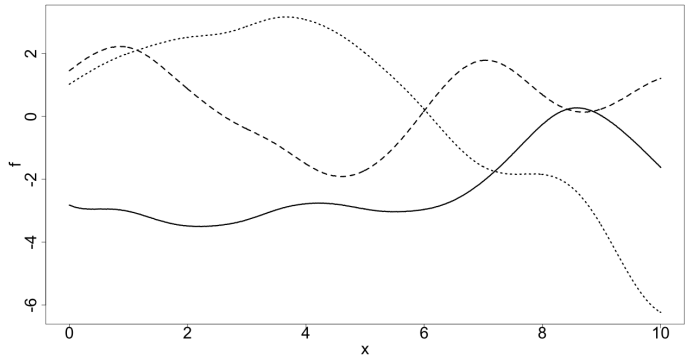
Before developing GP methods for stochastic epidemic models, we first illustrate GPs and Bayesian nonparametric methods with a basic regression example. Suppose



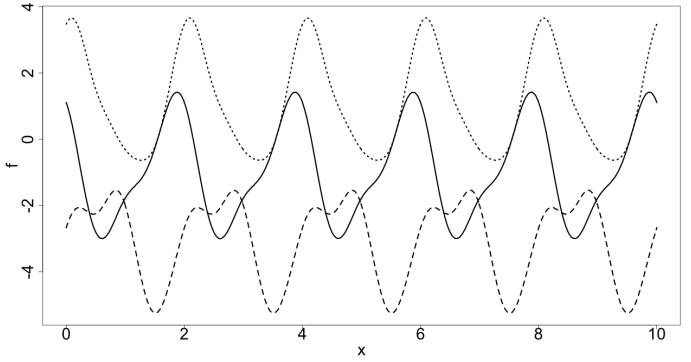
(a) Three samples from a GP prior distribution with squared exponential covariance function with $\alpha = 3$, and $l = 2$.



(b) Three samples from a GP prior distribution with Mátern covariance function with $\nu = 5/2$, and $l = 2$.



(c) Three samples from a GP prior distribution with rational quadratic covariance function with $\alpha = 3$, $l = 2$, and $\gamma = 1$.



(d) Three samples from a GP prior distribution with periodic covariance function with $\alpha = 3$, $l = 2$, and $\rho = 2$.

Figure 1.3: Samples drawn from GP prior distributions with different covariance functions.

we observe n pairs of points, $(\mathbf{x}, \mathbf{y}) = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, and we wish to fit the following model:

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2).$$

We model f by placing a GP prior distribution on it, and, for this example, we assume σ^2 is known. The likelihood function of $\mathbf{y}$ given the function f and the variance σ^2 is given by:

$$\pi(\mathbf{y}|f, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{ -\sum_{i=1}^{n} \frac{(y_i - f(x_i))^2}{2\sigma^2} \right\}.$$

We place the following GP prior distribution function on f using the squared exponential covariance function:

$$f \sim \mathcal{GP}(0, \Sigma), \quad \Sigma_{ij} = k(x_i, x_j; \alpha, l).$$

By Bayes' theorem, the posterior distribution is given by:

$$\begin{aligned} \pi(f|\mathbf{x}, \mathbf{y}, \sigma^2) &\propto \mathcal{GP}(f; 0, \Sigma)\pi(\mathbf{y}|\mathbf{x}, f, \sigma^2) \\ &\propto \mathcal{GP}(f; 0, \Sigma)\mathcal{N}(\mathbf{y}; f, \sigma^2 I) \\ &\propto \mathcal{GP}\left(f; \frac{1}{\sigma^2}(\Sigma^{-1} + \frac{1}{\sigma^2} I)^{-1}\mathbf{y}, (\Sigma^{-1} + \frac{1}{\sigma^2} I)^{-1} \right), \end{aligned}$$

where $\mathcal{N}(\mathbf{x}; \mathbf{0}, \Sigma)$ denotes the density function of a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix Σ, evaluated at the vector $\mathbf{x}$.

By way of example, we generate 80 points using the function $y = 0.5\log(x) - 4\Phi(x - 5)$ and add noise from a $N(0, 0.5^2)$ distribution, where $\Phi(x)$ is the cumulative distribution function of a standard normal density. We choose this function as it is challenging to model parametrically with no prior knowledge of the parametric form. For the GP prior distribution, we set $\alpha = 10$ and $l = 3$. We show the posterior mean and credible interval in figure 1.4. The posterior mean gives a reasonable approximation of the genera ting function and the generating function is contained entirely within the credible interval.

Assessing the model fit when using GPs is challenging as the number of parameters in a Bayesian Nonparametric model has a different meaning to the number of parameters in a parametric model. This means we cannot use typical measures of

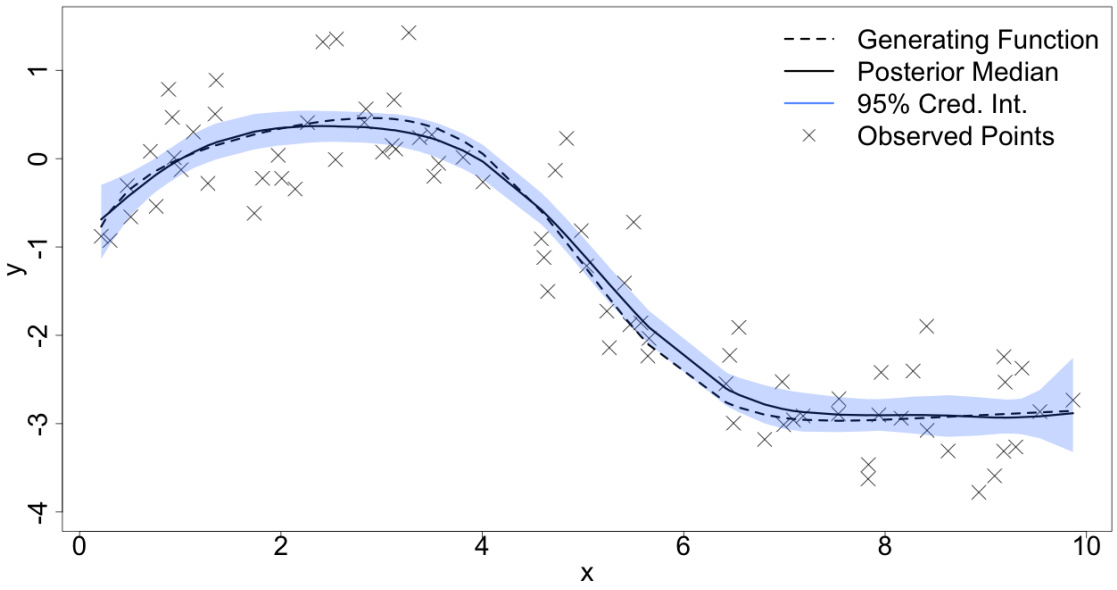


Figure 1.4: The regression example with GP prior distribution and simulated data. We show the true function (dotted), the posterior median function (solid), and the 95% credible interval (in blue).

model assessment such as Information Criteria or likelihood ratio tests. Diagnostics for checking the posterior distribution are also challenging as the GP models a function at a number of points, and we would need to check the posterior distribution at each point. For example, in this regression example, we evaluate the function at 80 points, resulting in a posterior distribution with 80 dimensions, or 80 marginal distributions. Model assessment for Bayesian Nonparametric models is an open area of research and we use several different methods to assess our models. We run simulation studies with a large number of independent simulations, where we simulate data from a given model and fit the model to the simulated data; the aim of this is that MCMC chains from well fitting and mixing models will converge to the same answer. We also make use of the posterior predictive distribution to compare our results to observed data. These methods are not perfect and are in some respects imprecise, but do allow us to compare models and reassure us that the Markov chains have converged to the posterior distribution.

1.5.4 Gaussian Process Notation

Much of the work on GPs has been carried out from a Machine Learning and Computer Science standpoint, indeed, one of the most widely read texts in the area is titled 'Gaussian Processes for Machine Learning' (Rasmussen and Williams, 2006). Although the methodology in Statistics and Machine Learning is identical, there are differing paradigms in the two areas. In Statistics, we are interested in inferring a function, and as such, the notation describes functions. Researchers in machine learning are often interested in making predictions about the value of a function in an area of the domain where no data is observed. Examples of this include audio reconstruction, where a missing part of a sound wave needs to be predicted given the parts of the wave that were observed, or GP emulation, where given a small number of outputs from a very complex model that is time consuming to run, a large number of outputs are inferred. Since the aim in machine learning is mostly to make predictions about the value of a function given the observed data, the notation and language used there is different to that used in statistical literature.

We now outline typical notation used in Machine Learning. At some input points $\mathbf{x} = \{x_1, \ldots, x_n\}$, we observe the values of the function $\mathbf{f} = \{f_1, \ldots, f_n\}$, where $f_i = f(x_i)$. We assume the mean of the GP prior distribution at the points of $\mathbf{x}$ is given by $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_n\}$. To construct a covariance matrix, we choose a covariance function k, and compute the covariance matrix, Σ, based on k evaluated at the input points $\mathbf{x}$. The prior distribution is therefore given by the following GP prior distribution, although we can view it as a multivariate normal distribution:

$$\mathbf{f} \sim \mathcal{GP}\left(\boldsymbol{\mu}, \Sigma\right).$$

From a statistical standpoint, we consider the problem of inferring functions. We consider a function f and after choosing suitable prior mean and covariance functions, μ and k respectively, we write the GP prior distribution as:

$$f \sim \mathcal{GP}(\mu, k).$$

However, as we consider the function f as an infinite dimensional object, we cannot evaluate or perform computations with this distribution. We introduce the set of

locations at which we wish to evaluate the function, $\mathbf{x} = \{x_1, \ldots, x_n\}$. Using this set, we construct a covariance matrix Σ, and evaluate the mean function at these points. The GP prior distribution becomes a multivariate normal distribution, although as we are inferring functions, we still choose to view it as a GP prior distribution over a space of functions. We denote it as

$$f \sim \mathcal{GP}(\mu, \Sigma).$$

We choose to use Σ over k as in chapter 3 we investigate Multi-Output GP prior distributions, where we use the same covariance function, k, over different input sets to construct difference covariance matrices and functions.

1.6 Structure of the Thesis

This thesis is divided into three chapters detailing the methodological advancements and one chapter implementing the methods to real life data sets. In chapter 2 we outline our nonparametric method for modelling infection rate functions where the infection rate between any two individuals depends on a single, continuous covariate, such as the Euclidean distance between the individuals. We do this using GP prior distributions, and we describe the Bayesian nonparametric approach in this chapter. We also develop a MCMC algorithm in this chapter to generate samples from this appropriate posterior distribution. In chapter 3, we turn our attention to multi-type epidemics, where the infection rate between any two individuals not only depends on a single, continuous covariate, but also the type of the susceptible individual. For example, the types in an outbreak of Foot and Mouth disease may be sheep farms and cattle farms. Our main method of modelling this is through Multi-Output Gaussian Processes, although we also outline a fixed-effects approach for this problem. The final methodological developments are given in chapter 4, where we model infection rate functions which depend on multiple continuous covariates, for example, Euclidean distance and the size of the susceptible individual. We do this by extending our GP technique into n dimensions.

In chapter 5, we apply our methods to two data sets and compare the results with previously used methods. The first data set is an outbreak of Avian Influenza in the Netherlands in 2003, in which 233 farms were infected with the disease, 30 million birds were culled, and one person died after contracting the disease. We infer the infection rate function for this outbreak as well as estimating which farms were infected with the disease and for how long. This has not previously been investigated for this outbreak. We then investigate various possible culling strategies for the outbreak by way of the posterior predictive distribution. We apply our methods to data taken from an outbreak of Foot and Mouth Disease in the UK in 2001. We use our multi-type methods to understand the difference in susceptibility between different types of animals.

CHAPTER 2

Bayesian Nonparametric Methods for Infection Rate Functions with One Covariate

2.1 Introduction

Individual-level models allow us to specify the infection rate between each pair of individuals, and have been successfully implemented in a number of cases (see e.g. Keeling, 2001; Boender et al., 2007; Jewell et al., 2009). To model the spread of a disease, we assume an infected individual i infects a susceptible individual j at the time points of Poisson process with rate β_{ij}. Often, even for small populations, the number of pair-wise infection rate parameters is too large to estimate. Instead, we model the pair-wise infection rates by an infection rate function. The infection rate function models the pair-wise infection rate between any two individuals as a function of the relationship between them. For example, in an outbreak of Swine Fever, we may assume the infection rate from one farm to another depends on the number of pigs on each farm, the type of housing for the pigs, and the location of the farm.

It is challenging to quantify or describe any biological rationale in these functions. For example, in Alexandersen et al. (2003), the authors state that under certain

weather conditions the Foot and Mouth disease virus can be carried of gusts on wind for short distances, and that cattle are more susceptible to this form of transmission than pigs. Constructing a mathematical function which contains this information can be challenging.

Parametric infection rate functions can also lack justification from the data. This is particularly acute when there is a spatial element to the spread of a disease. Consider an outbreak of Avian Influenza among poultry farms, where we model the infection rate between farms as a function of the distance between them. In Boender et al. (2007), the authors chose to propose functions of the form

$$f(d) = \frac{\beta_0}{1 + (d/\beta_1)^{\beta_2}}, \quad \beta_i > 0$$

for this problem. Although proposing a function form may be straightforward, justifying it is difficult given the observed data. The parametric forms are also strict assumptions about the infection rate functions. In this chapter, we develop a Bayesian nonparametric method for modelling infection rate functions which depend on a single covariate, for example, the Euclidean distance. This will allow us to build a much more flexible model and learn the infection rate function from the data.

We will use GPs to nonparametrically model the infection rate functions. GPs have been used to model functions in a wide variety of contexts, for example, modelling the number of births on each day of the year (Gelman et al., 2013, §21), or the price of houses in the UK (Hensman et al., 2013). Using GPs will allow us to learn the infection rate functions from the data without requiring specific biological justification or making strict assumptions about how the disease spread.

We will follow previous Bayesian methods such as those in O'Neill and Roberts (1999), Xu (2015) and Stockdale et al. (2018), and develop efficient, data augmentation, MCMC methods. This will allow us to infer the infection rate function alongside other parameters, as well as infer the times individuals were infected.

2.1.1 Layout of Chapter

We begin the chapter by defining an individual-level stochastic epidemic model and deriving an appropriate likelihood function. In section 2.3 we outline inference for parametric models, and a number of model selection methods. Section 2.4 concerns GPs and their use in our nonparametric methodology and inference for the length scale parameter of the commonly used squared exponential covariance function. We then develop an MCMC algorithm for our Bayesian nonparametric method in section 2.5. This method can have a high time and memory cost when implemented for an outbreak in a large population, so we outline approximation methods in section 2.6 and discuss methods for improving the MCMC framework for outbreaks of diseases in large populations. Section 2.8 gives a comprehensive overview of the methodology applied to various simulated data sets.

This chapter sets out the Bayesian nonparametric framework we will be using throughout the thesis and provides an exposition of the fundamental of this novel method. It serves to bridge two areas: individual-level stochastic epidemic models and GPs. It outlines a method for how to include monotonicity information in the model and provides a framework to carry out efficient MCMC methods. A simulation study confirms the accuracy of our method.

2.2 Transmission Model and Likelihood

This section concerns the standard stochastic epidemic model (see e.g. Andersson and Britton, 2000). During an outbreak of a disease individuals are susceptible to the disease, infected, or removed from the process. The data concerning the outbreak is the times at which individuals entered the infected and removed compartments. We first outline the model and then derive an appropriate likelihood function given the data.

2.2.1 Individual-Level Stochastic Epidemic Model

Consider an outbreak of a disease in a population with N individuals, where the individuals are labeled arbitrarily from $1, \ldots, N$. At some time $t \geq 0$, we assume each individual is either susceptible to contracting the disease, infected with the disease, or removed as they are no longer infectious. Consider an infected individual j and a susceptible individual k. We assume individual j makes infectious contact with individual k at the time points given by a Poisson process with rate $\beta_{j,k}$. If individual k is infected, they remain so for a time period drawn from a Gamma distribution with shape parameter α and rate parameter γ. At the end of the infectious period, the individual enters the removed class and cannot be reinfected. We assume the Poisson processes between each pair of individuals are independent and the infectious period distribution for each individual is independent of any other individual. The outbreak is declared over when there are no infected individuals remaining and we define n to be the final size or total number of individuals infected in the outbreak.

To construct the appropriate likelihood function given the times individuals were infected and removed: we denote the time individual j was infected and removed by i_j and r_j respectively, we label the infected individuals from $1, \ldots, n$ by removal time, such that $r_1 < r_2 < \ldots r_n$, we label the remaining individual in the population $n+1, \ldots, N$, and define their infection and removal times to be $i_j = r_j = \infty$. We assume a single individual becomes infected initially, who we label κ, and we define i_κ to be their infection time, and we define the sets $\mathbf{i} = \{i_1, \ldots, i_{\kappa-1}, i_{\kappa+1}, \ldots i_n\}$ and $\mathbf{r} = \{r_1, \ldots, r_n\}$.

When constructing the likelihood function, we split the outbreak into two processes – the infection process and the removal process. The infection process contributes to the likelihood function in two ways: the first is through individuals avoiding infection and the second is individuals becoming infected. Consider a susceptible individual k, the contribution of them avoiding infection from individual j is given by

$$\psi_{j,k} = \exp\{-\beta_{j,k}((r_j \wedge i_k) - (i_j \wedge i_k))\},$$

where $a \wedge b = \min\{a, b\}$. The difference of minimums is the length of time individual j could have infected individual k. This notation condenses the following three cases where individual j is infected:

$$(r_j \wedge i_k) - (i_j \wedge i_k) = \begin{cases} r_j - i_j & \text{if } j \text{ is removed before } k \text{ is infected, or } k \text{ is not infected,} \\ i_k - i_j & \text{if } k \text{ is infected after } j \text{ is infected and before } j \text{ is removed,} \\ 0 & \text{otherwise.} \end{cases}$$

A diagram for the first two cases is shown in figure 2.1.

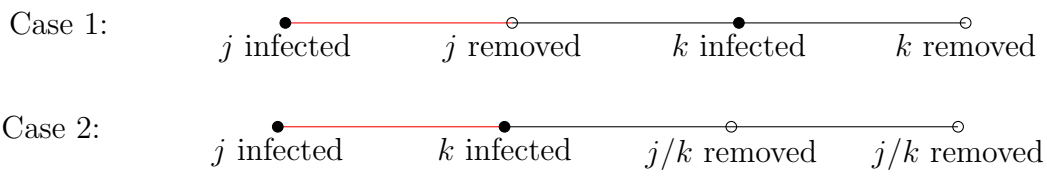


Figure 2.1: Two of the cases for the infection and removal times of two individuals j and k. The black and white circles correspond to infections and removals respectively. The red segment is the time individual j can infect individual k.

The event individual j becomes infected contributes to the likelihood through the overall hazard rate of the infection:

$$\phi_j = \sum_{k \in \mathcal{Y}_j} \beta_{k,j},$$

where $\mathcal{Y}_j = \{k : i_k < i_j < r_k\}$ is the set of individuals who were infectious immediately before j became infected. We do not consider the contribution of the initially infected individual, labelled κ, becoming infected as we do not model how the epidemic started, instead conditioning on the label κ and the corresponding infection time.

For the removal process, we consider how long each infected individual remained infectious with respect to the infectious period distribution. The removal process part of the likelihood function is given by:

$$\pi_{rem} = \prod_{j=1}^{n} h(r_j - i_j | \lambda, \gamma),$$

where $h(\cdot|\lambda,\gamma)$ is the Gamma probability density function, with shape parameter λ and rate parameter γ. Combining these three parts gives the full likelihood function, which is given by:

$$\begin{aligned}\pi(\mathbf{i},\mathbf{r}|\beta,\lambda,\gamma,\kappa,i_\kappa) &= \left(\prod_{j=1}^{n}\prod_{k=1}^{N}\psi_{j,k}\right)\left(\prod_{\substack{j=1\\ j\neq\kappa}}^{n}\phi_j\right)\prod_{j=1}^{n}h(r_j-i_j|\lambda,\gamma) \qquad (2.1)\\ &= \exp\left\{-\sum_{j=1}^{n}\sum_{k=1}^{N}\beta_{j,k}\left((r_j\wedge i_k)-(i_j\wedge i_k)\right)\right\}\\ &\quad\times\prod_{\substack{j=1\\ j\neq\kappa}}^{n}\left(\sum_{k\in\mathcal{Y}_j}\beta_{k,j}\right)\prod_{j=1}^{n}h(r_j-i_j|\lambda,\gamma).\end{aligned}$$

2.3 Bayesian Inference for Parametric Models

Current methods for individual level stochastic epidemic models are parametric. For parametric models, we propose one or more plausible parametric forms for the infection rate function and estimate the parameters from the data. These parametric forms are often based on previous models used in the literature or because of their mathematical convenience. In this section, we argue that once parametric forms have been proposed, there are a wide variety of model fitting and model assessment methods to choose from, and although they may not be straightforward to implement or suffer from their own limitations, this is not a flaw with the parametric approach. The flaw comes when proposing the parametric forms, as we have no justification from the data for these forms. . We no give a brief introduction of how to implement a parametric approach. We explicitly state the functional form of the infection rate. We outline possible infection rate functions, then provide a framework for Bayesian inference for the parametric infection rate. We end by giving an example of this method for a logistic model. These methods we describe are by no means the most suitable or optimal in every case, but are given as an example of how parametric methods can be applied.

2.3.1 The Choice of Infection Rate

We begin by assuming the infection rate from individual j to individual k depends on some characteristic of the individuals j and k. This characteristic is defined through a covariate x which describes the relationship between individuals j and k. The infection rate function is

$$\beta_{j,k} = \beta(x_{j,k}).$$

We can propose many parametric forms for the function β, using, for example, logistic or exponential functions. However, these proposals are arbitrary and we may not be able to justify our choices from the data. Choosing a parametric form for β is also a strong assumption, as we are explicitly stating the infection rate between each pair of individuals. In this section, we show that several methods of model fitting and model assessment can be used, and although they may be challenging to implement in practice, the main flaw with parametric models is proposing suitable functions.

Suppose we have chosen a parametric form for β and wish to infer the parameters of this function, denoted by the vector $\boldsymbol{\beta}$. We place a prior distribution, $\pi(\boldsymbol{\beta})$, on these parameters and a prior distribution on the infectious period distribution rate parameter γ. We follow Jewell et al. (2009) and fix the rate parameter $\lambda > 1$ as this gives a distribution the mean of which can be defined through γ. We assume the infection times are observed as we are demonstrating methods for the infection rate function. The posterior distribution is given by:

$$\pi\left(\boldsymbol{\beta}, \gamma | \mathbf{i}, \mathbf{r}, \lambda, i_\kappa, \kappa\right) \propto \pi(\mathbf{i}, \mathbf{r} | \boldsymbol{\beta}, \lambda, \gamma, \kappa, i_\kappa)\pi(\boldsymbol{\beta})\pi(\gamma) \tag{2.2}$$

$$\pi\left(\boldsymbol{\beta}, \gamma | \mathbf{i}, \mathbf{r}, \lambda, i_\kappa, \kappa\right) \propto \exp\left\{ -\sum_{j=1}^{n}\sum_{k=1}^{N} \beta(x_{j,k}) \left((r_j \wedge i_k) - (i_j \wedge i_k)\right) \right\}$$
$$\times \prod_{\substack{j=1 \\ j\neq\kappa}}^{n} \left(\sum_{k\in\mathcal{Y}_j} \beta(x_{k,j}) \right) \prod_{j=1}^{n} h(r_j - i_j | \lambda, \gamma)\pi(\boldsymbol{\beta})\pi(\gamma).$$

For the majority of parametric forms for β, we will need to develop an MCMC algorithm to infer the model parameters.

2.3.2 Inference for a Logistic Infection Rate

We now demonstrate an inference method for a logistic infection rate. We adopt a similar approach to Boender et al. (2007) and suppose the infection rate between individuals j and k depends on the Euclidean distance between them, $d_{j,k}$, and the infection rate function is given by:

$$\beta(x_{j,k}) = \frac{\beta_0}{1 + (d_{j,k}/\beta_1)^{\beta_2}},$$

where $\boldsymbol{\beta} = \{\beta_0, \beta_1, \beta_2\}$ are parameters to be inferred. We place independent exponential prior distributions on β_0, β_1 and β_2, with respective rates χ_{β_0}, χ_{β_1} and χ_{β_2}. We place an exponential prior distribution on the rate parameter γ, which has rate χ_γ. The posterior distribution is given by:

$$\begin{aligned}
\pi(\boldsymbol{\beta}, \gamma | \lambda, \mathbf{i}, \mathbf{r}, \kappa, i_\kappa) \propto \exp &\left\{ -\beta_0 \sum_{j=1}^{n} \sum_{k=1}^{N} \frac{1}{1 + \left(\frac{d_{j,k}}{\beta_1}\right)^{\beta_2}} \left((r_j \wedge i_k) - (i_j \wedge i_k) \right) \right\} \\
&\times \beta_0^{n-1} \prod_{\substack{j=1 \\ j \neq \kappa}}^{n} \left(\sum_{k \in \mathcal{Y}_j} \frac{1}{1 + (d_{j,k}/\beta_1)^{\beta_2}} \right) \frac{\gamma^{n\lambda}}{\Gamma(\lambda)^n} \prod_{j=1}^{n} (r_j - i_j)^{\lambda - 1} \\
&\times \exp\left\{ -\gamma \sum_{j=1}^{n} (r_j - i_j) \right\} \exp\{-\beta_0 \chi_{\beta_0}\} \exp\{-\chi_{\beta_1}\} \\
&\times \exp\{-\beta_2 \chi_{\beta_2}\} \exp\{-\gamma \chi_\gamma\}.
\end{aligned}$$

We use the following MCMC algorithm (algorithm 4) to infer the values of the infection rate parameters $\beta_0 > 0$, $\beta_1 > 0$ and $\beta_2 > 0$ and the infectious period distribution rate parameter γ.

We now outline each of the sampling steps from the MCMC framework in more detail.

2.3.2.1 Sampling γ

The full conditional posterior distribution for γ is given by:

$$\pi(\gamma | \lambda, \mathbf{i}, \mathbf{r}, \chi_\gamma) \propto \gamma \exp\{-\gamma \chi_\gamma\} \prod_{j=1}^{n} h(r_j - i_j | \lambda, \gamma).$$

Algorithm 4 Structure of the MCMC algorithm for the logistic model

1: Initialise the chain with estimates $\gamma^{(0)}$, $\beta_0^{(0)}$, $\beta_1^{(0)}$, and $\beta_2^{(0)}$.

Repeat the following steps

2: Sample γ from the conditional distribution $\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r}, \chi_\gamma)$ using a Gibbs step

3: Sample β_0 from the conditional distribution $\pi(\beta_0|\beta_1, \beta_2, \lambda, \gamma, \mathbf{i}, \mathbf{r}, \chi_{\beta_0})$ using a Gibbs step

4: Sample the values of β_1 using a Metropolis Hastings Random Walk step

5: Sample the values of β_2 using a Metropolis Hastings Random Walk step

As we have used a conjugate prior distribution, the posterior distribution has the following closed form:

$$\gamma|\lambda, \mathbf{i}, \mathbf{r}, \mathbf{x}, \lambda, \chi_\gamma \sim \Gamma\left(1 + n\lambda,\ \chi_\gamma + \sum_{j=1}^{n}(r_j - i_j)\right).$$

Although the full conditional distribution for γ is known and has a closed form, we still include it in our MCMC approach as a demonstration for methods later on.

2.3.2.2 Sampling β_0

The conditional posterior distribution for β_0 has a closed form due to prior conjugacy, and is given by:

$$\pi(\beta_0|\beta_1, \beta_2, \mathbf{i}, \mathbf{r}, \chi_{\beta_0}) \propto \beta_0^{(n+1)-1} \exp\left\{-\beta_0\left(\chi_{\beta_0} + \Psi\right)\right\}, \tag{2.3}$$

where $\Psi = \sum_{j=1}^{n}\sum_{k=1}^{N} \frac{1}{1+(d_{j,k}/\beta_1)^{\beta_2}}((r_j \wedge i_k) - (i_j \wedge i_k))$. Hence the posterior distribution is given by the following Gamma distribution:

$$\beta_0|\beta_1, \beta_2, \mathbf{i}, \mathbf{r}, \chi_{\beta_0} \sim \Gamma\left(n + 1, \chi_{\beta_0} + \Psi\right).$$

2.3.2.3 Sampling β_1 and β_2

We update the values of β_1 and β_2 jointly in the MCMC algorithm. We use a Metropolis Hastings Random Walk algorithm and given the current values β_1 and β_2, we propose updating them to β_1' and β_2' by drawing these values from the proposal

Parameter	Value
β_0	0.01
β_1	0.3
β_2	3
λ	2
γ	0.75

Table 2.1: Parameter values used to simulate the outbreak with the logistic infection rate shown in section 2.3.2.

distribution:

$$\begin{pmatrix}\beta_1'\\ \beta_2'\end{pmatrix} \sim N\left(\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right),$$

where σ_1^2 and σ_2^2 are the proposal variances for β_1 and β_2 respectively and ρ is the proposal correlation between β_1 and β_2. We choose values for the parameters in the proposal distribution, ρ, σ_1 and σ_2, such that we achieve a good mixing of the Markov chain, often basing them on trial runs of the MCMC algorithm. In more sophisticated algorithms, we can use an adaptive algorithm. We accept these proposals with probability

$$p_{acc} = \frac{\pi(\mathbf{i}, \mathbf{r}|\beta_0, \beta_1', \beta_2', \lambda, \gamma, \kappa, i_\kappa)}{\pi(\mathbf{i}, \mathbf{r}|\beta_0, \beta_1, \beta_2, \lambda, \gamma, \kappa, i_\kappa)} \wedge 1.$$

2.3.3 Results

We simulate an outbreak of a disease in a population of size 200 where the coordinates are generated uniformly at random on a unit square, and the parameter values shown in table 2.1.

One realisation of this outbreak generated 117 infected individuals, we recorded the infection and removal times of the individuals. We then infer the model parameters given these times. We propose the forms shown in table 2.2. The first model assumes homogeneous mixing, whereas the remaining models depend on the pair-wise

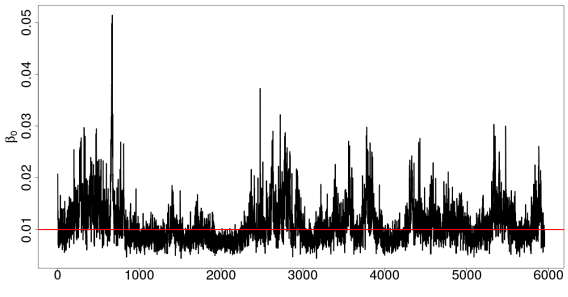
Model	Form
1	β_0
2	$\frac{\beta_0}{1+d_{j,k}^2}$
3	$\frac{\beta_0}{1+(d_{j,k}/\beta_1)^2}$
4	$\frac{\beta_0}{1+(d_{j,k}/\beta_1)^{\beta_2}}$

Table 2.2: Proposed parametric forms for the infection rate, $\beta(d_{j,k})$ for the example in section 2.3.2.

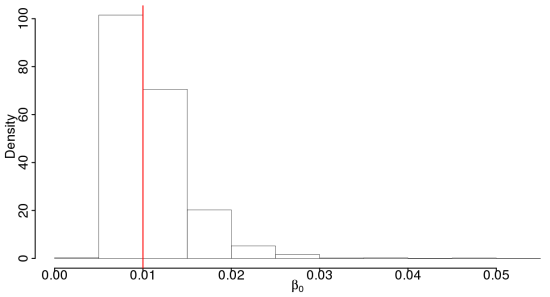
Euclidean distance. One advantage of choosing these parametric forms is that they are nested, enabling us to use likelihood-ratio tests for model comparison.

For models one and two, we use the exact posterior distributions and compute the median and 95% credible intervals. For models three and four, we run the MCMC algorithm for each model for 100,000 iterations, and remove the first 1,000 iterations as a burn-in period and thin the results by keeping every 20^{th} sample. The posterior median values and 95% credible intervals are shown in table 2.3. Figure 2.2 shows the trace plots and histograms for the parameters in model four. We can see that the Markov chain for β_2 mixes very slowly and that the 95% credible interval is very large. As the infection times are known, the trace plot for γ shows the chain for this parameter mixes well. Figure 2.3 compares the median infection rate to the true rate. We see large uncertainty around the estimate which is due to the population size being small, $N = 200$. In situations where we do not have much data about the outbreak, even with strong parametric assumptions and observing the times individuals were infected, it can be challenging to make precise estimates about the infection rate function.

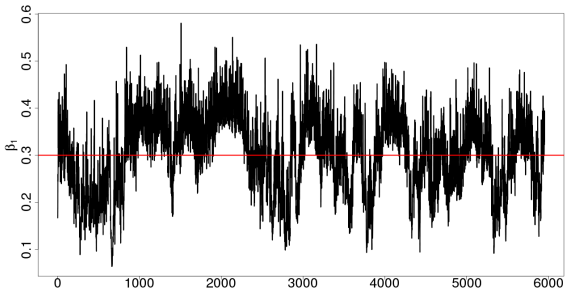
We now use the values in table 2.3 to compute the value of the corresponding likelihood functions and select the best model. There are a wide variety of model assessment methods, each with their different advantages and disadvantages. We now outline two methods: Bayes' Information Criterion and the Likelihood ratio test. We choose these by way of an example model assessment methods, and not because they



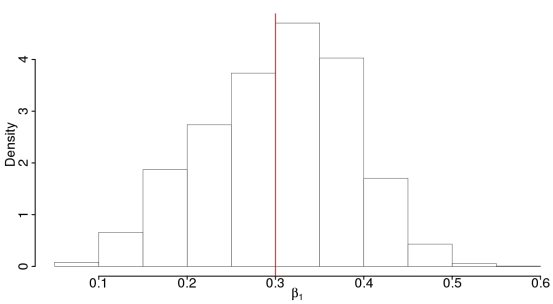
(a) Trace plot for β_0



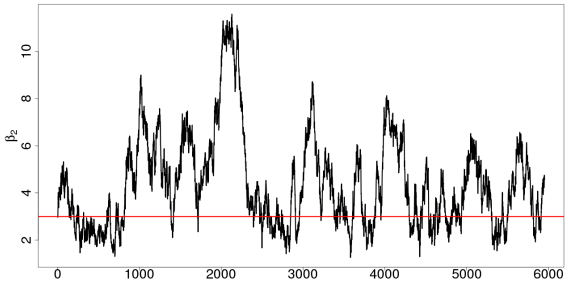
(b) Histogram for β_0



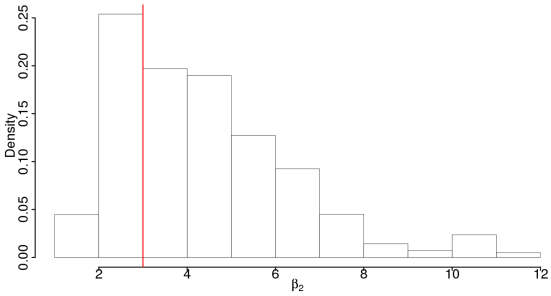
(c) Trace plot for β_1



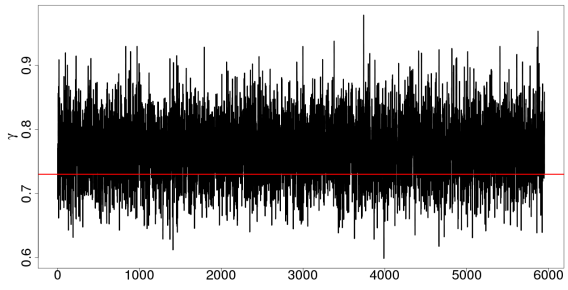
(d) Histogram for β_1



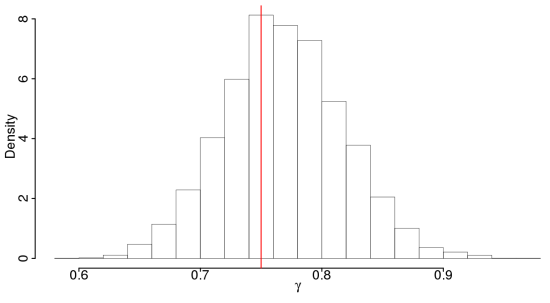
(e) Trace plot for β_2



(f) Histogram for β_2



(g) Trace plot for γ



(h) Histogram for γ

Figure 2.2: Trace plots and histograms for the model parameters in model 4 in the logistic infection rate example. The red lines give the true values of each parameter.

Model	γ	β_0	β_1	β_2
1	0.772 (0.676, 0.874)	0.00287 (0.00230, 0.00337)	-	-
2	0.772 (0.676, 0.874)	0.00363 (0.003, 0.00432)	-	-
3	0.769 (0.671, 0.872)	0.00356 (0.00790, 0.0311)	0.171 (0.102, 0.309)	-
4	0.7686 (0.00610, 0.0215)	0.00998 (0.00634, 0.0193)	0.310 (0.141, 0.449)	4.02 (1.877, 10.3)

Table 2.3: The results for the logistic infection rate function example. These are the posterior median value and 95% credible intervals for the values in each proposed function in table 2.2.

are optimal. We define Bayes' Information Criterion (BIC) (Schwarz, 1978) by:

$$\text{BIC} = k\log(N) - 2\log\pi(\mathbf{i},\mathbf{r}|\boldsymbol{\beta},\lambda,\gamma,\kappa,i_\kappa) \tag{2.4}$$

where k is the number of parameters in the model and N is the number of individuals in the population. The model with the smallest BIC is model 4, the true model. However, as shown in table 2.4, there are only minor differences between the BIC values for models two, three and four. Another model choice method is the likelihood-ratio test. Given a model with parameter vector $\boldsymbol{\beta}_1$, which is nested in the model with parameter vector $\boldsymbol{\beta}_2$, the likelihood-ratio test statistic between them is given by:

$$\Lambda = -2\left(\log\pi(\mathbf{i},\mathbf{r}|\boldsymbol{\beta}_1,\lambda,\gamma,\kappa,i_\kappa) - \log\pi(\mathbf{i},\mathbf{r}|\boldsymbol{\beta}_2,\lambda,\gamma,\kappa,i_\kappa)\right).$$

We can then test how significant this difference is as $\Lambda \sim \chi^2_{k_2-k_1}$, where k_1 and k_2 are the number of parameters in models one and two respectively. Models one and two are nested in model three and model three is nested in model four. This method shows that model four is significantly better at modelling the infection rate than model three. The likelihood-ratio statistics and corresponding p values for models two and three and three and four are given in table 2.4. Both BIC and the likelihood-ratio

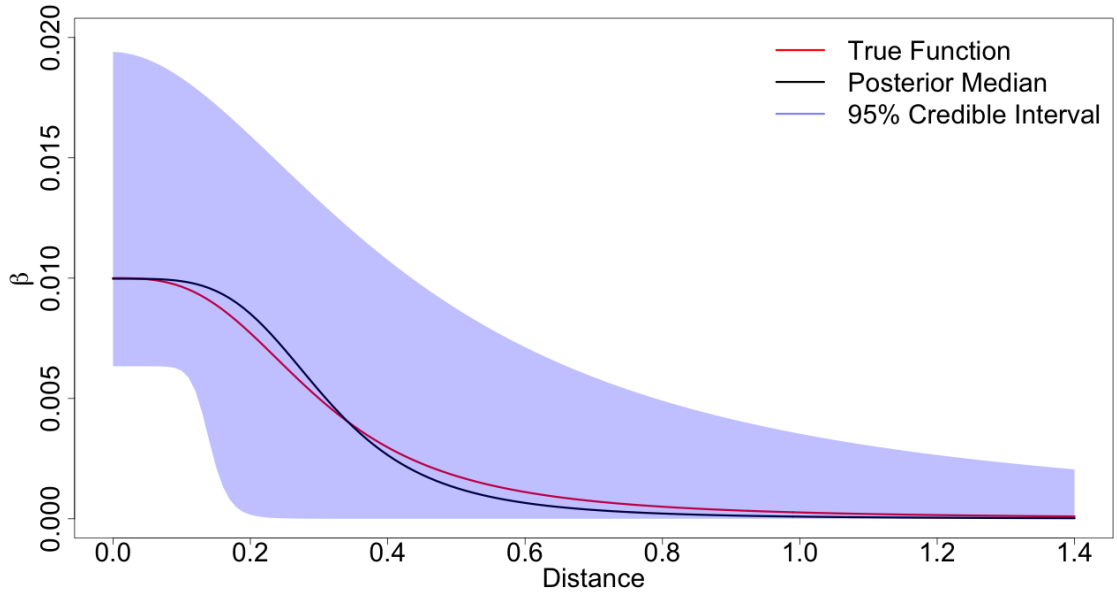


Figure 2.3: The true infection rate function compared to the infection rate functions constructed using the posterior median parameter values for model 4.

test correctly choose model four as the most suitable function with which to model the infection rate.

2.3.4 Difficulties and Constraints

Although proposing infection rate functions is straightforward, the choice is often arbitrary and has little justification from the data. One way of mitigating this is to propose a variety of functional forms and use a measure of goodness of fit to choose the best fitting function. Choosing a particular function is a restrictive assumption to make about the infection rate. We avoid this assumption by modelling the infection rate with GPs. We now develop our nonparametric modelling approach.

Model	Log-likelihood	BIC	Likelihood-Ratio	p value
1	-11821	23650	-	-
2	-667.9	1341	-	-
3	-654.8	1320	26.144	< 0.0001
4	-647.7	1312	14.06	0.0002

Table 2.4: The values of the log-likelihood and BIC for each of the models. The likelihood-ratio value in row m is the likelihood-ratio test value of model m compared to model $m-1$.

2.4 Gaussian Processes for Stochastic Epidemic Models

Instead of specifying the exact parametric form for the infection rate function β, we model this using a Bayesian nonparametric method and place a GP prior distribution on it. As the codomain of infection rate function and the GP prior distribution differ, we put a GP prior on a dummy function f and then transform this into the infection rate using a bijective function, g, with a positive codomain. The infection rate function is therefore given by:

$$\beta = g(f),$$

and the explicit choice of g is arbitrary. There are many choices for g, including the soft-plus function, $g(x) = \log(1 + \exp(x))$, for which $g(x) \approx x$ for large enough values of x. For the computation in this thesis, we used $g = \exp$ as this requires a single transformation to switch between f and β in the MCMC algorithm. A suitably chosen function g will not impact the estimate for β as the GP will adapt to the choice of g.

Given the set of all covariate values, $\mathbf{x} = \{x_{j,k}\}$, we put a GP prior on the infection rate via the function g as follows:

$$\beta = g(f), \quad f \sim \mathcal{GP}(0,\, \Sigma)$$

The covariance matrix, Σ, contains the information about how the infection rate will change over the input space. This information is partly gained from the data, but also the choice of covariance function and hyperparameters. It is important to choose a covariance function that models functions with suitable characteristics, such as smoothness and periodicity. We also need to set the hyperparameters to suitable values, which can either be done by hand or by learning them from the data. There are many choices of covariance function, and the resulting estimates will depend on this choice, so it is important to consider what type of function we are modelling. We have outlined several covariance functions in section 1.5.2. If we are modelling an outbreak of a disease where there is a spatial component, we can disregard the periodic covariance function. The squared exponential, Matérn and rational quadratic covariance functions are more suited to modelling spatial functions.

Making a choice about the covariance function is challenging given the little data we observe. For example, there is little to no justification from the data about the differentiability class of the infection rate function or how quickly moving the function is. Having previously declared lack of justifiability as a problem in the parametric framework, we also encounter it in the Bayesian nonparametric framework. However, in the parametric framework, we make strict assumptions about the infection rate function which need to be justified from a biological standpoint, however in the Bayesian nonparametric framework, we need to make mathematical assumptions about the function, for example, its differentiability class. As modellers and statisticians, we are more able to make such assumptions. We can also choose to include weak but plausible assumptions in other parts of the model, for example in section 2.7.1, we include assumptions about monotonic infection rate functions in the likelihood function.

In this thesis, we choose the squared exponential covariance function, as this models a wide range of smooth functions. We also choose this function as we wish to learn plausible values for the GP hyperparameters. In doing so, we are better able to justify our choice of covariance function. The squared exponential function has two hyperparameters, α and l, which control the signal variance and length

scale respectively. In the next section, we outline how we can learn the length scale parameter l. Both the rational quadratic and Matérn covariance functions include an extra hyperparameter. In the rational quadratic function, we can specify the relative weighting of short to long length scale variation and in the Matérn covariance function we can specify how the differentiability class. Given we typically only observe removal times and these will no be informative about possible values of these parameters, we choose not to use these covariance functions. From this point of the thesis, when we refer to any covariance function k, we are referring to the squared exponential covariance function unless another is specified.The full prior distribution is given by:

$$\beta = g(f), \quad f \sim \mathcal{GP}(0,\ \Sigma), \quad \Sigma_{jk} = k(x_j, x_k; \alpha, l)$$

One drawback with the GP prior distribution is that we require substantially more data to infer the infection rate than with the parametric model. This is because we are inferring the infection rate at each point in the input space. For example, suppose we observe an outbreak of a disease among a population of N individuals where the j^{th} individual has coordinates (x_j, y_j) and $x_j, y_j \sim \mathrm{U}[0, 1]$. We suppose the infection rate is distance-dependent and model the infection rate function nonparametrically over the space of all the pair-wise distances. We will observe many medium-sized pair-wise distances, but very few long range distances or very small distances. This means the Bayesian nonparametric method may struggle to learn the infection rate for the very small or very large distances and revert to the prior mean.

2.4.1 Learning the Gaussian Process Hyperparameter

One difficulty of the GP is the need to choose sensible values for the hyperparameters. The variance parameter α controls how much the function can vary. For our inference method, we can fix this parameter to a value so that the prior distribution generates functions which have a plausible codomain. The length scale is more challenging to set and small changes in the length scale can result in substantially different posterior distribution. It may be possible to learn plausible values of the length scale from the model's application. For example when looking at spatial covariates for an outbreak

of Avian Influenza, we expect the infection rate to differ when farms are several kilometres, not metres, apart. Although this may give an approximate range of values, we require more precision and we aim to learn the length scale from the data.

We follow Murray and Adams (2010) and Titsias et al. (2011) derive the following density function of l, f and the data $\mathbf{i}$ and $\mathbf{r}$:

$$\begin{aligned}\pi(f, \mathbf{i}, \mathbf{r}, l) &= \pi(l)\pi(f|l)\pi(\mathbf{i}, \mathbf{r}|g(f), l) \\ &= \pi(l)\mathcal{GP}(f; 0, \Sigma)\pi(\mathbf{i}, \mathbf{r}|g(f), \lambda, \gamma),\end{aligned}$$

From the joint distribution, we derive the following posterior conditional distributions:

$$\pi(l|f, \mathbf{i}, \mathbf{r}) \propto \pi(l)\mathcal{GP}(f; 0, \Sigma) \tag{2.5}$$

$$\pi(f|l, \mathbf{i}, \mathbf{r}) \propto \mathcal{GP}(f; 0, \Sigma)\pi(\mathbf{i}, \mathbf{r}|g(f), \lambda, \gamma). \tag{2.6}$$

We give details of how we sample from these distributions in an MCMC framework in section 2.5.

As written above, we choose to fix the signal variance parameter α. It is possible to learn this parameter, and placing a conjugate inverse-Γ prior distribution on this parameter results in an accessible, closed form posterior conditional distribution. However, for the sake of the mixing of the Markov Chain, it is convenient to fix this parameter. The value of this can be chosen through trial-runs and inspecting the prior distribution. We need to ensure α is large enough such that the samples from the prior distribution cover a large enough range and can mimic the true infection rate function. However, choosing a value that is too large results in a very poorly mixing Markov chain. We can compare this to the variance parameter in a one-dimensional normal prior distribution. Choosing a small value for the variance parameter results in a constrictive prior distribution, however choosing a value that is very large results in a prior distribution that is very vague places too much weight outside of the plausible range of values of the parameter. In both this one-dimensional example and our GP prior distribution, we need to choose a variance parameter that is suitably vague. In the following section we discuss underrelaxed MCMC, where we can in effect control the step size of the proposal distribution, which can mitigate the effects of a very vague prior distribution.

2.5 MCMC Implementation

In an outbreak, we often only observe the times individuals were removed, and the infection times are unobserved. As such, we treat the infection times as missing data and infer these values. By Bayes' theorem, the posterior distribution given the removal times $\mathbf{r}$ is:

$$\pi(\beta, l, \gamma, \mathbf{i}, \kappa, i_\kappa | \mathbf{r}, \lambda) \propto \pi(\mathbf{i}, \mathbf{r} | \beta, \lambda, \gamma, \kappa, i_\kappa)\pi(\beta|l)\pi(l)\pi(\gamma)\pi(\kappa)\pi(i_\kappa|\kappa). \tag{2.7}$$

To estimate the infection rate function β, the infectious period distribution rate parameter γ, the label and time of the initially infected individual, κ and i_κ, as well as the infection times of the other infected individuals, we develop an MCMC algorithm. The data we require is the removal times $\mathbf{r} = \{r_1, \dots r_n\}$, centred such that $r_1 = 0$. We follow Kypraios (2007) and Jewell et al. (2009) and fix the infectious period distribution scale parameter λ. We place the following prior distributions on the model parameters:

$$\begin{aligned}
&f \sim \mathcal{GP}(0,\, \Sigma), \qquad \Sigma_{jk} = k(x_j, x_k; \alpha, l), \qquad \beta = g(f) \\
&l \sim \text{Exp}(\chi_l) \\
&\gamma \sim \text{Exp}(\chi_\gamma) \\
&\kappa \sim \text{U}[1, \dots, n] \\
&i_\kappa \sim -z, \qquad z \sim \text{Exp}(\chi_\kappa)
\end{aligned}$$

The prior distribution for f is a GP prior distribution evaluated at the set of all pairs covariates, for example all pairwise distances. We place an exponential prior distribution on the length scale parameter l. We use a vague exponential prior distribution by setting the rate parameter to be small. We place a conjugate prior distribution on the infectious period distribution rate parameter γ, again, using a small parameter for the prior distribution so that it is vague. As we defined $\mathbf{i}$ to be the set of infection times excluding the initial infective, we must place a prior distribution on the initial infective and their infection time. For κ, the label of the initial infective, we place a discrete uniform prior distribution over the labels of

all infected individuals, and for the corresponding infection time i_κ, we place an exponential prior distribution on a dummy variable z and set $i_\kappa = -z$. This restricts i_κ to be less than 0, so that the first infection occurs before the first removal. Using the augmented likelihood function in equation (2.1), we derive the following posterior density using Bayes' theorem:

$$\begin{aligned}\pi(\beta, l, \gamma, \mathbf{i}, \kappa, i_\kappa|\mathbf{r}, \lambda) &\propto \pi(\mathbf{i}, \mathbf{r}|\beta, \lambda, \gamma, \kappa, i_\kappa)\pi(\beta|l)\pi(l)\pi(\gamma)\pi(\kappa)\pi(i_\kappa|\kappa) \\ &\propto \exp\left\{-\sum_{j=1}^{n}\sum_{k=1}^{N} g(f(x_{j,k}))\left((r_j \wedge i_k) - (i_j \wedge i_k)\right)\right\} \\ &\times \prod_{\substack{j=1 \\ j\neq\kappa}}^{n}\left(\sum_{k\in\mathcal{Y}_j} g(f(x_{k,j}))\right)\prod_{j=1}^{n} h(r_j - i_j|\lambda, \gamma)\mathcal{GP}(f; 0, \Sigma) \\ &\times \chi_l \exp\left\{-l\chi_l\right\}\chi_\gamma \exp\left\{-\gamma\chi_\gamma\right\}\chi_\kappa \exp\left\{i_\kappa\chi_\kappa\right\}.\end{aligned}$$

In the posterior distribution, we have replaced the infection rate function β with its inferred form $g(f)$. The full MCMC algorithm is shown in algorithm 5 and we now

Algorithm 5 Structure of the MCMC algorithm

1: Initialise the chain with estimates $\gamma^{(0)}$, $f^{(0)}$, $l^{(0)}$, and $\mathbf{i}^{(0)}$

Repeat the following steps

2: Sample γ from the conditional distribution $\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r}, \chi_\gamma)$ using a Gibbs step

3: Sample f using an underrelaxed proposal mechanisms for a Metropolis Hastings step

4: Sample l using a Metropolis Hastings step

5: Update an infection time

outline each of the steps in the algorithm.

2.5.1 Sampling the Infectious Period Rate Parameter

The full conditional posterior distribution for γ is given by:

$$\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r}, \chi_\gamma) \propto \chi_\gamma \exp\{-\gamma\chi_\gamma\}\prod_{j=1}^{n} h(r_j - i_j|\lambda, \gamma).$$

As we have used a conjugate prior distribution, the full conditional posterior distribution is given by:

$$\gamma|\mathbf{i},\mathbf{r},\lambda,\chi_\gamma \sim \Gamma\left(1+n\lambda,\ \chi_\gamma+\sum_{j=1}^{n}(r_j-i_j)\right).$$

2.5.2 Sampling the Infection Rate

According to equation (2.7), the full conditional posterior distribution for β is

$$\pi(f|\lambda,\gamma,\mathbf{i},\mathbf{r},\mathbf{x}) \propto \mathcal{GP}(f;0,\Sigma)\exp\Big\{-\sum_{j=1}^{n}\sum_{k=1}^{N} g(f(x_{j,k}))\big((r_j\wedge i_k)-(i_j\wedge i_k)\big)\Big\}$$
$$\times\prod_{j=1}^{n}\Big(\sum_{k\in\mathcal{Y}_j} g(f(x_{k,j}))\Big).$$

There are various methods to sample from this distribution. Hamiltonian Monte Carlo is one of the more efficient methods, but is more analytically complex than other methods (Heinonen et al., 2016). Splitting the function into blocks and conditionally updating one block given the values of the function in the other blocks reduces computational complexity, as this reduces the size of the covariance matrix, but increases the computation time, as we need to propose function values for each block. We instead sample from this distribution using the underrelaxed MCMC method. This method was first proposed by Neal (1995), and applied to GPs by Adams et al. (2009). We use it as it allows us to update the function as one block and as it is intuitive. To propose a new function, β', we use the form

$$\beta' = g(f'), \quad f' = \sqrt{1-\delta^2}f+\delta\nu, \tag{2.8}$$

where ν is drawn from the GP prior distribution, and $\delta\in(0,1]$. The parameter δ is a tuning parameter, similar to the variance parameter in a Metropolis Hastings Random Walk algorithm. For values of δ close to zero, proposals are more similar to the current function values, and for values of δ close to one, samples are much closer to samples from the prior distribution. Once we have proposed a new value of f, we accept or reject the proposal by computing the acceptance probability:

$$p_{acc} = \frac{\pi(\mathbf{i},\mathbf{r}|g(f'),\lambda,\gamma)}{\pi(\mathbf{i},\mathbf{r}|g(f),\lambda,\gamma)}\frac{\mathcal{GP}(f';0,\Sigma)}{\mathcal{GP}(f;0,\Sigma)}\frac{q(f|f')}{q(f'|f)}\wedge 1,$$

where $q(f'|f)$ is the probability of proposing f' given the current state f. One of the advantages of the underrelaxed proposal mechanism is that the proposal ratio can be simplified. Rearranging equation (2.8), this probability is given by the following multivariate normal density:

$$\begin{aligned} q(f'|f) &= \mathcal{GP}\left(\frac{1}{\delta}\left(\sqrt{1-\delta^2}f - f'\right); 0, \Sigma\right) \\ &\propto \exp\left\{-\frac{1}{\delta^2}\left(\sqrt{1-\delta^2}f - f'\right)^t \Sigma^{-1}\left(\sqrt{1-\delta^2}f - f'\right)\right\} \\ &\propto \exp\left\{-\frac{1}{\delta^2}\left[(1-\delta^2)f^t\Sigma^{-1}f - \sqrt{1-\delta^2}(f^t\Sigma^{-1}f') - \sqrt{1-\delta^2}(f'^t\Sigma^{-1}f) + f^t\Sigma^{-1}f'\right]\right\}. \end{aligned}$$

The proposal ratio is therefore given by:

$$\begin{aligned} \frac{q(f|f')}{q(f'|f)} &= \frac{\exp\left\{-\frac{1}{\delta^2}\left[(1-\delta^2)f'^t\Sigma^{-1}f' - \sqrt{1-\delta^2}(f'^t\Sigma^{-1}f) - \sqrt{1-\delta^2}(f^t\Sigma^{-1}f') + f'^t\Sigma^{-1}f\right]\right\}}{\exp\left\{-\frac{1}{\delta^2}\left[(1-\delta^2)f^t\Sigma^{-1}f - \sqrt{1-\delta^2}(f^t\Sigma^{-1}f') - \sqrt{1-\delta^2}(f'^t\Sigma^{-1}f) + f^t\Sigma^{-1}f'\right]\right\}} \\ &= \frac{\exp\left\{-\frac{1}{\delta^2}\left[(1-\delta^2)f'^t\Sigma^{-1}f' + f'^t\Sigma^{-1}f\right]\right\}}{\exp\left\{-\frac{1}{\delta^2}\left[(1-\delta^2)f^t\Sigma^{-1}f + f^t\Sigma^{-1}f'\right]\right\}} \\ &= \frac{\exp\left\{-f^t\Sigma^{-1}f\right\}}{\exp\left\{-f'^t\Sigma^{-1}f'\right\}} \\ &= \frac{\mathcal{GP}(f; 0, \Sigma)}{\mathcal{GP}(f'; 0, \Sigma)}. \end{aligned}$$

Hence the proposal ratio is the inverse of the prior density ratio. This reduces the acceptance probability to

$$\begin{aligned} p_{acc} &= \frac{\pi(\mathbf{i}, \mathbf{r}|g(f'), \lambda, \gamma)}{\pi(\mathbf{i}, \mathbf{r}|g(f), \lambda, \gamma)} \frac{\mathcal{GP}(f; 0, \Sigma)\mathcal{GP}(f'; 0, \Sigma)}{\mathcal{GP}(f'; 0, \Sigma)\mathcal{GP}(f; 0, \Sigma)} \wedge 1 \\ &= \frac{\pi(\mathbf{i}, \mathbf{r}|g(f'), \lambda, \gamma)}{\pi(\mathbf{i}, \mathbf{r}|g(f), \lambda, \gamma)} \wedge 1. \end{aligned}$$

This acceptance probability is convenient in an MCMC algorithm as we need to compute neither the prior distribution ratio nor the proposal distribution ratio.

2.5.3 Sampling the Length Scale

The conditional posterior distribution for l is given in equation (2.5). We use a random walk Metropolis algorithm to target this distribution, proposing new values of l by $l' = l + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$. We propose to update the covariance matrix

by $\Sigma'_{jk} = k(x_j, x_k;\, \alpha,\, l')$. The acceptance probability is given by:

$$p_{acc} = \frac{\mathcal{GP}(f; 0, \Sigma')}{\mathcal{GP}(f; 0, \Sigma)} \wedge 1.$$

We choose to fix σ based on trial runs, but it is possible to update this and use and adaptive MCMC algorithm.

2.5.4 Sampling Infection Times

As the infection times are unobserved, we treat them as parameters and infer these in the MCMC algorithm. We need to infer these as they are required to evaluate the likelihood function, and they also contain useful information in their own right when examining how an outbreak spread.

We propose a new label for κ uniformly at random and then propose a value of i_κ by $i'_\kappa = r_\kappa - t$, where $t \sim \Gamma(\lambda, \gamma)$. We write $\mathbf{i} + i_k$ to denote the set of infection times with i_k included and $\mathbf{i} - i_k$ to be the set with i_k removed. We accept this proposal with probability

$$p_{acc} = \frac{\exp\{-\chi_\kappa i'_\kappa\}}{\exp\{-\chi_\kappa i_\kappa\}} \frac{h(r_\kappa - i_\kappa|\lambda, \gamma)}{h(r_\kappa - i'_\kappa|\lambda, \gamma)} \frac{\pi(\mathbf{i} - i_\kappa + i'_\kappa, \mathbf{r}|\beta, \lambda, \gamma)}{\pi(\mathbf{i}, \mathbf{r}|\beta, \lambda, \gamma)} \wedge 1.$$

We infer the remaining infection times using a Metropolis-Hastings algorithm developed by O'Neill and Roberts (1999). We use the infectious period distribution to infer these times, as for any infected individual j

$$r_j - i_j = t_j, \quad t_j \sim \Gamma(\lambda, \gamma). \tag{2.9}$$

We uniformly at randomly choose an infected individual j and propose a new infection time by $i'_j = r_j - t_j$, where $t_j \sim \Gamma(\lambda, \gamma)$. We accept the proposal with probability

$$p_{acc} = \frac{h(r_j - i_j|\lambda, \gamma)}{h(r_j - i'_j|\lambda, \gamma)} \frac{\pi(\mathbf{i} - i_j + i'_j, \mathbf{r}|\beta, \lambda, \gamma)}{\pi(\mathbf{i}, \mathbf{r}|\beta, \lambda, \gamma)} \wedge 1.$$

2.6 Inference for Large Populations

Large populations pose two difficulties for our inference method. The first is the GP prior distributions and its covariance matrix. Large populations can require large

covariance matrices that are slow to decompose and invert, or require large memory allocations. The second difficulty is the speed of the MCMC algorithm, in particular updating the infection times. In this section we address both problems.

2.6.1 Methods for approximating GPs

Computational problems due to the dimension of the covariance matrix are well acknowledged in the GP literature, see for example Csato and Opper (2002); Hensman et al. (2013); Quinonero-Candela and Rasmussen (2005). Given N data points, a corresponding $N \times N$ covariance matrix is generated, which can be inverted and decomposed in time $O(N^3)$ and requires memory $O(N^2)$. For a spatially dependent infection rate, the dimension of the covariance matrix quickly becomes large; for example in a population with 500 individuals; there are almost 150,000 pair-wise distances. Although there are efficient ways to store, decompose and invert large matrices, we can approach this problem from a statistical standpoint. We now outline several methods for fitting a GP to a large data set.

Bui et al. (2017) divide the approximation methods into two sets: those using exact inference for approximate models, and those using approximate inference for exact models. The first set are methods that create an approximate model, similar to the original model and then perform inference on the approximate model, with the assumption that the inference for the approximate model gives similar results to the inference of the full model. Methods in this set include the *Fully Independent Training Conditional* approximation (Quinonero-Candela and Rasmussen, 2005) and the *Projected Process* approximation (Csato and Opper, 2002). The second set of methods take the opposite approach and use the exact model, but with relaxed assumptions on the inference algorithms. *Variational Free Energy* (Titsias, 2009) is one of the most commonly used methods of this kind.

2.6.1.1 The Mean Projection Approximation

The method we will use is a projection method used to construct many approximations. It creates a similar model on a small set of pseudo data and performs exact inference

for this new model. The *Mean Projection Approximation* (MPA) forms the basis of many other projection approximations such as the Fully Independent Training Conditional (Csato and Opper, 2002) approximation, the Partially Independent Training Conditional approximation (Quinonero-Candela and Rasmussen, 2005), the Subset of Regressors approximation (Rasmussen and Williams, 2006) and the SOLVE-GP method (Shi et al., 2019). These methods are primarily designed for regression and prediction problems and their differences are due to the formulation of the covariance matrix and posterior predictive distribution. However, they all contain the same mean projection method, which we now outline.

The MPA places a GP prior distribution on a function over a pseudo data set, then projects it onto the full data set, as such we can use the likelihood function in our inference algorithms. Suppose we have a data set $\mathbf{x}$ of size n, where n is prohibitively large. To perform inference, we create a pseudo data set of size $m < n$, and denote it by $\bar{\mathbf{x}}$. The chosen input points must cover a sufficient proportion of the domain and any areas we expect the inferred function to have interesting properties. The pseudo data set should be similar to the original data set, but smaller in size. One method is to construct a set of uniformly placed points over the original input space. If we expect the function to have particular qualities, such as turning points, then we need to ensure the pseudo set has sufficient input points near the locations of these points. As the purpose of this is to infer the function, deciding if these features exist and their locations can be difficult. We further examine the impact of the size of the pseudo set on computational time and error in section 2.8.2.

We denote the original function of interest by f and the function over the pseudo data set by $\bar{f}$. We place a joint GP prior distribution on f and $\bar{f}$ such that

$$\begin{pmatrix} f \\ \bar{f} \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{x,x} & \Sigma_{x,\bar{x}} \\ \Sigma_{\bar{x},x} & \Sigma_{\bar{x},\bar{x}} \end{pmatrix} \right), \tag{2.10}$$

where $\Sigma_{x,x}$ is the covariance matrix for the original data set, $\Sigma_{\bar{x},\bar{x}}$ is the covariance matrix for the pseudo data set, and $\Sigma_{x,\bar{x}}$ the covariance between the original and pseudo data set. The MPA method exploits the conditional distribution of the

multivariate normal distribution, as the distribution of the function f given $\bar{f}$ is:

$$f|\bar{f} \sim \mathcal{GP}\left(\Sigma_{x,\bar{x}}\Sigma_{\bar{x},\bar{x}}^{-1}\bar{f},\ \Sigma_{x,x} - \Sigma_{x,\bar{x}}\Sigma_{\bar{x},\bar{x}}^{-1}\Sigma_{\bar{x},x}\right).$$

This projects $\bar{f}$ onto the full data set $\mathbf{x}$, but only requires time $O(m^3)$ to compute $\Sigma_{\bar{x},\bar{x}}^{-1}$. For dense input spaces, the Eigenvalues of the covariance matrix may decay to 0 quickly making this matrix difficult to decompose or invert computationally. Due to this, and also as f is not a real observation, we compute it by:

$$f = \Sigma_{x,\bar{x}}\Sigma_{\bar{x},\bar{x}}^{-1}\bar{f}. \tag{2.11}$$

The MPA allows us to approximate f and greatly reduce the dimension of the covariance matrix, yet still evaluate the likelihood function and perform exact inference. We choose the MPA method as it requires fewer operations than other similar methods, which will be computationally advantageous in the MCMC algorithm. The second advantage to the MPA method is that it requires only minimal adjustments to the MCMC algorithm, instead of targeting f, we now target $\bar{f}$. This means in proposal distributions and acceptance probabilities we replace f with $\bar{f}$ and Σ by $\Sigma_{m,m}$. In the likelihood function we replace $g(f)$ with $g(\Sigma_{x,\bar{x}}\Sigma_{\bar{x},\bar{x}}^{-1}\bar{f})$.

2.6.2 The Conditional Model

In some modelling circumstances, we can assume *a priori* that the function we are modelling tends towards a horizontal asymptote as x tends to some value. We can then choose a critical point, such that for any input point larger than the critical point, the function can be modelled by the asymptote. For example, we may be interested in modelling an infection rate which depends on the distance between two individuals. It may be reasonable to assume that there is some critical distance, d^*, such that if any two individuals are further away from each other than d^* the infection rate between them can be assumed to be effectively 0 and one farm will almost never infect another farm more than d^*km away.

Suppose we are modelling a function f over an input region χ, and there exists some $c \in \mathbb{R}$ such that as $x \longrightarrow \infty$, $f(x) \longrightarrow c$. We approximate this by assuming

there exists some $x^* \in \chi$ such that for all $x > x^*$, $f(x) = c$. We partition χ into two sets χ_1, which contains the elements less than or equal to x^* and χ_2 for the elements strictly greater than x^*. We then construct two functions f_1 and f_2, which correspond to the functions over χ_1 and χ_2 respectively. We place a joint GP prior distribution on f_1 and f_2 such that:

$$\begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix}\right),$$

where $\Sigma_{i,j}$ is the covariance matrix specifying the covariance between sets $\mathbf{x}_i$ and $\mathbf{x}_j$. The conditional distribution of f_1 given $f_2 = c$ is

$$f_1|(f_2 = c) \sim \mathcal{GP}(\tilde{\mu}, \tilde{\Sigma}),$$

where $\tilde{\mu} = \Sigma_{1,2}\Sigma_{2,2}^{-1}c$ and $\tilde{\Sigma} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$. We can use the underrelaxed proposal mechanism to target f_1 in the MCMC algorithm, and proposing new values by:

$$f_1' = \sqrt{1-\delta^2} f_1 + \delta\nu, \quad \nu \sim \mathcal{GP}(\tilde{\mu}, \tilde{\Sigma}).$$

Proposed values of β are then given by:

$$\beta' = g(f'), \quad f' = \{f_1', c\}.$$

This will increase the speed of the MCMC algorithm, as when computing the likelihood function, many elements will remain the same when updating the function. This method can improve the mixing of the Markov chain as fewer elements of the GP need to be updated. The main disadvantage to this method is that we need to *a priori* identify both the limiting value, c, and the point where this value begins x^*. For example, when modelling an infection rate function which tends to 0 as x tends to infinity, we need to identify a value x^* for which the infection rate is close enough to 0 to be approximated by 0. On a computational level, we need $g^{-1}(0)$ to be a well-defined, real value. In the case where the log function is used as the transformation function, we have found that setting $c = -20$ works well.

The most suitable application of this method is when the domain of f is large, or the value of x^* is a small value in χ. This is because we will be modelling a

large number of points which will be constant. This can also create difficulties when inferring the length scale parameter, as f may fluctuate greatly over χ_1, but not at all over χ_2.

2.6.3 Methods for the MCMC Algorithm

To improve the speed of the MCMC algorithm, we are able to make use of parallel threading and rewrite several equations to make them more suitable for computational evaluation.

The main difficulty is updating the infection times, as they need to be updated sequentially and they are slow to converge. We can reduce the time required to run the algorithm by updating 30-50% of the infection times for each iteration of the MCMC algorithm, choosing infection times with a random scan and accepting and rejecting the proposals individually.

Another difficulty is repeatedly computing the likelihood function (2.1), in particular the double sum

$$\Psi = \sum_{j=1}^{n} \sum_{k=1}^{N} \exp\{-\beta(x_{j,k})((r_j \wedge i_k) - (i_j \wedge i_k))\}.$$

To reduce the requirements for this computation, we define the following matrices:

$$\mathbf{B} = \begin{pmatrix} 0 & \beta(x_{1,2}) & \dots & \beta(x_{1,N}) \\ \beta(x_{2,1}) & 0 & \dots & \beta(x_{2,N}) \\ \vdots & \vdots & \ddots & \vdots \\ \beta(x_{N,1}) & \beta(x_{N,2}) & \dots & 0 \end{pmatrix},$$

$$\mathbf{\Delta} = \begin{pmatrix} r_1 - i_1 & (r_1 \wedge i_2) - (i_1 \wedge i_2) & \dots & (r_1 \wedge i_N) - (i_1 \wedge i_N) \\ (r_2 \wedge i_1) - (i_2 \wedge i_1) & r_2 - i_2 & \dots & (r_2 \wedge i_N) - (i_2 \wedge i_N) \\ \vdots & \vdots & \ddots & \vdots \\ (r_N \wedge i_1) - (i_N \wedge i_1) & (r_N \wedge i_1) - (i_N \wedge i_1) & \dots & r_N - i_N \end{pmatrix}.$$

The double sum is now given by

$$\Psi = \sum_{j=1}^{N} \sum_{k=1}^{N} (\mathbf{B} \circ \mathbf{\Delta})_{j,k},$$

where $\circ$ is the component-wise multiplication operation. On the surface, we have increased the double sum from $O(nN)$ to $O(N^2)$, however this double sum is no longer ordered by infected then susceptible individuals, and the elements of $\mathbf{\Delta}$ corresponding to susceptible-susceptible interaction will be 0. The main advantage to this formulation is that when updating that infection time i_k to i'_k, we now only need to update the k^{th} row and column of $\mathbf{\Delta}$, instead of computing the whole matrix from scratch. Furthermore, given the current infection matrix $\mathbf{\Delta}$ and the proposed matrix $\mathbf{\Delta}'$, the proposed value of the double sum is

$$\Psi' = \Psi + \sum_{j=1}^{N} \left(\mathbf{B} \circ (\mathbf{\Delta}' - \mathbf{\Delta})\right)_{j,k} + \sum_{j=1}^{N} \left(\mathbf{B} \circ (\mathbf{\Delta}' - \mathbf{\Delta})\right)_{k,j}. \tag{2.12}$$

We can also make performance improvements by choosing a suitable programming language. We have written our algorithm in C using the GCC compiler, and make use of OpenMP to parallelise the double sum and product elements of the likelihood function. This works by transferring different iterations of the sum and product to different compute cores. For example, when computing equation (2.12) with 16 available cores, we can choose to compute the values of the two sums for $j = 1, \ldots, \frac{N}{16}$ on core one, for $j = \frac{N}{16} + 1, \ldots, \frac{2N}{16}$ on core two and so on. Once all iterations of the sum have been computed, we then add all the values from the 16 cores to compute the values of the two sums. We have found that dynamic allocation of the likelihood function elements, that is allowing the OpenMP processor to allocate which elements go to which cores, reduces the time for the MCMC algorithm by up to 30%, with larger time savings for larger population sizes.

2.7 Monotonicity and Gaussian Processes

The main advantage of Bayesian nonparametric methods is the lack of assumptions we need to make. However, if we can make general, accurate assumptions about the model, this will improve the accuracy of the results. One assumption we can often make about infection rate functions is that they are monotone functions. Consider an outbreak of a disease among livestock on farms. We may assume the more livestock

present on a farm, the more susceptible the farm is to contracting the disease, and wish to model the infection rate function as monotonically increasing. We now discuss a method for enforcing monotonicity in GPs in a regression setting proposed in Riihimäki and Vehtari (2010), and then apply this to our epidemic model.

Enforcing monotonicity is challenging, especially as we cannot restrict the prior distribution in such a way that we only draw samples which are monotonic. Instead of specifying that the function must be monotonic, In Riihimäki and Vehtari (2010), the authors developed a method where they specify the sign of the gradient at a limited number of points. They define the joint prior distribution of the function f and its derivative f', and then only consider functions whose derivative meets a pre-determined monotonicity constraint.

To implement this method, we first need to derive the GP gradient. Suppose the function of interest is given by f over an input set $\mathbf{x} = \{x_1, \ldots, x_n\}$. We begin by placing a GP prior distribution on f:

$$f \sim \mathcal{GP}(0,\, \Sigma), \quad \Sigma_{ij} = k(x_i, x_j).$$

The covariance between f_i and f_j is given by the covariance function evaluated at x_i and x_j:

$$\begin{aligned} \operatorname{cov}(f_i, f_j) &= k(x_i, x_j) \\ &= \alpha^2 \exp\left\{-\frac{(x_i - x_j)^2}{l^2}\right\}. \end{aligned} \tag{2.13}$$

As differentiation is a linear operator, the covariance between f_i and f_j', the value of the GP evaluated at x_i and the value of the derivative of the GP at x_j is

$$\begin{aligned} \operatorname{cov}(f_i, f_j') &= \frac{\partial k(x_i, x_j)}{\partial x_j} \\ &= -\alpha^2 \frac{2}{l^2}(x_i - x_j) \exp\left\{-\frac{(x_i - x_j)^2}{l^2}\right\} \\ &= -\frac{2}{l^2}(x_i - x_j) k(x_i, x_j). \end{aligned} \tag{2.14}$$

And the covariance between the values of the derivative, f_i' and f_j', is

$$\begin{aligned}\text{cov}(f_i', f_j') &= \frac{\partial^2 k(x_i, x_j)}{\partial x_i \partial x_j} \\ &= \frac{2\alpha^2}{l^2}\left[1 - \frac{2}{l^2}(x_i - x_j)^2\right]\exp\left\{-\frac{(x_i - x_j)^2}{l^2}\right\} \\ &= \frac{2}{l^2}\left[1 - \frac{2}{l^2}(x_i - x_j)^2\right]k(x_i, x_j). \end{aligned} \tag{2.15}$$

As we are using a zero mean prior distribution, we have

$$\mathbb{E}[f_i] = 0 \implies \mathbb{E}[f_i'] = 0.$$

Hence the joint distribution of the values of the function f and the values of the derivatives f' is

$$\begin{pmatrix} f \\ f' \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \frac{\partial}{\partial x}\Sigma \\ \frac{\partial}{\partial x}\Sigma & \frac{\partial^2}{\partial x \partial x}\Sigma \end{pmatrix}\right), \tag{2.16}$$

where Σ is the matrix given by the non-differentiated function in equation (2.13), $\dfrac{\partial}{\partial x}\Sigma$ specifies the covariance between the function and its first derivative and is computed using equation (2.14), and $\dfrac{\partial^2}{\partial x \partial x}\Sigma$ is the covariance matrix for the differentiated function and is constructed in equation (2.15).

Given the joint prior distribution, we now wish to enforce a monotonicity. The method proposed in Riihimäki and Vehtari (2010) requires us to specify the sign of f' at a limited number of points. We follow Riihimäki and Vehtari (2010) and introduce a monotonicity parameter $m = \pm 1$, where the value of m depends on whether we assume the gradient of f to be positive (+1) or negative (-1). In Riihimäki and Vehtari (2010), the authors suggest for one dimensional problems that derivative points are chosen uniformly on a grid. As we use a squared exponential covariance function, we must take the length scale into account when placing these points, as we must ensure the length scale and spacing of the derivative points are agreeable.

To enforce monotonicity, we include an extra term in the likelihood function to only consider contributions from functions where the sign of the gradient at the chosen points satisfy the monotonicity condition. We follow Riihimäki and Vehtari

(2010) and include a probit term in the likelihood function, which is given by:

$$\pi(m|f_i') = \Phi(mf_i'\frac{1}{\tau}), \quad \Phi(z) = \int_{-\infty}^{z} N(u;\, 0, 1)du,$$

where τ is a tolerance parameter. In general, this assigns 0 probability to derivative values of the opposite sign to m and a probability of 1 when the signs are the same. The tolerance parameter τ allows for small errors, that is gradients which are of the opposite sign to m but close to 0. The choice of τ is problem specific and can either be fixed or considered a hyperparameter, the value of which can be learned in the MCMC algorithm. Another, stricter option, would be to use an indicator function, which rejects any function which does not meet the constraint. The value of τ allows us to choose how strict the monotonicity constraint is, as when $\tau \longrightarrow 0$, the probit term tends to a step function and any samples which have a derivative with the wrong sign are rejected. In many cases, this is too strict and although we are enforcing the sign of the derivative a small error in the sign can be tolerated. Both the probit and indicator functions are shown in figure 2.4.

In practice, there is little difference seen in results when using different values of τ. We find that in areas where there is a lot of data, this condition and the value of τ makes no difference to the results as we can learn the function directly from the data. In areas where there is little data, we do see there is some change in the results, however we only see considerable change in results comparing values of τ which are an order of magnitude different. The effect seems to control how quickly the posterior reverts to the prior distribution in large areas of little data.

2.7.1 Monotonic GP Regression

We now demonstrate the monotonic GP method in a regression problem, where

$$y_i = f(x_i) + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2).$$

Given some observations $\mathbf{y}$ at locations $\mathbf{x}$, and the assuming the direction of the monotonicity, the posterior distribution for the function f and its derivative f' is

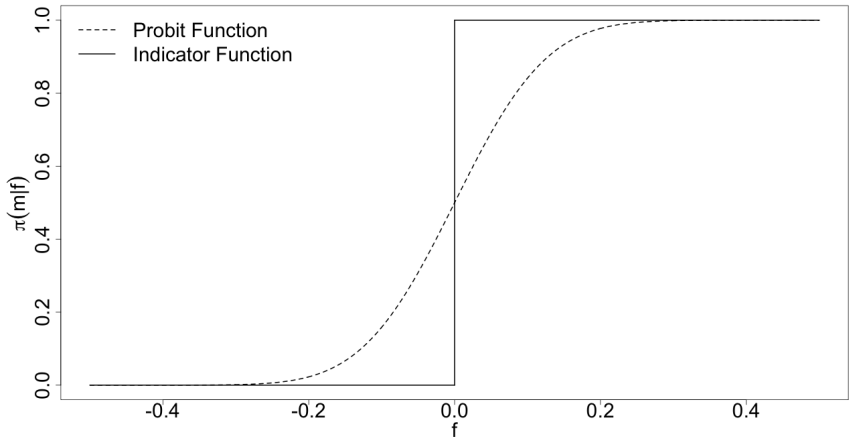


Figure 2.4: Two possible options for the monotonicity term in the likelihood function in section 2.7.1. The dashed line is the probit function with $\tau = 10^{-2}$ and the solid line is an indicator function.

given by:

$$\pi(f, f'|\mathbf{y}, m, \sigma^2) \propto \pi(f, f')\pi(\mathbf{y}|f, \sigma^2)\pi(m|f'),$$

To generate samples from the posterior distribution we use an MCMC algorithm, using the underrelaxed proposal mechanism. Given the current values of f and f', we propose values $f^{(prop)}$ and $f'^{(prop)}$ by:

$$\begin{pmatrix} f^{(prop)} \\ f'^{(prop)} \end{pmatrix} = \sqrt{1-\delta^2} \begin{pmatrix} f \\ f' \end{pmatrix} + \delta \begin{pmatrix} \nu \\ \nu' \end{pmatrix}, \quad \begin{pmatrix} \nu \\ \nu' \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \frac{\partial}{\partial x}\Sigma \\ \frac{\partial}{\partial x}\Sigma & \frac{\partial^2}{\partial x \partial x}\Sigma \end{pmatrix} \right).$$

Using the underrelaxed proposal mechanism, the acceptance probability is given by:

$$p_{acc} = \frac{\pi(f^{(prop)}|\mathbf{y})\pi(m|f'^{(prop)})}{\pi(f|\mathbf{y})\pi(m|f')}.$$

By way of an example, we generate ten input points $\mathbf{x} = \{x_1, \ldots, x_{10}\}$ uniformly at random on the interval [0, 10], and then generating 10 observations points from the model:

$$y_i = \exp\{-x_i\} - \Phi(x_i - 5) + \varepsilon_i, \quad \varepsilon \sim N(0,\, 0.5^2).$$

We wish to infer the value of the function at the input points $\mathbf{x}^* = \{0, 0.01, 0.02, \ldots, 10\}$. We enforce a negative gradient at the points $\mathbf{x}_m = \{0, 0.5, \ldots, 10\}$. We place the

following GP prior distribution on f and f':

$$\begin{pmatrix} f \\ f' \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \frac{\partial}{\partial x}\Sigma \\ \frac{\partial}{\partial x}\Sigma & \frac{\partial^2}{\partial x \partial x}\Sigma \end{pmatrix}\right).$$

and set $\alpha^2 = 5$ and $l^2 = 8$. The likelihood function for the regression part is given by:

$$\pi(\mathbf{y}|f, \mathbf{x}) = \frac{1}{\sqrt{(2\pi\sigma^2)^{10}}} \exp\left\{-\sum_{i=1}^{10} \frac{(y_i - f_i)^2}{2\sigma^2}\right\},$$

and the probit likelihood function is given by:

$$\pi(-1|f') = \prod_{i=1}^{20} \Phi\left(\frac{-1}{10^{-4}} f_i'\right).$$

From this, the likelihood contributions to functions which has positive gradient at the monotonicity points is 0. We choose the value of τ based on Riihimäki and Vehtari (2010). The full posterior is given by:

$$\pi(f, f'|\mathbf{x}, \mathbf{x}_m, \mathbf{y}, \alpha, l, \sigma^2) \propto \mathcal{GP}\left(\begin{pmatrix} f \\ f' \end{pmatrix}; 0, \begin{pmatrix} \Sigma & \frac{\partial}{\partial x}\Sigma \\ \frac{\partial}{\partial x}\Sigma & \frac{\partial^2}{\partial x \partial x}\Sigma \end{pmatrix}\right) \times N(\mathbf{y}; f, \sigma^2 I) \prod_{i=1}^{20} \Phi\left(\frac{-1}{10^{-4}} f_i'\right).$$

We then compare the standard GP regression model to the monotone GP regression model, the results are shown in figure 2.5. We can see the uncertainty around the estimate for the standard GP prior distribution is much larger than for the monotone GP, especially in the interval [0, 1.5]. This is because there are no observations in this interval, so we are almost entirely dependent on the prior distribution. With the monotonic GP prior distribution, we assume $f(0) \geq f(0.5) \geq \ldots \geq f(2)$, which forces the function to be decreasing, whereas in the standard method, this assumption is not made and the samples revert to the prior distribution. As we are including a strong assumption in the prior distribution, the monotone assumption reduces uncertainty in areas where data is not observed.

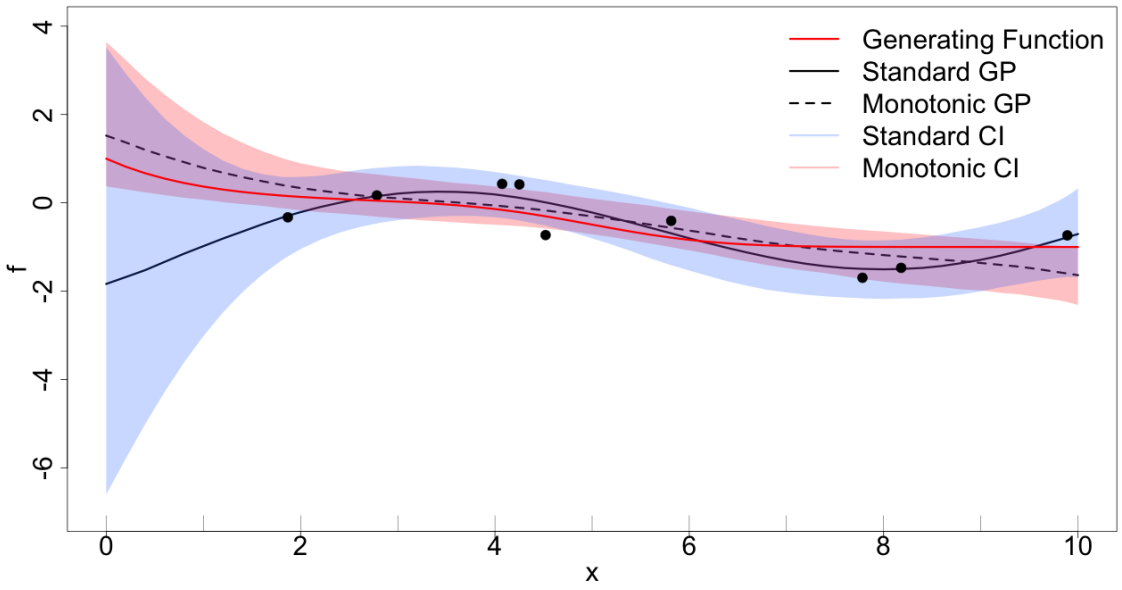


Figure 2.5: The standard GP regression model compared to the monotone GP regression model developed in section 2.7.1. The tick marks are the points at which monotonicity is enforced.

2.7.2 Monotone GPs for Epidemic Models

We now apply this method to our epidemic model and assume the infection rate function is an increasing function of the covariate, that is:

$$\beta' \geq 0.$$

The case when the function is monotonically decreasing is analogous with $m = -1$. We model the infection rate through the dummy function f, which is transformed into the infection rate through the function g. By the chain rule:

$$\beta = g(f) \implies \beta' = g'(f)f'.$$

As we choose the form of g, the sign of g' is known, and we can infer the required sign of f' for β' to be positive.

We then construct a set of q points, $\mathbf{x}_m$, at which we will enforce monotonicity by either forming a grid through $\mathbf{x}$ or choosing points from $\mathbf{x}$ uniformly at random. The

value of the infection rate at the points $\mathbf{x}$, is given by β, and the prior distribution for β is

$$\beta = g(f), \quad \begin{pmatrix} f \\ f' \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \frac{\partial}{\partial x}\Sigma \\ \frac{\partial}{\partial x}\Sigma & \frac{\partial^2}{\partial x \partial x}\Sigma \end{pmatrix} \right).$$

The full likelihood function, including the contributions from the probit model, is

$$\begin{aligned} \pi(\mathbf{i}, \mathbf{r} | f, f', \lambda, \gamma, \kappa, i_\kappa) = & \exp\left\{ -\sum_{j=1}^{n} \sum_{k=1}^{N} g(f(x_{j,k})) \left((r_j \wedge i_k) - (i_j \wedge i_k) \right) \right\} \\ & \times \prod_{\substack{j=1 \\ j \neq \kappa}}^{n} \left(\sum_{k \in \mathcal{Y}_j} g(f(x_{k,j})) \right) \prod_{j=1}^{n} h(r_j - i_j | \lambda, \gamma) \prod_{j=1}^{q} \Phi\left(f'_j \frac{1}{10^{-4}} \right). \end{aligned}$$

We then continue by deriving the posterior distribution in the same fashion as in section 2.5, and generating samples from this distribution using the MCMC algorithm outlined in algorithm 5. This method can be combined with the MPA method given in section 2.6, which reduces the dimension of Σ and therefore the prior distribution covariance matrix. We do not need to project f' onto the full data set as we are only concerned with the values of f.

2.8 Simulation Studies

In this section, we give the results of two simulation studies to show the effectiveness of the nonparametric method. In the first study, we demonstrate the nonparametric method on a small data set with 100 individuals, and compare the results to our approximation method. We then run a simulation study for outbreaks in a large population of 1000 individuals using the MPA method to infer the infection rate.

To simulate outbreaks of a disease, we use an infection rate that is distance dependent. The exact form of the rate is

$$\beta(d_{i,j}) = \beta_0 \exp\{-\beta_1 d_{i,j}\}, \tag{2.17}$$

where $d_{i,j}$ is the Euclidean distance between individuals i and j. This gives an infection rate close to β_0 when the distance is small, but tends to 0 as the distance

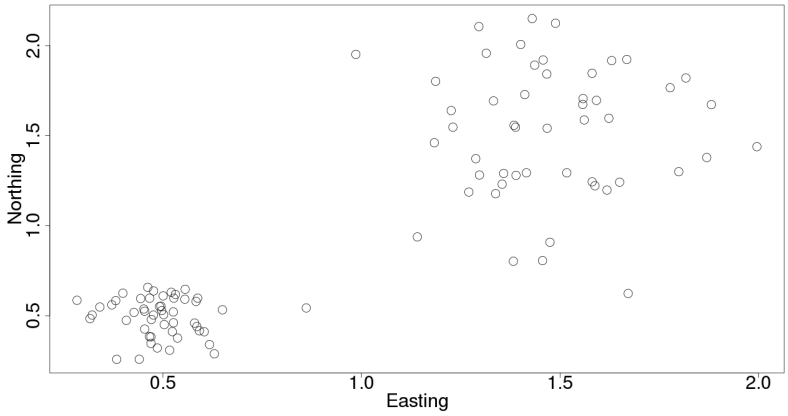


Figure 2.6: The positions used to simulate the outbreaks in the small simulation study with 100 individuals in section 2.8.1.

between the infected and susceptible individuals grows. This is a characteristic that we would expect to see in outbreaks where the distance is a key factor.

We choose to run a simulation study where we infer the model parameters for 250 independent outbreaks, using a different random seed for each outbreak. This is for two reasons. The first is so we cannot cherry pick results for a 'nice' outbreak. We present the results of all 250 outbreaks with the aim that the inference method performs well for all 250 outbreaks. The second reason is that it is difficult to study the mixing of the infection rate function. As we are evaluating the function at a large number of points we would need to investigate trace plots for all of the points of the function. This would be unfeasible, and although we can instead look at collections of points or the sum of the points of the function, these often disguise poor mixing in parts of the function. Instead we use a different initial condition for the function in the MCMC algorithm for each outbreak. If the posterior estimates for all outbreaks are similar then given that the starting points are different, we can deduce the Markov chains are mixing well.

2.8.1 Inference for a Small Population

We generate the positions of 100 individuals by

$$x_i \sim N(0.5, 0.1^2),\ y_i \sim N(0.5, 0.1^2), \quad i = 1, \dots 50,$$
$$x_i \sim N(1.5, 0.25^2),\ y_i \sim N(1.5, 0.3^2), \quad i = 51, \dots 100.$$

This creates two clusters with 50 individuals each, the positions of which are shown in figure 2.6. We choose a population with two clusters as we are want to analyse how the model is affected by intra- and intercluster infections. We are likely to see many infections between pairs of individuals in the same cluster, but fewer between pairs of individuals in different clusters because they are further apart. This will make it more difficult for the model to learn the infection rate function for larger distances. We simulate 250 outbreaks of a disease with infection rate shown in equation (2.17) in this population. We ensure that there are more than 20 infecteds in each outbreak, this is to ensure there is sufficient data from which the model can learn. The parameters we use in the model are shown in table 2.5. The value of N was chosen to be 100 so that we can compare the GP and MPA methods. This is because the data set generates a covariance matrix small enough to use the GP method, and large enough to the MPA method. The infection rate parameters give a function that is plausible to observe in a real outbreak. We chose the infectious period shape parameter to be two so that the distribution is not exponential. The value of the GP variance, α, was chosen so that the samples prior distribution vary over a large range and produce a wide variety of possible values for β_0. This generates 250 sets of infection and removal times.

For each set, we implement the MCMC algorithm to infer the infection rate, infection times and the rate parameter of the Gamma infectious period distribution. As sampling the length scale involves repeatedly inverting and decomposing the covariance matrix, we only use the method with the MPA and do not fit a GP using the full data set, and as such, fix the length scale for this study. We run the MCMC algorithm for 100,000 iterations and remove the first 5,000 as a burn-in period. We thin the results by keeping every 5^{th} sample.

Parameter	Value
N	100
β_0	0.2
β_1	2
λ	2
γ	3
α	10
l	8

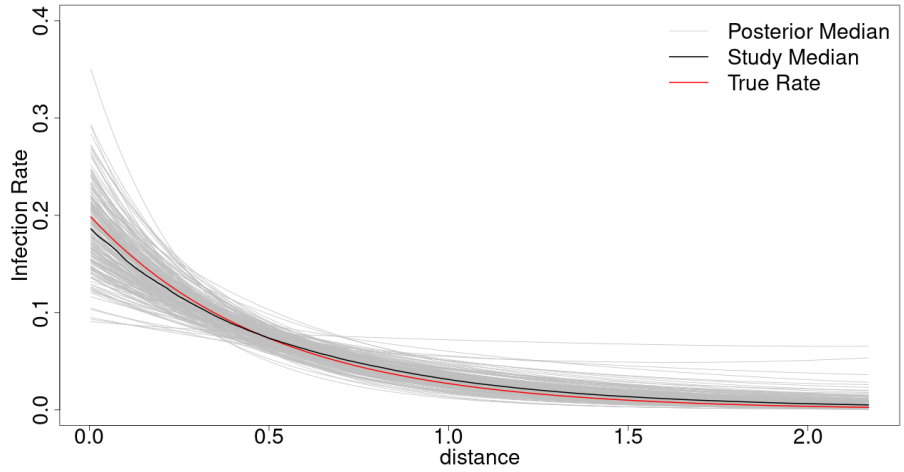
Table 2.5: Parameter values used to simulate outbreaks in the small simulation study

2.8.1.1 Observed Infection Times

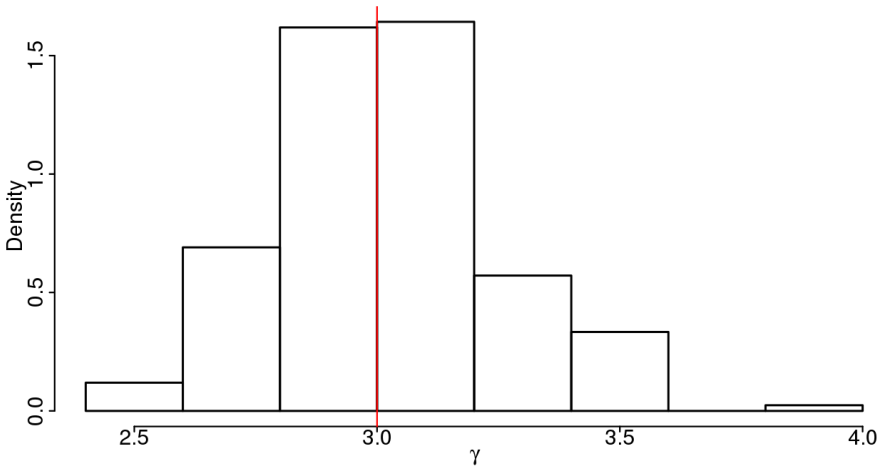
We first estimate the infection rate with known infection times because this should yield better estimates for the infection rate function. Figure 2.7(a) shows that we can estimate the infection rate function with minimal error. We estimate the shape and scale of the infection rate well using our new nonparametric method. By using this nonparametric method, we remove any identifiability issues. As the exact infection times are known, estimating γ is not challenging, and we estimate this parameter well as shown in figure 2.7(b).

2.8.1.2 Unobserved Infection Times

We now repeat the study except we assume the times individuals were infected were unobserved and infer these using data augmentation within the MCMC algorithm. Figure 2.8(a) shows the posterior median estimates for the true infection rate functions for each data set as well as the median of the posterior medians. This shows we can recover the true infection using our nonparametric method. As figure 2.8(c) shows, the infection times are estimated well, but there is a tendency to underestimate the infection times. We monitor the success of the infection time estimates by comparing the sum of the infection time estimates to the sum of the true infection times. This



(a) Estimates for the infection rate function



(b) Estimates for the infectious period distribution rate parameter

Figure 2.7: Results of the simulation study in section 2.8.1 for a small population of size 100 with known infection times. Figure (a) shows estimates for the infection rate, with each grey line representing the posterior median for one of the 250 data set. The black line is the median of all 250 data sets and the red line is the true infection rate. Figure (b) is a histogram for the median estimate of γ for each of the 250 data sets.

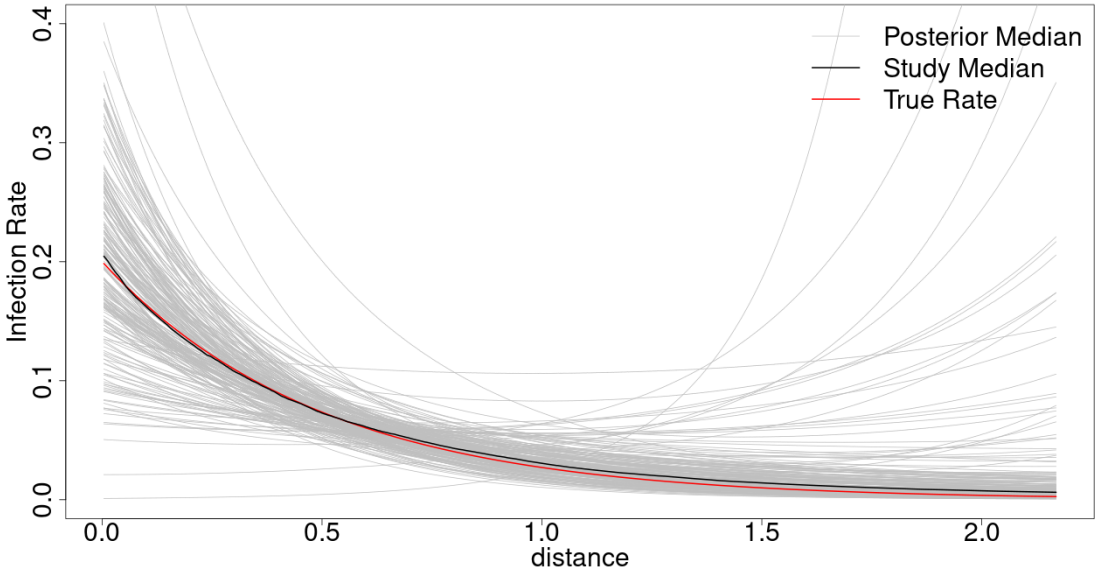
is given by:

$$\tilde{i} = \frac{\sum_{j=1}^{n} \hat{i}_j - \sum_{j=1}^{n} i_j}{\sum_{j=1}^{n} i_j},$$

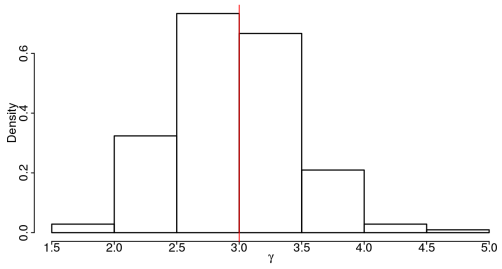
where $\hat{i}_j$ is the posterior median infection time for individual j. We use this statistic as it allows us to compare all the data sets, despite them having different numbers of infected individuals. We incur a small relative error when estimating the median value of γ, shown in figure 2.8(b), which leads to an overestimate in the infection times. For some data sets, the estimated infection rate is an increasing function. These very poor estimates correspond to data sets where the final size of the outbreak is small and the infected individuals lie almost exclusively in one cluster. When the infected individuals all lie in the bottom left cluster in figure (2.6), we only observe infections over small distances, and the GP cannot learn about the infection rate in long range transmissions.

2.8.1.3 Comparison of the full and MPA methods

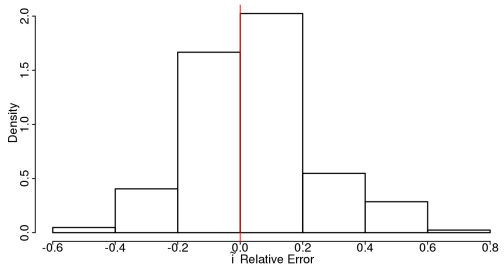
We now repeat the simulation study without infection times with the MPA method to compare to the full GP method. We use an input set of size 32 consisting of the smallest pair-wise distance, the largest pair-wise distance, and 30 other randomly chosen pair-wise distances. Figure 2.9(a) shows using the MPA does not affect the results and estimates the infection rate well. Figure 2.9(b) confirms this and shows that the MPA method infers the infection rate well, and introduces only a small amount of error. The mean absolute error for the full method is 0.0035, compared to 0.0055 for the MPA method. Figures 2.9(c-d) show that we continue to estimate the infection times and infectious period distribution rate parameter well.



(a) Estimates for the infection rate function

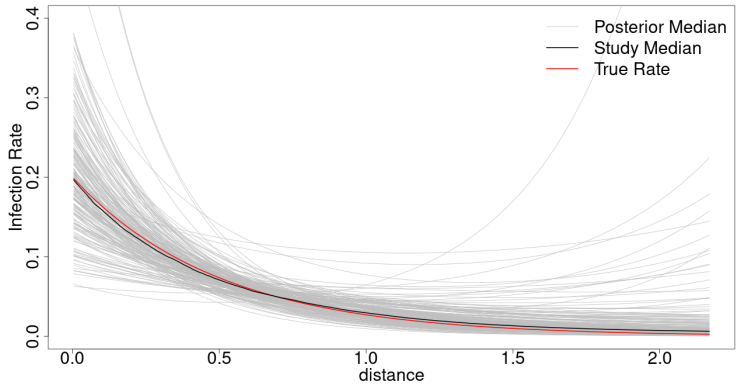


(b) Estimates for the infectious period distribution rate parameter

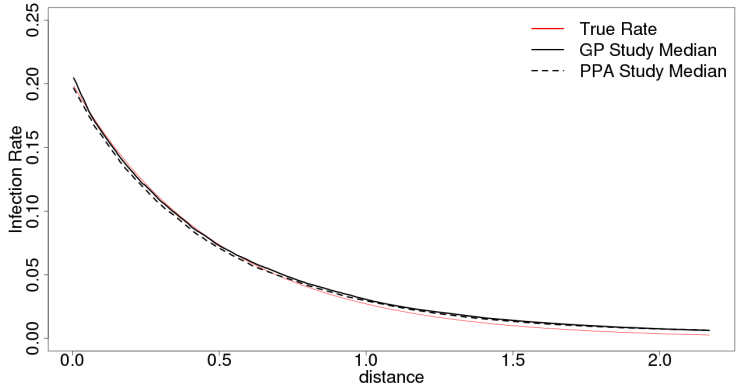


(c) The relative error for the sum of the infection times

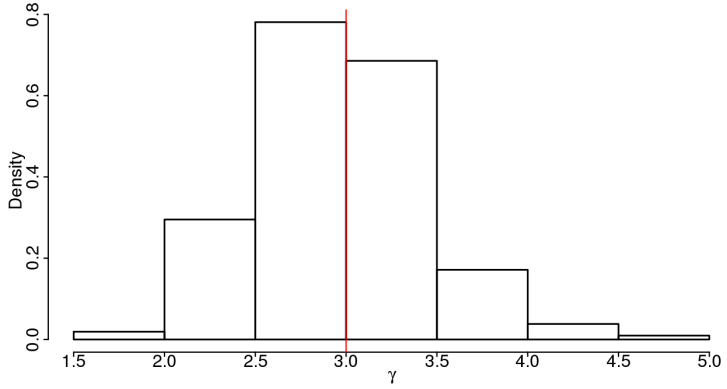
Figure 2.8: Results of the simulation study for a small population with unobserved infection times. Figure (a) shows the estimates for the infection rate function for each data set. Figure (b) shows the distribution of the infectious period distribution rate parameter estimates, and (c) shows the relative error in the sum of the infection times for the 250 data sets.



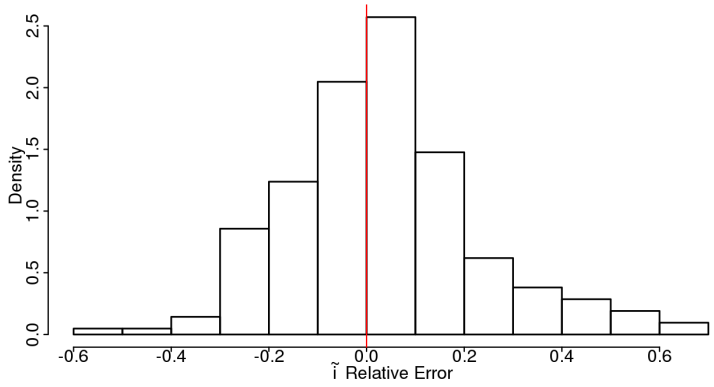
(a) Estimates for the infection rate function



(b) The GP study median compared to the MPA study median



(c) Estimates for the infectious period distribution rate parameter



(d) The relative error for the sum of the infection times

Figure 2.9: Results of the simulation study for a small population of size 100 in section 2.8.1 with unknown infection times and the MPA method.

Scenario	Parameter	Median	95% CI
Observed Infection Times	β_0	0.186	(0.117, 0.271)
	γ	3.00	(2.45, 3.713)
Unobserved Infection Times	β_0	0.203	(0.080, 0.365)
	γ	2.921	(1.903, 4.608)
	$\tilde{i}$	0.028	(-0.311, 0.522)
MPA with unobserved Infection Times	β_0	0.198	(0.078, 0.355)
	γ	2.925	(1.897, 4.612)
	$\tilde{i}$	0.029	(-0.342, 0.510)

Table 2.6: The median and 95% credible interval from the 250 medians for the three different scenarios in the small simulation study.

2.8.2 The MPA and the Pseudo Set

We now examine the effectiveness of the MPA when changing the number of input points in the pseudo set. We choose the first 50 data sets from the previous study and run the MPA inference algorithm with 8, 16, 32, 64, 128, 256, 512, 1024, and 2048 input points. The pseudo set consists of the smallest and largest pair-wise distances and the remainder are uniformly chosen. The use of the smallest and largest points in the set of pair-wise distances ensures the MPA method does not need to interpolate to cover the full data set. We choose the remaining points uniformly, as this ensures equal coverage across the entire domain. As we outline in section 2.6, we need to ensure there are sufficient input points at locations where we see changes in behaviour in the function, and in this study we assume the function behaves similarly across the domain, i.e. it has no turning points or singularities. We compare the results for each pseudo set to the results of the full algorithm. The full algorithm is equivalent to the MPA method with the whole data set as the input set.

Figure 2.10 shows the computational time required to run 30,000 iterations of the MPA inference algorithm, and the root mean square error (RMSE) of the study

median against the true rate. We define RMSE to be

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{N}\sum_{k=1}^{j}\left(\hat{\beta}(x_{j,k}) - \beta(x_{j,k})\right)^2}{N(N-1)/2}},$$

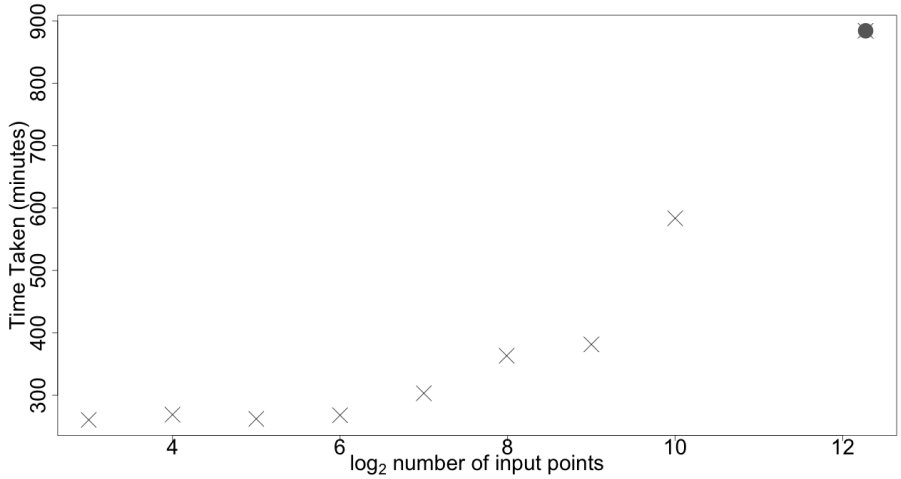
where $\hat{\beta}$ is the posterior median estimate for the true function β. The data point with the largest number of pseudo points corresponds to the full GP method. We can see from figure 2.10(a) that as the number of pseudo points increases, the computational time required increases exponentially, this is driven by the need to repeatedly project the pseudo function onto the complete data set. We can see from figure 2.10(b) the logistic type decrease in the RMSE incurred as the number of pseudo points increases. In order to obtain results with a similar order of error to that when using the full data set, we need to implement the MPA with a large number of pseudo points. However, including a large number of input points results in a similar computational time to that of the method with the full data set. Using fewer pseudo points reduces the computation time by up to 33% when compared to using the full data set and, although this method incurs the most error, the increase in root mean square error is less than 0.003.

2.8.3 Inference for a Large Population

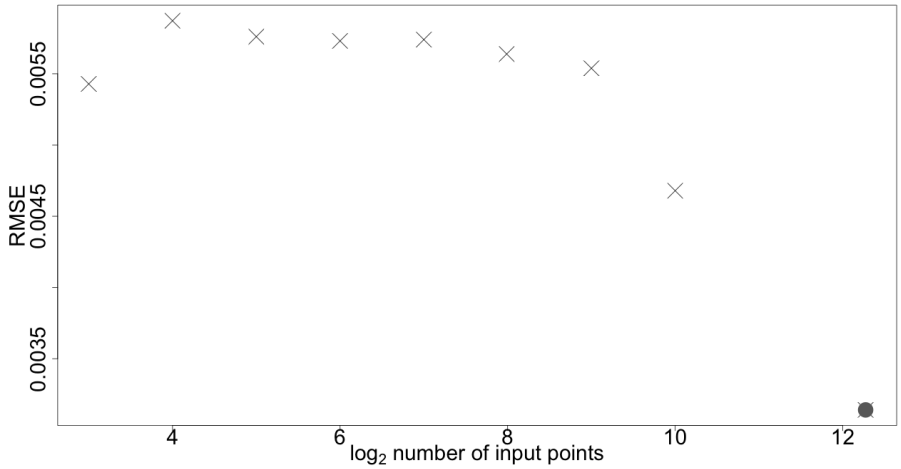
We generate the positions of 1000 individuals uniformly on $[0, 4]^2$, and simulate 100 outbreaks using the distance dependent infection rate in equation (2.17). To simulate the outbreaks, we used the parameter values shown in table 2.8.

Given 1000 individuals, there are 499,500 pair-wise distances between the individuals, so this data set is too large to use the full GP method. We apply the MPA method with 32 pseudo points chosen uniformly between 0 and $4\sqrt{2}$, the smallest and largest possible pairwise distances.

For each of the simulations, we use the MPA method to infer the infection rate, using only the distances and removal times as observed data. Figure 2.12 shows that we can estimate the infection rate effectively across the entire domain. The estimate for the large-scale simulation study is more accurate compared to the small study



(a) Size of pseudo data set against computational time



(b) Size of pseudo data set against RMSE

Figure 2.10: The results for the MPA simulation study in section 2.8.2. We show the computational time taken and the RMSE for the MPA method used to infer the infection rate, with different number of input points. The shaded point corresponds to the standard GP method.

Parameter	Value
N	1000
β_0	0.01
β_1	2
λ	3
γ	6
α	10

Table 2.7: Parameter values used to simulate outbreaks in the large population simulation study.

for two reasons: the number of data points has increased, and we are estimating the length scale. We can also estimate plausible values of the length scale hyperparameter, as shown in figure 2.11(a). As there is no true value for this parameter, there is no truth we can compare it to. As with the previous studies, we can estimate the infectious period distribution parameter well, which leads to a good estimate of the sum of the infection times. This is shown in figures 2.11(b-c) and the median estimates for the inferred parameters are shown in table 2.8.

2.8.4 Remarks on the Simulation Studies

In both the large and small simulation studies we were able to estimate the infection rate function well. We can also infer the infectious period distribution rate parameter and the times individuals were infected. Using an approximation method allows us to perform inference on much larger data sets as well as learn plausible values for the length scale parameter. Both of these allow us to give much better estimates for the infection rate. The method learns the infection rate function well over the entire set of distances, and is not affected by the cluster structure in the population.

One disadvantage to modelling the infection rate function using a Bayesian non-parametric method is the computational time required to run the MCMC algorithm. To estimate the infection rate function, infectious period distribution rate parameter

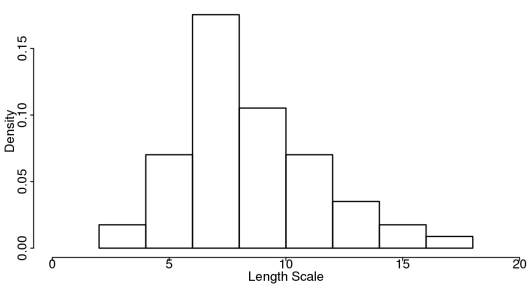
and infection times for a single data set in the large simulation study with population size 1,000 took on average 57.2 hours. This was using the University of Nottingham's High Performance Compute service alongside using the MPA, efficient computation of the likelihood function, and OpenMP software. This is compared to around 5 hours for a parametric alternative. There are three reasons for the large computational time: the length scale parameter, the likelihood function, and the infection times. Inference for the length scale parameter in the GP prior distribution requires us to repeatedly decompose and invert the covariance matrix which, even when using the MPA to reduce its dimension, takes a large amount of time. The likelihood function also takes a considerable amount of time to compute, and as we are using a nonparametric method, updating the infection rate function means updating every value of $\beta_{j,k}$ individually. As in the parametric method, estimating the infection times is very time consuming, especially when the population is large.

Parameter	True Value	Study Median	Study Credible Interval
β_0	0.01	0.01	(0.00782, 0.013)
γ	6	5.86	(4.79, 7.78)
$\tilde{i}$ Relative Error	0%	0.215%	(-5.88%, 4.78%)

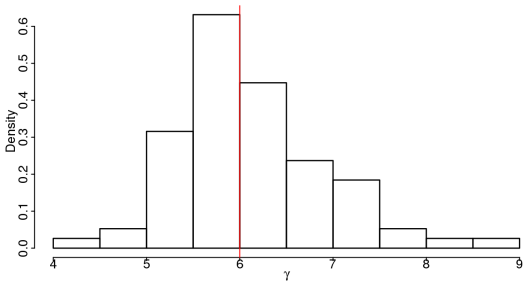
Table 2.8: Parameter values for the large simulation study. The study median refers to the median of the 100 posterior median estimates, and the study credible interval is a 95% credible interval of the posterior median estimates.

2.9 Conclusion

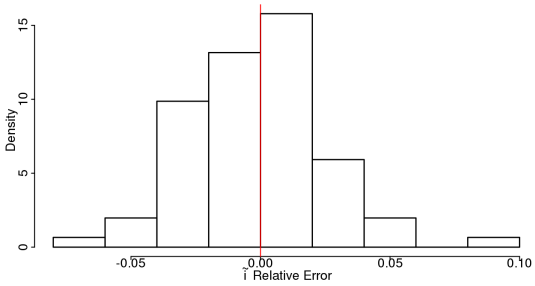
In this chapter, we have developed a novel method for inferring the infection rate function of a heterogeneously mixing infectious disease. Our method is the first Bayesian nonparametric method to infer infection rate functions for infectious disease models that are not time dependent. This method assumes that the infection rate depends on one characteristic of the relationship between the infector and susceptible individual, e.g. Euclidean distance. The first method uses a GP prior distribution



(a) Estimates for the GP prior distribution length scale parameter



(b) Estimates for the infectious period distribution rate parameter



(c) The relative error for the sum of the infection times

Figure 2.11: Results of the large simulation study with unknown infection times for the simulation study in section 2.8.3. In figure (a) we show the distribution of the median estimate for l, the length scale parameter, for each of the 100 data sets. Figure (b) shows the distribution of the infectious period distribution rate parameter estimates, and figure (c) shows the relative error in the sum of the infection times for the 100 data sets.

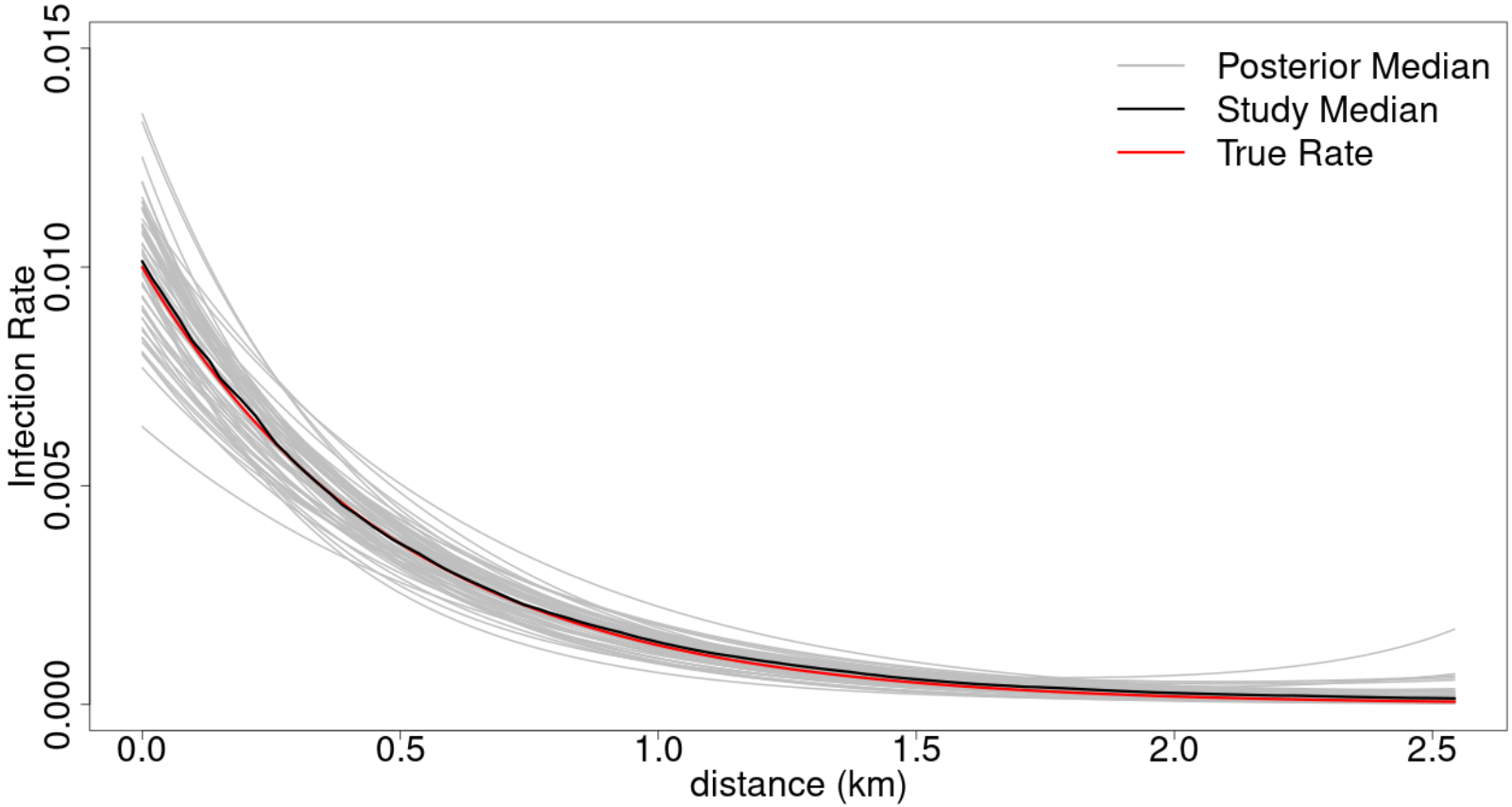


Figure 2.12: The estimated infection rate functions for the 100 data sets in the large simulation study. The grey lines are the posterior median estimates for the individual data sets. The black line is the median of the posterior medians, and the red line is the true rate.

and an MCMC algorithm to estimate the rate. This gives good results for small populations. However the computational requirements can become prohibitive for large data sets. To overcome this problem we implemented a standard approximation from the GP literature. The MPA method fits a GP to a small subset of the data and projects this onto the full data set, which allows us to use the same likelihood function and MCMC framework as in the original method. We then implemented a method for inferring the prior distribution length scale, improving the results and reducing the precision required in specifying the prior distribution. We were able to successfully use these methods alongside standard methods for inferring parameters of the infectious period distribution and infection times. Finally, we showed the effectiveness of our method using simulated data.

The framework we have developed reduces the need for arbitrary, parametric assumptions. Our Bayesian nonparametric model nevertheless requires some assumptions to be made. In particular, we have made decisions about the smoothness of the function through the choice of covariance function and length scale parameter. We can remove the need for specific assumptions about the length scale parameter by estimating plausible values for it. Although our method requires some assumptions to be made, these are less specific than and not as arbitrary as in parametric modelling. Results from parametric methods have less uncertainty associated with them than our Bayesian nonparametric methods. This is because the assumptions in parametric methods are much stronger, which give results that underestimate uncertainty. In some cases, this can lead to incorrect predictions. By using Bayesian nonparametric methods, we are able to avoid making such strong assumptions and better estimate the uncertainty.

One drawback of our method is that we need to be confident about the variance of the GP prior distribution. Underestimating this value yields slow convergence times for the Markov chain, or even no convergence in extreme cases. Overestimating this parameter also gives poor mixing of the chain as too few proposals are accepted. We have also assumed that the infection rate is a well-behaved function with no discontinuities, and our choice of covariance function outputs infinitely differentiable

functions – a characteristic unlikely to be observed in nature. One method of overcoming this would be to use a different covariance function, for example, the Màtern covariance function, where the number of times sample paths generated by the GP are differentiable can be specified. When using the MPA method, we need to ensure the pseudo dataset covers the entire domain and there are sufficient points in areas where the infection rate function has interesting features. These areas can be difficult to determine as the entire function is unknown. Bayesian nonparametric method also requires substantially more data than equivalent parametric methods, as we are inferring the infection rate at every point of the domain. Compared to the parametric methods, the uncertainty concerning our nonparametric estimate is much larger. This is an underlying property of nonparametric methods, however we can argue that parametric methods are too certain as we are including much stricter assumptions that may be incorrect. When estimating infection rate functions for individual outbreaks our uncertainty will be larger than if we had used parametric methods, and this will be carried across into the posterior predictive distribution.

We have shown that our method works well for one-dimensional infection rate functions, and we now move on to models where the infection rate depends on multiple parameters; for example size and distance or multi-type epidemics. This will be crucial when modelling diseases such as Foot and Mouth Disease, where the infection rate may depend factors such as distance between farms, type of animal, and number of animals on a farm.

CHAPTER 3

Bayesian Nonparametric Methods for Individual-Level Multi-Type Epidemics

3.1 Introduction

To more realistically model the spread of an infectious disease, we may want to allow the infection rate to differ between types of individuals as well as over a continuous covariate. For example, in an outbreak of Avian Influenza, the infection rate may not only depend on the distance between farms, but also on whether the type of animal on the farm were caged chickens or free-range turkeys.

These models are known as multi-type models; see Andersson and Britton (2000, §6) for a comprehensive overview. There are two main types of multi-type models: those where the infection rate depends on the social structure in the population, and those where infection rates differ due to the population consisting of different types. To take the social structure of a population into account, we consider which people share a household, workplace or school. In Ball et al. (1997), the authors propose a household model where there is an infection rate parameter that controls infections between individuals in the same household, as well as a parameter controlling infections

between members of different households. In this chapter, we are concerned with multi-type models where the infection rate differs between types of individuals. In Britton (1998), the author proposes a multi-type model where each individual in the population can be assigned a type, for example sex or age group. It is assumed that individuals of the same type act in an identical fashion, and the infection rate from one individual to another depends on their type. We consider the multi-type model proposed in Britton (1998) as opposed to Ball et al. (1997) as in chapter five, we will model outbreaks of disease among livestock, where we can assign each farm a type, for example cattle farm or sheep farm.

The multi-type model we consider is an extension of the *general* epidemic model as we allow for the population to consist of different types. The there are several possible ways of allowing for different types in the population. We outline an infectivity, susceptibility and dual model. The infectivity model is where the infection rate from individual i to j depends on the type of infectious individual i. Conversely, the infection rate in the susceptibility model depends on the type of the susceptible individual j. The dual model is where the infection rate depends on both the type of the infected individual i and the susceptible individual j, and this is the type of model analysed in Britton (1998).

To model on outbreak of an infectious disease in a population of multiple types, we suppose the population consists of m types. In an infectivity model, we assume that infectious contact occurs between a given type τ individual and another individual of any type at the time points of a homogeneous Poisson process with rate β_τ. In the susceptibility model, we assume the infectious contact occurs between a given infected individual of any type and a susceptible type τ individual according to a homogenous Poisson process with rate β_τ. Finally, in the dual model, the infection rate from a given individual, which is type τ, to a given individual, which is type ζ, is given by $\alpha_\tau\beta_\zeta$. Typically, the infectious period distribution is identical across all types, however it is possible to have distinct distributions for each type.

It is possible to allow for the infection rate to depend on other variables, as well as the type, and we do this by modelling the pair-wise infection process as an

inhomogeneous Poisson process. In Xu (2015, §3.4), the author develops a Bayesian nonparametric framework for modelling type-dependent infection rates in multi-type susceptibility models. In this model, the infection rate from a given individual of any type to a given type τ individual at time t is modelled by the function $\beta^{(\tau)}(t)$. The author uses a Sigmoidal Gaussian Cox Process to model the time-dependent infection rate in a Bayesian nonparametric framework. They then developed an MCMC algorithm for inferring the infection rate function, based on a method proposed in Adams et al. (2009) where the infection times are thinned.

In this chapter, we develop a multi-type susceptibility model and incorporate the methods developed in chapter two. We assume the infection rate function from a given individual of any type to an individual of type τ is given by the function $\beta^{(\tau)}$. Our method is simple to implement for both infectivity and dual models, however both of these models require more data to implement. In practice are more challenging to work with than susceptibility models. We now develop several methods for inferring the infection rate function for each type in a susceptibility model using GPs and then Multi-Output GPs (MOGPs).

MOGPs are a method for modelling several functions simultaneously and allowing for the functions to be dependent on or correlated to each other (Boyle and Frean, 2005). In Álvarez et al. (2012), the authors implement a MOGP framework where the interactions between the functions are described through the covariance matrices for the prior distributions for the functions. MOGPs have a wide variety of applications, particularly when there is limited data or the data is not very informative. This is because the MOGP framework allows information to be shared between the functions. For example in Alvarez et al. (2009), the authors infer the concentrations of various metals in a mine using a MOGP framework. One of the primary parameters of interest is the concentration of copper at various locations, and they supplement this with information about the concentration of lead, nickel and zinc. By using a MOGP framework and allowing the concentration of each of the four metals to be correlated, they are able to make better predictions about where copper-rich areas are located.

We now develop a MOGP framework for multi-type models. We are interested

where the infection rate functions in an individual-level stochastic epidemic model differs between types, as well as across a covariate such as the Euclidean distance. An example for this is in an outbreak of Avian Influenza. The infection rate between any two farms may depend on the distance between them and whether the farms house chickens or ducks. We model the infection rate function for each type nonparametrically using a GP prior distribution and using a MOGP framework allows the functions, which we assume to be similar, to share information across types.

3.1.1 Layout of Chapter

We begin the chapter by defining our multi-type model extending the likelihood function in the previous chapter to account for multiple types of individuals. In section 3.3 we introduce fixed effects models and implement a fixed effects approach for modelling multi-type epidemic models nonparametrically. A key piece of our nonparametric method is Multi-Output GPs, we describe them in section 3.4 and then apply them to a stochastic epidemic model framework in section 3.5.

This chapter contributes three methods for modelling multi-type outbreaks nonparametrically. A fixed-effects type method can be used when there is a lack of data. Two MOGP methods are then developed to allow for a more flexible model, where the infection rate functions between types are dependent but not identical. Most current methods for mutli-type models assume the infection rate functions for types are of the same form, and our method allows us to relax this assumption. We aim our two MOGP methods at different audiences: mathematical modeller and veterinary practitioners.

3.2 Multi-Type Transmission Model and The Likelihood Function

To allow for different types of individuals in our models, we define a multi-type model and construct the corresponding likelihood function. We recall the heterogeneously

mixing stochastic epidemic model from section 2.2. An infected individual j makes infectious contact with a susceptible individual k at the time points of a Poisson process with rate $\beta_{j,k}$. If individual k becomes infected, they remain so for a time period which is distributed according to a random variable following a $\Gamma(\lambda, \gamma)$ distribution, where λ is the shape parameter, and γ is the rate parameter.

In the previous chapter, we assumed the infection rate from individual j to k can be modelled as a function of some continuous covariate describing their relationship, for example the Euclidean distance between the individuals. We now extend this to a suite of functions which depend on the same covariate, but where the choice of function depends on the type of the susceptible individual. This is known as a susceptibility model. However our methodology will also work for infectivity models, where the infection depends on individual j as well as models where the infection rate depends on both j and k. We have chosen a susceptibility model as our methods require a large amount of data and removal times observed from infectivity and dual models are less informative than susceptibility models.

3.2.1 Multi-Type Transmission Model

We suppose that in a population of size N, each individual is one of $m << N$ types, and the infection rate differs between types. We consider the type to be a categorical variable. For modelling an outbreak of a disease among humans, examples of suitable categorial variables are the sex and ethnicity of the individuals. For diseases on farms, an example of a categorical variable is the type of livestock on the farm. We model the infection rate from individual j to k, which depends on the covariate $x_{j,k}$, by the following functions:

$$\beta_{j,k} = \begin{cases} \beta^{(1)}(x_{j,k}) & \text{if } k \text{ is type one,} \\ \quad \vdots & \\ \beta^{(m)}(x_{j,k}) & \text{if } k \text{ is type } m. \end{cases}$$

We recall the likelihood function for the heterogeneously mixing stochastic epi-

demic model in equation (2.1):

$$\pi(\mathbf{i},\mathbf{r}|\beta,\lambda,\gamma,\kappa,i_\kappa) = \exp\left\{-\sum_{j=1}^{n}\sum_{k=1}^{N}\beta(x_{j,k})\left((r_j \wedge i_k)-(i_j \wedge i_k)\right)\right\} \tag{2.1}$$
$$\times \prod_{\substack{j=1\\ j\neq\kappa}}^{n}\left(\sum_{k\in\mathcal{Y}_j}\beta(x_{k,j})\right)\prod_{j=1}^{n}h(r_j-i_j|\lambda,\gamma),$$

where $\mathcal{Y}_j = \{k; i_k < i_j < r_k\}$ is the set of individuals who are infectious immediately before individual j becomes infected.

Let $\mathbf{c} = \{c_1,\ldots,c_N\}$ be the set of types of the individuals, where individual j has type c_j. We assume these types to be fixed and known. Replacing the pair-wise infection rate $\beta_{j,k}$ for the type dependent infection rate function, $\beta^{(c_j)}$, gives us the following heterogeneously mixing multi-type likelihood function:

$$\pi(\mathbf{i},\mathbf{r}|\beta^{(1)},\ldots,\beta^{(m)},\lambda,\gamma,\kappa,i_\kappa,\mathbf{c}) = \exp\left\{-\sum_{j=1}^{n}\sum_{k=1}^{N}\beta^{(c_k)}(x_{j,k})\left((r_j \wedge i_k)-(i_j \wedge i_k)\right)\right\} \tag{3.1}$$
$$\times \prod_{\substack{j=1\\ j\neq\kappa}}^{n}\left(\sum_{k\in\mathcal{Y}_j}\beta^{(c_j)}(x_{k,j})\right)\prod_{j=1}^{n}h(r_j-i_j|\lambda,\gamma).$$

We can decompose the likelihood function into a product of terms, each element corresponding to a likelihood contribution for each type. Let C_τ be the set of individuals who are type τ and $\tilde{C}_\tau$ be the set of type τ individuals who were infected, then the contribution for type τ individuals is given by:

$$\pi_\tau = \exp\left\{-\sum_{j=1}^{n}\sum_{k\in C_\tau}\beta^{(\tau)}(x_{j,k})\left((r_j \wedge i_k)-(i_j \wedge i_k)\right)\right\}\prod_{\substack{j\in\tilde{C}_\tau\\ j\neq\kappa}}\left(\sum_{k\in\mathcal{Y}_j}\beta^{(\tau)}(x_{k,j})\right) \tag{3.2}$$
$$\times \prod_{j\in\tilde{C}_\tau} h(r_j-i_j|\lambda,\gamma).$$

We will use the decomposed form both analytically for the full conditional distributions of some parameters and computationally. The full likelihood function can be recovered by computing

$$\pi(\mathbf{i},\mathbf{r}|\beta^{(1)},\ldots,\beta^{(m)},\lambda,\gamma,\kappa,i_\kappa,\mathbf{c}) = \prod_{\tau=1}^{m}\pi_\tau.$$

3.3 Fixed Effects Model

The first model we propose is a fixed effects model. Fixed effects models are commonly used in regression problems where the data can be split into groups, the model allows for population-wide trends as well as group-specific trends. This idea can be used to model infection rate functions in a multi-type epidemic model as we allow for the infection rate function to have the same shape across all types, but a different scale for each type. We first outline nonparametric methods for a fixed effects model in a regression setting, then apply this to a stochastic epidemic model.

3.3.1 Fixed Effects Regression Model

Suppose we observe n points $\{y_{1,t}, \ldots, y_{n,c}\}$, for each of m categories, where $c = 1, \ldots, m$, at the input points $\{x_{1,c}, \ldots, x_{n,c}\}$. We fit the linear fixed effects model:

$$y_{i,c} = f(x_{i,c}) + \alpha_c + \varepsilon_{i,c}, \quad \varepsilon_{i,c} \overset{i.i.d}{\sim} N(0, \sigma^2), \quad i = 1, \ldots, n. \tag{3.3}$$

We assume there is an underlying generating function f that is common to all types. However the location parameter, α, which controls the y-intercept, is different for each type. We model the location type-effect for type j by including α_j. We can represent this model in the following matrix form:

$$\begin{pmatrix} y_{1,1} \\ \vdots \\ y_{n,1} \\ y_{1,2} \\ \vdots \\ y_{n,2} \\ \vdots \\ y_{1,m} \\ \vdots \\ y_{n,m} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & f(x_{1,1}) \\ \vdots & \vdots & & \vdots & \\ 1 & 0 & \cdots & 0 & f(x_{n,1}) \\ 0 & 1 & \cdots & 0 & f(x_{1,2}) \\ \vdots & \vdots & & \vdots & \\ 0 & 1 & \cdots & 0 & f(x_{n,2}) \\ \vdots & \vdots & & \vdots & \\ 0 & 0 & \cdots & 1 & f(x_{1,m}) \\ \vdots & \vdots & & \vdots & \\ 0 & 0 & \cdots & 1 & f(x_{n,m}) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \\ 1 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,1} \\ \vdots \\ \varepsilon_{n,m} \end{pmatrix}, \tag{3.4}$$

which can be written as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. We will refer to $\mathbf{Y}$ as the vector of observations, $\mathbf{X}$ as the design matrix, $\boldsymbol{\beta}$ as the vector of parameters, and $\boldsymbol{\varepsilon}$ as the error vector.

The likelihood function for the data is given by:

$$\pi(\mathbf{Y}|\boldsymbol{\beta}, f, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\beta)^T(\mathbf{Y}-\mathbf{X}\beta)\right\}.$$

Given data $\mathbf{Y}$, we can perform Bayesian inference for the values of the parameters, $\boldsymbol{\beta}$ and the function f. The posterior distribution is given by:

$$\pi(\beta, f|\mathbf{Y}, \sigma^2) \propto \pi(\mathbf{Y}|\beta, \sigma^2)\pi(\beta)\pi(f).$$

We put the following conjugate prior distribution on $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} \sim \mathcal{N}(0,\ \lambda^2\boldsymbol{I}). \tag{3.5}$$

We model f nonparametrically, by placing a GP prior distribution on the function as follows:

$$f \sim \mathcal{GP}(0,\ \Sigma), \quad \Sigma_{ij} = k(x_i, x_j, \alpha^2, l^2).$$

By Bayes' theorem, the posterior distribution is given by:

$$\begin{aligned}
\pi(\beta, f|\mathbf{Y}, \sigma^2) &\propto \pi(\mathbf{Y}|\beta, \sigma^2)\pi(\beta)\pi(f).\\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right\}\mathcal{GP}(f;0,\Sigma)\exp\left\{-\frac{1}{2\lambda^2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right\}\\
&= \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) - \frac{1}{2\lambda^2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right\}\mathcal{GP}(f;0,\Sigma)\\
&\propto \exp\left\{(\boldsymbol{\beta}-\boldsymbol{\mu}_n)^T\Sigma_n^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_n)\right\}\mathcal{GP}(f;0,\Sigma),\\
&\propto N(\boldsymbol{\beta};\ \boldsymbol{\mu}_n, \Sigma_n)\mathcal{GP}(f;0,\Sigma),
\end{aligned}$$

where $\Sigma_n = \left(\frac{1}{\sigma^2}\mathbf{X}^T\mathbf{X} + \frac{1}{\lambda^2}\mathbf{I}\right)^{-1}$, and $\boldsymbol{\mu}_n = \frac{1}{\sigma^2}\Sigma_n\mathbf{X}^T\mathbf{Y}$. The penultimate step in the derivation comes by completing the square and removing any terms dependent solely on $\mathbf{Y}$.

To generate samples from this distribution, we can construct an MCMC algorithm. The conditional posterior distribution for $\boldsymbol{\beta}$ has a closed form, and is given by:

$$\boldsymbol{\beta}|\mathbf{Y}, f, \sigma^2 \sim N(\boldsymbol{\mu}_n,\ \Sigma_n).$$

We can use a Gibbs step to sample values from the distribution. To sample f, we can use a Metropolis Hastings step with an underrelaxed proposal mechanism, we

described in section 2.5. Given then current state f, we propose values by:

$$f' = \sqrt{1-\delta^2} f + \delta\nu, \quad \nu \sim \mathcal{GP}(0,\, \Sigma).$$

We accept f' with probability

$$p_{acc} = \frac{\pi(\mathbf{Y}|\boldsymbol{\beta}, f', \sigma^2)}{\pi(\mathbf{Y}|\boldsymbol{\beta}, f, \sigma^2)}$$

Figure 3.1 shows the results of this Bayesian inference method applied to a data set where six observations in each of two categories are observed. The data was generated as follows:

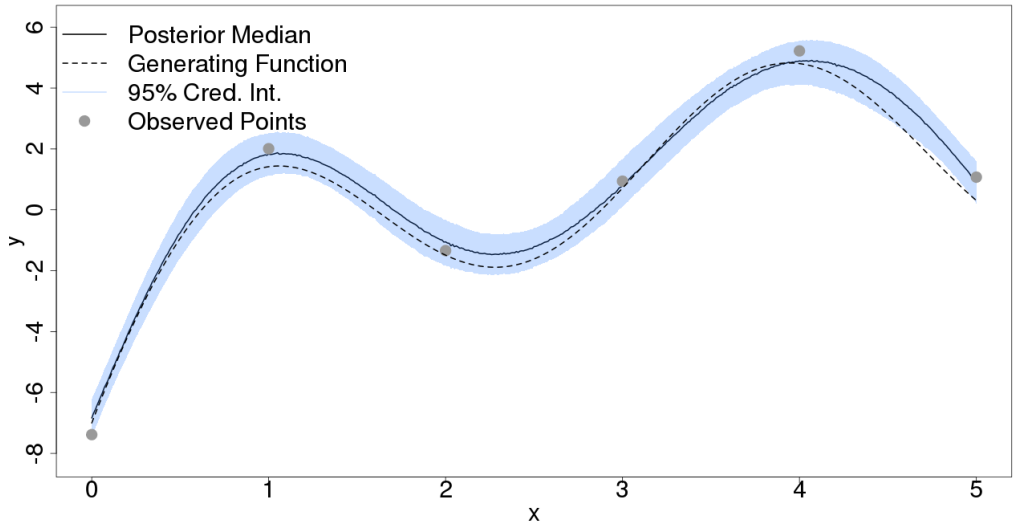
$$y_{i,1} = 2 + 3\sin(2x_{i,1}) + 3\exp(-3x_{i,1}) + \varepsilon_{i,1}$$
$$y_{i,2} = 4 + 3\sin(2x_{i,2}) + 3\exp(-3x_{i,2}) + \varepsilon_{i,2}$$

and the error term was given by a random variable following a $N(0, 0.5^2)$ distribution. By using a vague prior density in equation (3.5), where $\lambda^2 = 20^2$, we can infer the functions well. However, table 3.1 shows that none of the parameters are correctly identified. This is due to identifiability issues. Consider the points $\{(x_{1,1}, y_{1,1}),\, (x_{1,2}, y_{1,2})\} = \{(0, 5),\, (0, 7)\}$. We must solve the equations:

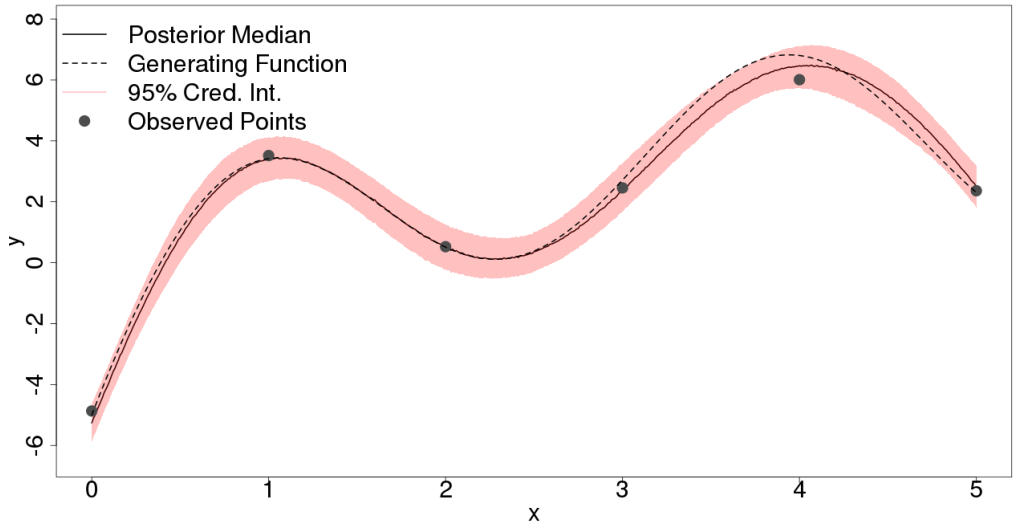
$$5 = \alpha_1 + f(0)$$
$$7 = \alpha_2 + f(0).$$

As there are three unknown parameters, α_1, α_2, and $f(0)$, and only two observations, we cannot fully identify the parameters individually. To resolve this, we need to reduce the number of degrees of freedom in the model by one, which is done by fixing some linear combination of the model parameters. The simplest way to do this is to fix $\alpha_1 = 0$, as the value we infer for α_2 is given by $\alpha_2 - \alpha_1$. This is relevant for the multi-type infection model as we will not be able to fully identify the susceptibility parameters, instead we can infer the relative susceptibility for one type compared to another.

We now apply this method to an epidemic model where the population can be split into different categories or types.



(a) Inferred values for function 1 compared to the true function.



(b) Inferred values for function 2 compared to the true function.

Figure 3.1: The fixed effects GP regression model in section 3.3.1 applied to a sample data set.

Parameter	True Value	Posterior Median	95% Credible Interval
α_1	2	6.86	(-2.17, 18.29)
α_2	4	8.43	(-0.603, 19.90)
$f(x_{1,1})$	3	-5.06	(-14.26, 20.03)

Table 3.1: Parameter estimates and credible intervals for the fixed effects GP regression model.

3.3.2 Fixed Effects Epidemic Model

Consider a population consisting of m types, labeled $1, \dots m$. We assume the infection rate from individual j to k can be split into two independent parts. The first part is the infection rate function, which depends on the covariate $x_{j,k}$, for example the distance between the individuals. The second part is a scaling parameter, which depends type of the susceptible individual k. We define the infection rate functions as:

$$\beta_{j,k} = \begin{cases} \beta_1 g(f(x_{j,k})) & \text{if individual } k \text{ is type 1,} \\ \quad \vdots & \\ \beta_m g(f(x_{j,k})) & \text{if individual } k \text{ is type } m. \end{cases}$$

This model is similar to the fixed effects regression model. Consider the log of the infection rate, which is given by:

$$\log \beta_{j,k} = \log \beta_{c_k} + \log f(x_{j,k}). \tag{3.6}$$

Given the removal times of the infected individuals and the covariate information, we wish to infer the scaling parameters $\beta_1, \dots, \beta_m$ and the function f. To resolve any identifiability problems, we set $\beta_1 = 1$, as this enables us to fully identify f and the susceptibility of each type relative to type one individuals, $\frac{\beta_2}{\beta_1}, \dots, \frac{\beta_m}{\beta_1}$. The form shown in equation (3.6) also allows us to simplify the decomposed likelihood function, as $\beta_2, \dots, \beta_m$ are constants. Allowing for this fixed effects form, the likelihood

contribution for type τ individuals in equation (3.2) becomes:

$$\pi_\tau = \exp\left\{-\beta_\tau \sum_{j=1}^{n} \sum_{k \in C_\tau} f(x_{j,k}) \left((r_j \wedge i_k) - (i_j \wedge i_k)\right)\right\} \beta_t^{n_t} \prod_{\substack{j \in \tilde{C}_\tau \\ j \neq \kappa}} \left(\sum_{k \in \mathcal{Y}_j} f(x_{j,k})\right) \quad (3.7)$$

$$\times \prod_{j \in \tilde{C}_\tau} h(r_j - i_j | \lambda, \gamma),$$

where n_τ is the number of type τ individuals infected. By Bayes' theorem, the posterior distribution is:

$$\pi(\beta_2, \ldots, \beta_p, f, \lambda, \gamma, \mathbf{i}, i_\kappa, \kappa | \mathbf{r}, \lambda, \alpha, \mathbf{c}) \propto \left(\prod_{\tau=1}^{m} \pi_\tau\right) \pi(f|l)\pi(l)\pi(\gamma)\pi(\gamma) \quad (3.8)$$

$$\times \pi(\kappa)\pi(i_\kappa|\kappa) \left(\prod_{j=2}^{m} \pi(\beta_j)\right).$$

3.3.3 MCMC Implementation

We place exponential prior distributions on $\beta_2, \ldots \beta_p$, a GP prior distribution on the function f over the space of covariate information, and a vague exponential prior distribution on the GP length scale. The complete set of prior distributions for this model is given by:

$$\begin{aligned}
&\beta_j \sim \text{Exp}(\chi_j), \qquad j = 2, \ldots, p \\
&\beta = g(f), \quad f \sim \mathcal{GP}(0,\, \Sigma), \qquad \Sigma_{jk} = k(x_j, x_k; \alpha, l) \\
&l \sim \text{Exp}(\chi_l) \\
&\gamma \sim \text{Exp}(\chi_\gamma) \\
&\kappa \sim \text{U}[1, \ldots, n] \\
&i_\kappa \sim -z, \qquad z \sim \text{Exp}(\chi_\kappa)
\end{aligned}$$

Typically, we will choose the parameters of the exponential prior distributions, $\chi_2, \ldots, \chi_p, \chi_l$, χ_γ and χ_κ to be small, e.g. 0.01, as this gives a vague prior distribution. As in the previous chapter, we fix $\lambda > 1$ to give a bell-shaped distribution for the infectious period distribution whose mean is defined by the rate parameter, γ.

3.3.3.1 Sampling the Scaling Parameter

The full conditional posterior distribution for the scaling parameter for type τ individuals, β_τ, is given by:

$$\pi(\beta_\tau|\mathbf{i},\mathbf{r},f,\kappa,i_\kappa) \propto \exp\left\{-\beta_\tau\left(\chi_\tau + \sum_{j=1}^{n}\sum_{k\in C_\tau} g(f(x_{j,k}))\left((r_j \wedge i_k) - (i_j \wedge i_k)\right)\right)\right\}\beta_\tau^{n_\tau+1}.$$

As we have placed a conjugate prior distribution on β_τ, the posterior distribution has a closed form given by the following gamma distribution:

$$\beta_\tau|\mathbf{i},\mathbf{r},f \sim \Gamma\left(n_\tau,\ \chi_\tau + \sum_{j=1}^{n}\sum_{k\in C_\tau} g(f(x_{j,k}))\left((r_j \wedge i_k) - (i_j \wedge i_k)\right)\right).$$

We can therefore use a Gibbs sampler step to sample values for $\beta_2,\ldots,\beta_m$.

3.3.3.2 Sampling the Infection Rate Function

As in the previous chapter, we propose a new function, f', using an underrelaxed proposal mechanism, where

$$f' = \sqrt{1-\delta^2}f + \delta\nu, \quad \nu \sim \mathcal{GP}(0,\ \Sigma),$$

and $\delta \in (0,1]$ is a tuning parameter. The acceptance probability for f' is

$$p_{acc} = \frac{\pi(\mathbf{i},\mathbf{r},\mathbf{c}|g(f'),\lambda,\gamma,\kappa,i_\kappa,\beta_2,\ldots,\beta_m)}{\pi(\mathbf{i},\mathbf{r},\mathbf{c}|g(f),\lambda,\gamma,\kappa,i_\kappa,\beta_2,\ldots,\beta_m)}.$$

We can use the Mean Projection Approximation (MPA) in the same way as the previous chapter in section 2.6. That is given the full set of n input points $\mathbf{x}$, we construct a set of m training points $\bar{\mathbf{x}}$. We then propose functions $\bar{f}$ over this set, and project it onto the full data set by:

$$f = \Sigma_{x,\bar{x}}\Sigma_{\bar{x},\bar{x}}^{-1}\bar{f}.$$

3.3.3.3 Sampling the Remaining Parameters

The remaining parameters are sampled as outlined in section 2.5. The outline of the MCMC algorithm is given in algorithm 6.

Algorithm 6 Structure of the MCMC algorithm

Initialise the chain with values $\gamma^{(0)}$, $\beta_2^{(0)}, \ldots, \beta_p^{(0)}$, $f^{(0)}$, $l^{(0)}$, and $\mathbf{i}^{(0)}$

Repeat the following steps

Sample γ from the conditional distribution $\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r})$ using a Gibbs sampler

for $t = 2, \ldots, p$ **do**

    Sample $\beta^{(t)}$ from $\pi(\beta^{(t)}|\mathbf{i}, \mathbf{r}, f, \kappa, i_\kappa)$ using a Gibbs step

end for

Sample f using an underrelaxed proposal method for a Metropolis Hastings step

Sample l using a Metropolis Hastings step

Sample κ and i_κ using a Metropolis Hastings step

Update an infection time

3.3.4 Simulation Study

We now test the method by using simulated data. We simulate the positions of 1,000 individuals on a unit square, and choose uniformly at random 350 individuals to be type one, 350 to be type two and the remaining 300 labelled type 3. We then simulate an outbreak of an infectious disease, where the infection rate is given by:

$$\beta(d_{j,k}) = \beta_{c_k} \exp\{-\mu d_{j,k}\},$$

where $d_{j,k}$ is the Euclidean distance between individuals j and k, β_{c_k} is the scaling parameter for individual k and μ is a rate parameter common to all types. To simulate the outbreaks, we used the parameters in table 3.2. We simulated 250 outbreaks of an infectious disease among the individuals conditioned on more than 100 being infected, and using only the removal times and positions, we inferred the infection rate, scaling parameters, infectious period distribution rate parameter and infection times. We ran the MCMC algorithm for 30,000 iterations for each data set, removing the first 1,000 iterations as a burn-in period. On a standard desktop PC, this took around 15 hours for each data set. We fixed the GP variance parameter to be $\alpha = 6$. The results for the shape of the infection rate are given in figure 3.2. Due to the identifiability problems, the rates have been rescaled so that we are comparing them

Parameter	Value
β_1	0.008
β_2	0.004
β_3	0.001
μ	2
λ	5
γ	3

Table 3.2: Values used to simulate outbreaks in the fixed effects model simulation study in section 3.3.4.

Parameter	True Value	Study Median	Study 95% Credible Interval
$\frac{\beta_2}{\beta_1}$	0.5	0.504	(0.379, 0.655)
$\frac{\beta_3}{\beta_1}$	0.125	0.127	(0.0938, 0.198)
γ	3	2.803	(2.065, 3.684)
$\tilde{i}$	0	0.0129	(-0.0321, 0.0637)
l	-	7.781	(4.695, 15.493)

Table 3.3: The fixed effects model simulation study median and credible intervals.

against $f(d) = \exp\{-2d\}$. These issues are because we are inferring the product of two parameters: the scaling parameter β_j and the function $g(f)$. As the value of $g(f)$ will depend on the value of β_j in each individual outbreak, and the outbreaks are independent we must fix one of the parameters to compare them. We therefore fix the y-intercept, $g(f(0))$, to be 1. We can see that, although the study median is close to the true rate, the variance is higher than for the standard, single-type method. This may be down to combined uncertainty in both the shape of the GP and the scale parameters. The estimates for the scale parameters are shown in figure 3.3, and we infer these parameters well. The numerical estimates and credible intervals are shown in table 3.3.

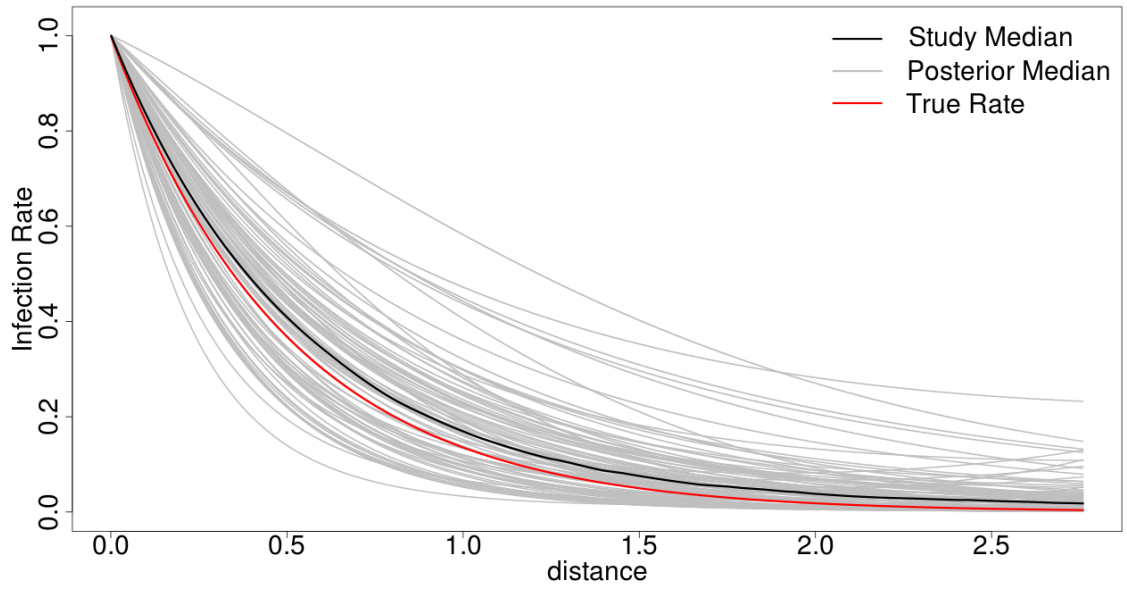
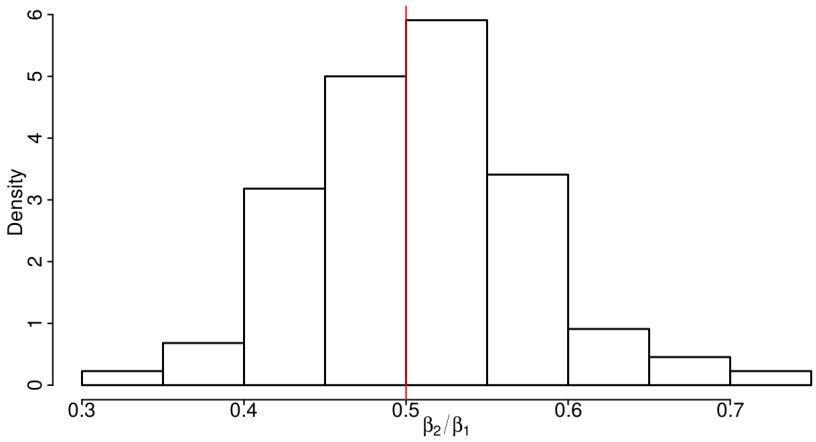


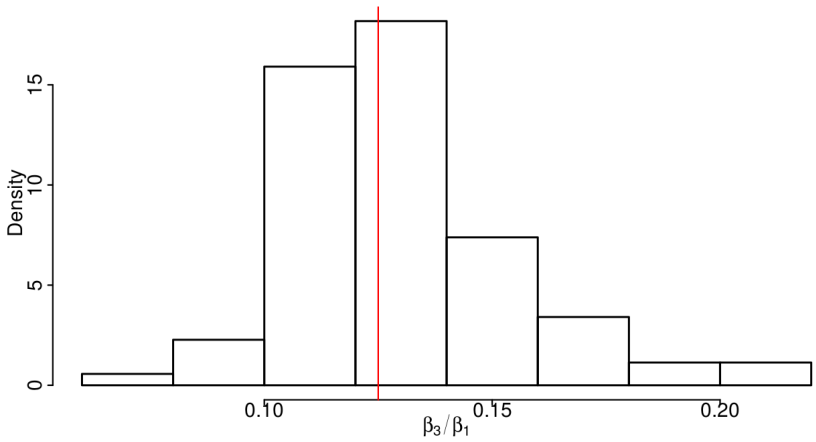
Figure 3.2: The 250 posterior medians for the infection rate, with the study median and the true rate using the simulated data for the fixed effects simulation study in section 3.3.1. As there are identifiability issues, the functions are rescaled so that the infection rate for immediate neighbours is 1.

3.4 Multi-Output Gaussian Processes

Multi-Output Gaussian Processes (MOGPs) are a method for learning multiple functions simultaneously, allowing for correlation and dependency between the functions. This allows the functions to collaborate and share the information learned (Nguyen and Bonilla, 2014). There are two main ways of introducing dependency between the functions. The first is to allow to noise or error terms in the different functions to be dependent on each other (Rasmussen and Williams, 2006, §5.4.3). This method requires the likelihood functions to be based on the normal distribution and is frequently used in regression models (Rasmussen and Williams, 2006, §9.2). The second method is to build a correlation structure into the prior distribution. This method has been studied in spatial statistics and is commonly referred to as co-kriging (Cressie, 1993). As our likelihood function does not have a Gaussian form, we will employ the second method.



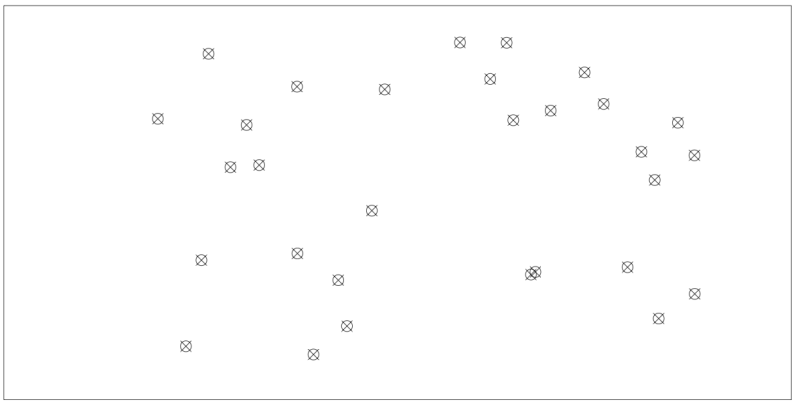
(a) The distribution of the median estimates for the 250 estimates of β_2/β_1.



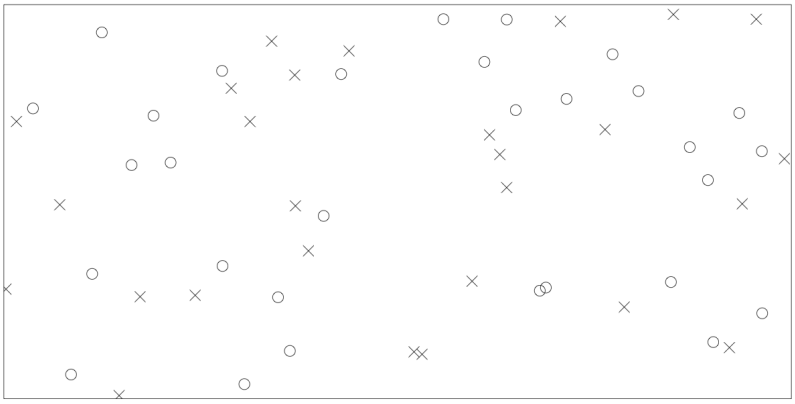
(b) The distribution of the median estimates for the 250 estimates of β_3/β_1.

Figure 3.3: The results of the fixed effects model applied to the 250 data sets from the simulated data in section 3.3.4.

There are several methods for using MOGPs. In Liu et al. (2018), the authors divide the MOGP with a correlated prior distribution methods into two sets: symmetric and asymmetric methods. Suppose we wish to infer m different functions, then using symmetric MOGPs, we allow each of the m prior distributions to depend on all of the other prior distributions. We can construct one large covariance matrix and draw samples from each of the prior distributions simultaneously. If we instead use an asymmetric method, the prior distributions are constructed sequentially, so



(a) Isotopic data



(b) Heterotopic data

Figure 3.4: Examples of heterotropic and isotropic data sets.

that the prior distribution for function j depends on the distribution for function $j-1$. We will develop both symmetric and asymmetric methods.

Before we outline the methods, we first define isotopic and heterotopic systems. If the functions we are modelling share the same input space, this is known as an isotopic system. In heterotopic systems, the functions have different input spaces. An illustration of both of these is given in figure 3.4. In this chapter we are concerned with heterotopic data.

Suppose we observe m correlated process, $f^{(1)}, \ldots, f^{(m)}$, which we wish to model

nonparametrically. We now outline three methods: the Independent GP model, the Multi-Output Covariance model (MOC) and a Discrepancy Based Model (DBM).

3.4.1 The Independent GP Model

We model the function the j^{th} function $f^{(j)}$ nonparametrically, each function is independent of the others. We place independent prior distributions on $f^{(j)}$ such that:

$$f^{(j)} \sim \mathcal{GP}(0,\, \Sigma^{(j)}), \quad \Sigma^{(j)}_{k,l} = k(x^{(j)}_k, x^{(j)}_l; \alpha, l), \quad j = 1 \ldots, m.$$

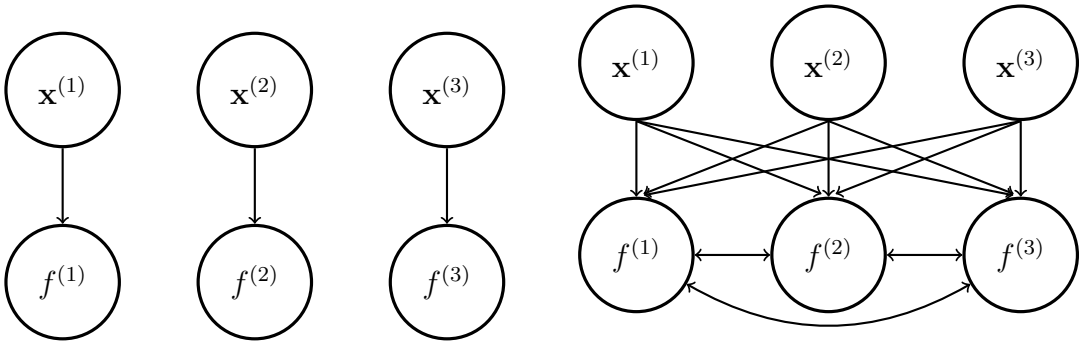
The advantage to this model is its simplicity, because we do not have to describe the relationship between any two functions. The independence of the prior distributions allows us to factorise the joint prior distribution over sets of observations, which is advantageous when the likelihood function can be similarly factorised. The simplicity of the model is also a drawback, as it does not capture the relationship between the functions.

3.4.2 The Multi-Output Covariance Model

For the MOC model, we place a joint GP prior distribution on the functions $f^{(1)}, \ldots, f^{(m)}$, such that

$$\begin{pmatrix} f^{(1)} \\ f^{(2)} \\ \vdots \\ f^{(m)} \end{pmatrix} \sim \mathcal{GP}\left(0,\, \begin{pmatrix} \Sigma^{(1,1)} & \rho_{1,2}\Sigma^{(1,2)} & \cdots & \rho_{1,m}\Sigma^{(1,m)} \\ \rho_{1,2}\Sigma^{(2,1)} & \Sigma^{(2,2)} & \cdots & \rho_{2,m}\Sigma^{(2,m)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,m}\Sigma^{(m,1)} & \rho_{2,m}\Sigma^{(m,2)} & \cdots & \Sigma^{(m,m)} \end{pmatrix}\right), \tag{3.9}$$

where $\Sigma^{(a,b)}_{l,m} = k(x^{(a)}_l, x^{(b)}_m; \alpha, l)$ and $\rho_{j,k}$ is the correlation between functions $f^{(j)}$ and $f^{(k)}$. Using the multi-output covariance block matrix allows us to draw samples for all functions simultaneously, as well as specify the correlation between the functions. This can be a disadvantage when there are a large number of functions, as there are a large number of correlation parameters. In our framework, we will assume that all functions have the same length scale parameter. This is because we would need



(a) Independent GPs (b) Multi Output Covariance Model

Figure 3.5: A graphical representation of the independent GP model and the Multi-Ouptut Covariance Model.

to learn the length scale parameters for different types, i.e. $\Sigma^{(j,k)}$, which are not immediately interpretable, and the data may not contain information about.

This model is a type of Common Principle Components Model, which allows for different population to consist of different groups, each with a unique covariance structure. The covariance structure dictates how the groups interact with each oher. This was first proposed in Flury (1984). In the MOC model, we assume the the covariance is given by the squared exponential function, and the relationships between the groups is given by a correlation parameter. Graphical representations of the independent GP model and the MOC model are given in figure 3.5.

3.4.2.1 MOC Regression Model

We now give an example of using the MOC model to fit curves to a regression data set. Suppose we observe the data points shown in figure 3.6, where $\mathbf{x}^{(1)} = \mathbf{x}^{(2)} = \{0, \ldots, 10\}$, $\mathbf{y}_1 = \sin(\mathbf{x}^{(1)}) + \log(\mathbf{x}^{(1)} + 1) + \boldsymbol{\varepsilon}$, $\mathbf{y}_2 = \sin(\mathbf{x}^{(2)}) + \log(\mathbf{x}^{(2)} + 1) + \frac{1}{2}\mathbf{x}^{(2)} + \boldsymbol{\varepsilon}$, and $\varepsilon_i \overset{i.i.d}{\sim} N(0, 0.1)$. We wish to fit curves $f^{(1)}$ and $f^{(2)}$ through $\mathbf{y}_1$ and $\mathbf{y}_2$ over the input points $\mathbf{x}^{(j)*} = \{0, 0.01, 0.02, \ldots, 10\}$. The posterior distribution is given by:

$$\pi(f^{(1)}, f^{(2)} | \mathbf{y}, \sigma^2, \rho) \propto \pi(\mathbf{y} | f^{(1)}, f^{(2)}, \sigma^2) \pi(f^{(1)}, f^{(2)} | \rho).$$

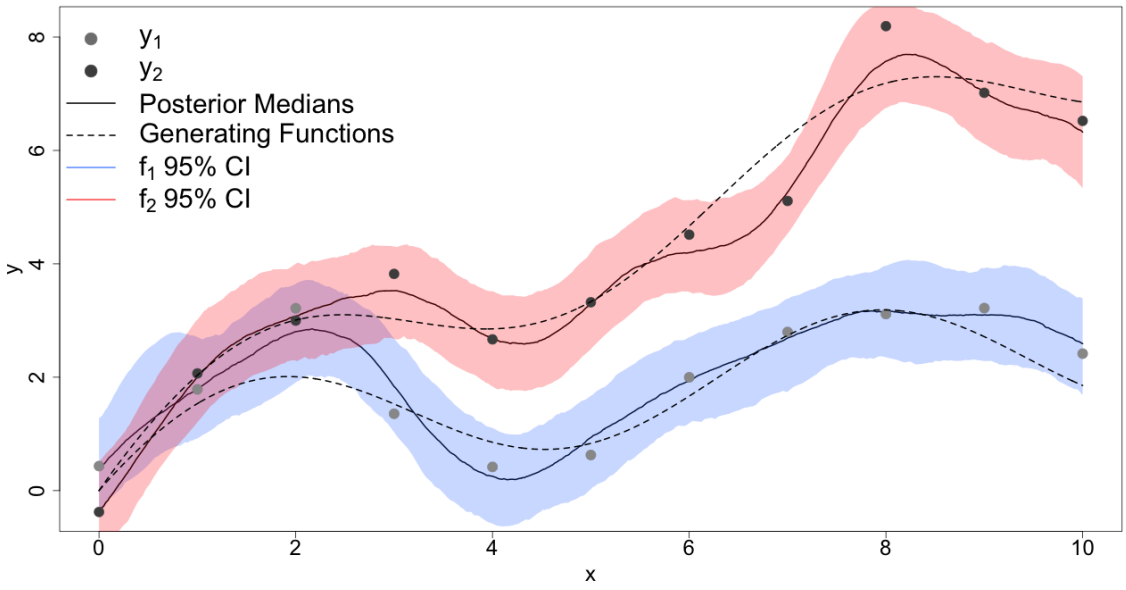


Figure 3.6: The results of the MOC model applied to a basic regression problem.

We place the following MOC GP prior distribution on $f^{(1)}$ and $f^{(2)}$:

$$\begin{pmatrix} f^{(1)} \\ f^{(2)} \end{pmatrix} \sim \mathcal{GP}\left(0, \begin{pmatrix} \Sigma^{(1,1)} & 0.3\Sigma^{(1,2)} \\ 0.3\Sigma^{(2,1)} & \Sigma^{(2,2)} \end{pmatrix}\right),$$

where $\Sigma^{(i,j)}_{l,m} = k({x^{(i)}}^*_l, {x^{(j)}}^*_m; \alpha, l)$. We set the GP length scale to be 2 and the variance to be 2. The results of 10,000 runs of a Metropolis-Hasting algorithm with an underrelaxed proposal mechanism is shown in figure 3.6. This method gives a reasonable fit to the data, although it is evident that the length scale is more suitable for $f^{(1)}$ than $f^{(2)}$.

3.4.3 The Discrepancy Based Model

The Discrepancy-Based Model (DBM) is constructed such that the prior distribution for function j depends on the prior distribution for function $j-1$. As this method requires isotropic data, we first create a pseudo data set $\bar{\mathbf{x}}$ based on $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$.

We then set the following prior distributions:

$$\begin{aligned}
\bar{f}^{(1)} &\sim \mathcal{GP}(0,\, \Sigma^{(1)}), \quad \Sigma^{(1)}_{j,k} = k(\bar{x}_j, \bar{x}_k;\, \alpha, l_1) \\
u^{(2)} &\sim \mathcal{GP}(0,\, \Sigma^{(2)}), \quad \Sigma^{(2)}_{j,k} = k(\bar{x}_j, \bar{x}_k;\, \alpha, l_2), \quad \bar{f}^{(2)} = \bar{f}^{(1)} + u^{(2)} \\
&\vdots \\
u^{(m)} &\sim \mathcal{GP}(0,\, \Sigma^{(m)}), \quad \Sigma^{(m)}_{j,k} = k(\bar{x}_j, \bar{x}_k;\, \alpha, l_m), \quad \bar{f}^{(m)} = \bar{f}_{m-1} + u^{(m)},
\end{aligned}$$

where $\bar{f}^{(j)}$ is the function for data set j over the pseudo data set. After sampling from the prior distributions and computing the discrepancies, we then project each function $\bar{f}^{(j)}$ onto its respective input data set $\mathbf{x}^{(j)}$. An example of this model with 3 functions is shown in figure 3.7.

The model allows us to asses how each function differs from the previous function. It also lets us choose one function as a baseline, but without the loss of information incurred when using the fixed effects model. One drawback of using this model is that the the input data must be isotopic, as each function is used in the prior distribution for the following function. However, as we are likely to be using the MPA method from the previous chapter, this is not a severe limitation as long as all functions share a similar input domain.

3.4.3.1 Discrepancy Based Regression Model

We now apply this model to fit two curves to a data set and we use the same data set as in the MOC model. The posterior distribution is given by:

$$\pi(f^{(1)}, f^{(2)}|\mathbf{y}, \sigma^2) \propto \pi(\mathbf{y}|f^{(1)}, f^{(2)}, \sigma^2)\pi(f^{(1)})\pi(f^{(2)}|f^{(1)}).$$

The results are shown in figure 3.8, and we can see the median estimates for both functions fit the data well. We can also compare the differences between the functions, as shown in figure 3.8(b). This shows the functions are substantially different after approximately one unit.

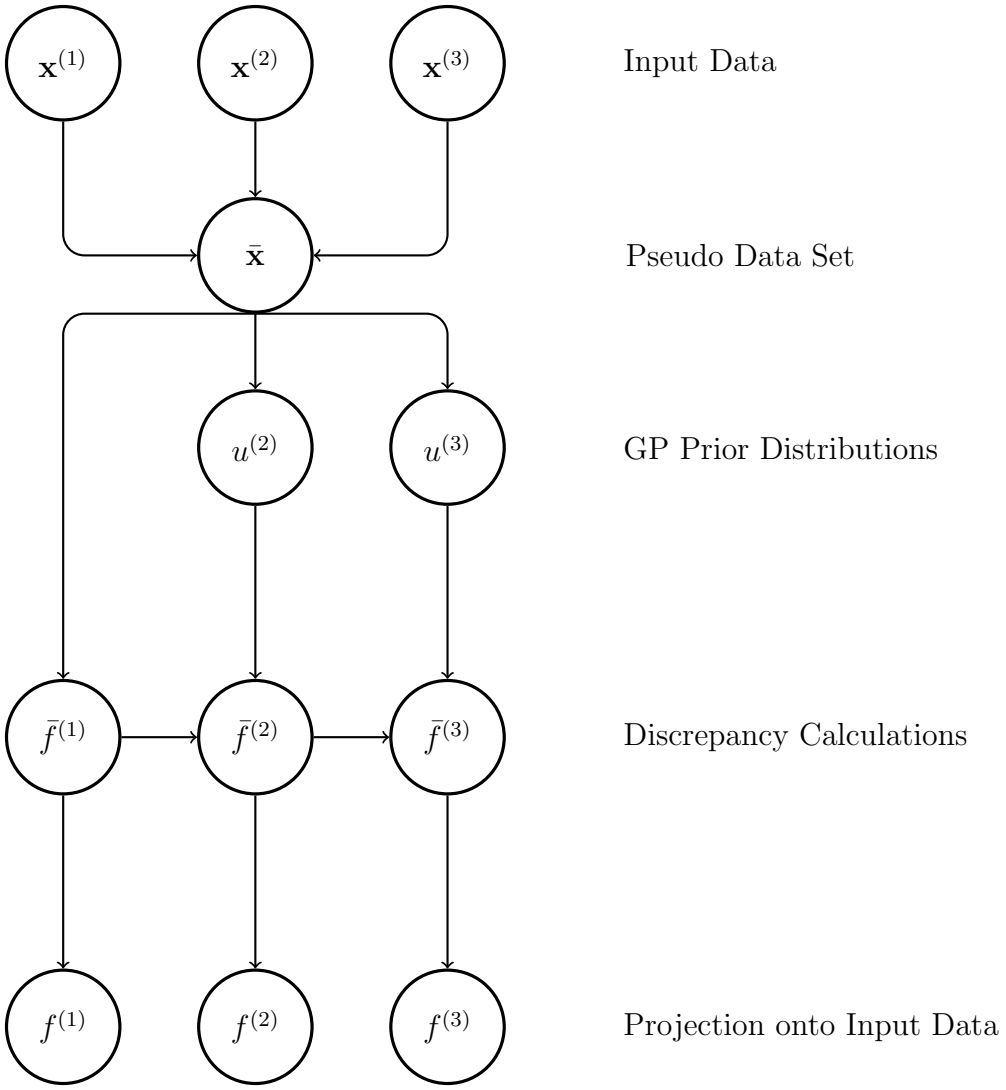
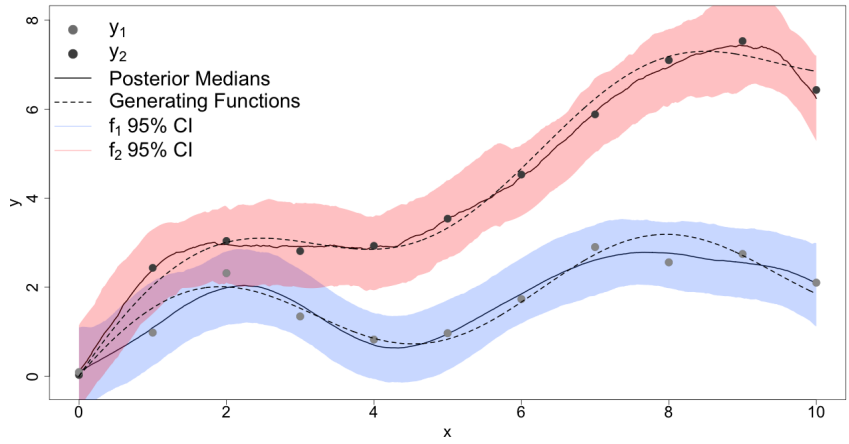


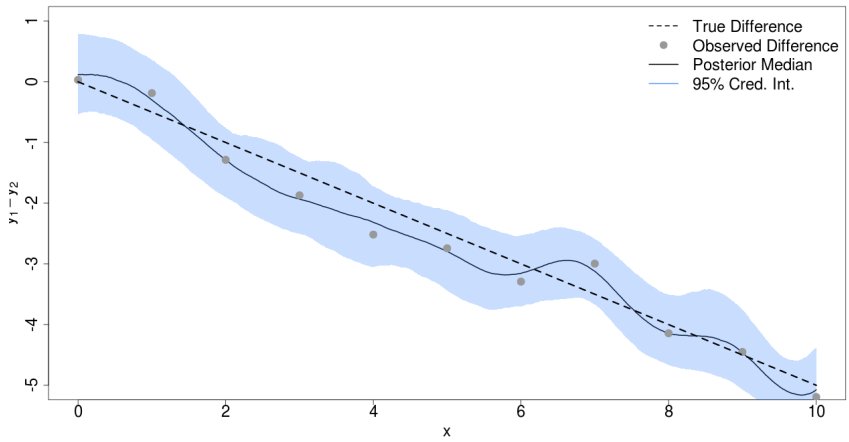
Figure 3.7: The prior distribution structure for the DBM with three functions.

3.4.4 Comparison of the Multi-Output Gaussian Process Models

The models offer different advantages and are suited to different settings. The MOC model allows us to describe the covariance between functions, whereas the DBM allows us to more directly compare estimates for functions. The independent model assumes the functions are independent of each other *a priori* and allows us to factorise the prior distributions. Another important difference is the length scales of the prior



(a) The results of the DBM prior distribution applied to a basic regression problem.



(b) The estimated difference between the two regression curves with 95% credible interval

Figure 3.8: The results of the DBM applied to a basic regression problem. The posterior medians for the functions are shown with a black solid line, and the functions which generated the data are shown with a dashed line.

distribution, the DBM and independent model allow us to set different length scales for each function. However, we assume the length scales are all identical for the MOC model.

Both the DBM and MOC model allow us to state the relationship between

the functions. The MOC model describes the relationship through the correlation parameters $\rho_{i,j}$, whereas the DBM allows us to compare the functions more directly through the discrepancy calculations. This form is likely to be more accessible to practitioners applying this method to data than by estimating the correlation in the MOC model.

The models also have different advantages computationally and in the MCMC algorithm. The MOC model allows us to draw samples for all functions jointly, however it does require a larger covariance matrix, which will require more time and memory to decompose. As the covariance matrix is a block matrix, the eigenvalues can be numerically difficult to obtain, and can result is numerical instabilities when inverting or decomposing the matrix, particularly when $p_{i,j}$ is close to 1 and the off-diagonal matrices are similar. For this we recommend estimating either the covariance ρ or the length scale l, but not both. The DBM has a smaller covariance matrix, so does not suffer from this problem. However the samples from the prior distributions must be drawn sequentially.

3.5 MOGPs for Multi-Type Models

We now apply the three methods to modelling the infection rate functions in an individual-level multi-type epidemic model. For notational convenience, we restrict ourselves to populations consisting of two types. It is, however possible for the methods to be extended to a larger number of types. We denote the vector of covariate information for type one and type two individuals by $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ respectively, and the functions for type one and type two individuals by $\beta^{(1)}$ and $\beta^{(2)}$ respectively.

3.5.1 Independent Gaussian Process Model

The first model assumes that the values of $\beta^{(1)}$ and $\beta^{(2)}$ are independent of each other, and that information about one of the vectors does not help in learning the

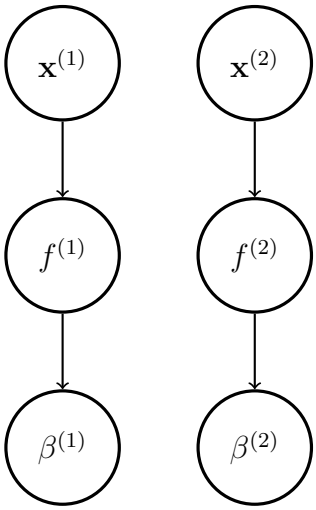


Figure 3.9: The independent GP prior distribution structure for epidemic models.

other. We place independent prior distributions on $\beta^{(1)}$ and $\beta^{(2)}$ such that:

$$\beta^{(1)} = g(f^{(1)}), \quad f^{(1)} \sim \mathcal{GP}(0, \Sigma^{(1)}), \quad \Sigma^{(1)}_{j,k} = k(x^{(1)}_j, x^{(1)}_k; \alpha, l). \tag{3.10}$$

$$\beta^{(2)} = g(f^{(2)}), \quad f^{(2)} \sim \mathcal{GP}(0, \Sigma^{(2)}), \quad \Sigma^{(2)}_{j,k} = k(x^{(2)}_j, x^{(2)}_k; \alpha, l). \tag{3.11}$$

We sample the infection rate functions using the underrelaxed proposal mechanism for a MH algorithm outlined in section 2.5. For large data sets we can use the MPA outlined in equation (2.11), repeating this step for each function. The remaining parameters and infection times are updated as before, and an outline of the MCMC algorithm is shown in algorithm 7.

3.5.2 Multi-Output Covariance Model

In the MOC model, we allow the infection rate functions for the two types to be correlated. A graphical illustration of this is shown in figure 3.10.We define the covariance between the two functions to be ρ, and put the following joint prior distribution on $\beta^{(1)}$ and $\beta^{(2)}$:

$$\beta^{(1)} = g(f^{(1)}),\ \beta^{(2)} = g(f^{(2)}), \quad \begin{pmatrix} f^{(1)} \\ f^{(2)} \end{pmatrix} \sim \mathcal{GP}\left(0, \begin{pmatrix} \Sigma^{(1,1)} & \rho\Sigma^{(1,2)} \\ \rho\Sigma^{(2,1)} & \Sigma^{(2,2)} \end{pmatrix}\right),$$
$$\Sigma^{(l,m)}_{j,k} = k\left(x^{(l)}_j, x^{(m)}_k; \alpha, l\right). \tag{3.12}$$

Algorithm 7 Structure of the Independent GP MCMC algorithm

1: Initialise the chain with estimates $\gamma^{(0)}$, $(f^{(1)})^{(0)}$, $(f^{(2)})^{(0)}$, $l_1^{(0)}$, $l_2^{(0)}$, and $\mathbf{i}^{(0)}$

Repeat the following steps

2: Sample γ from the conditional distribution $\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r})$ using a Gibbs step

3: Sample $f^{(1)}$ and $f^{(2)}$ using an underrelaxed proposal method for a Metropolis Hastings step

4: Sample l_1 using a Metropolis Hastings step

5: Sample l_2 using a Metropolis Hastings step

6: Sample κ and i_κ using a Metropolis Hastings step

7: Update an infection time

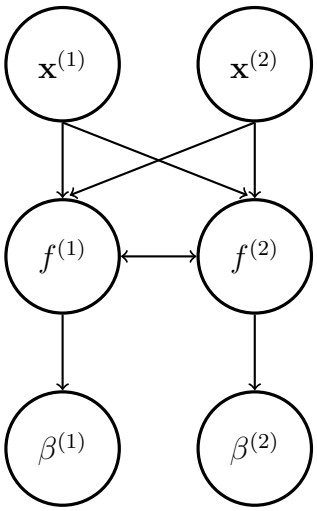


Figure 3.10: The MOC prior distribution for multi-type epidemic models.

We note that this model with $\rho = 0$ is equivalent to the independent GP model. A graphical representation for this model is shown in figure 3.10. To sample values for $\beta^{(1)}$ and $\beta^{(2)}$, we again use the Metropolis Hastings algorithm with an underrelaxed proposal mechanism, and we propose new values by:

$$\beta^{(1)\prime} = g(f^{(1)\prime}),\ \beta^{(2)\prime} = g(f^{(1)\prime}),$$
$$\begin{pmatrix} f^{(1)\prime} \\ f^{(2)\prime} \end{pmatrix} = \sqrt{1-\delta^2} \begin{pmatrix} f^{(1)} \\ f^{(2)} \end{pmatrix} + \delta \begin{pmatrix} \nu^{(1)} \\ \nu^{(2)} \end{pmatrix}, \begin{pmatrix} \nu^{(1)} \\ \nu^{(2)} \end{pmatrix} \sim \mathcal{GP} \left(0, \begin{pmatrix} \Sigma^{(1,1)} & \rho\Sigma^{(1,2)} \\ \rho\Sigma^{(2,1)} & \Sigma^{(2,2)} \end{pmatrix} \right).$$

We then accept the proposed values with probability

$$p_{acc} = \frac{\pi(\mathbf{i}, \mathbf{r}|\beta^{(1)\prime}, \beta^{(2)\prime}, \lambda, \gamma, \kappa, i_\kappa, \mathbf{c})}{\pi(\mathbf{i}, \mathbf{r}|\beta^{(1)}, \beta^{(2)}, \lambda, \gamma, \kappa, i_\kappa, \mathbf{c})}.$$

For large data sets, we again use the MPA outlined in equation. As each function has unique and independent training and full data sets, we project each function independently. We do this by computing

$$f^{(j)} = \Sigma_{j,\bar{x}} \Sigma^{-1}_{\bar{x},\bar{x}} \bar{f}^{(j)}, \tag{3.13}$$

where $\Sigma_{j,\bar{x}}$ is the covariance matrix constructed from set $\mathbf{x}_j$ and the pseudo input set $\bar{\mathbf{x}}$. We place a uniform prior distribution on ρ, such that

$$\rho \sim \mathrm{U}[-1, 1].$$

If it is believed *a priori* that the infection rate functions are positively correlated, it is possible to use a $\mathrm{U}[0, 1]$ prior distribution and similarly a $\mathrm{U}[-1, 0]$ if negative correlation is assumed *a priori*. By using a closed interval, we can estimate whether $\rho = 0$, in which case there is no correlation between the infection rate functions, or whether $\rho = 1$, in which case the infection rate functions for both types are identical. To sample this parameter we use a Metropolis Hastings Random walk algorithm, and propose new values by

$$\rho' \sim N(\rho, \sigma^2_\rho).$$

We accept ρ' with probability

$$p_{acc} = \frac{\frac{1}{\sqrt{|\Sigma'|}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} f^{(1)} & f^{(2)} \end{pmatrix} \Sigma'^{-1} \begin{pmatrix} f^{(1)} \\ f^{(2)} \end{pmatrix} \right\}}{\frac{1}{\sqrt{t|\Sigma|}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} f^{(1)} & f^{(2)} \end{pmatrix} \Sigma^{-1} \begin{pmatrix} f^{(1)} \\ f^{(2)} \end{pmatrix} \right\}} \wedge 1,$$

where $\Sigma' = \begin{pmatrix} \Sigma^{(1,1)} & \rho'\Sigma^{(1,2)} \\ \rho'\Sigma^{(2,1)} & \Sigma^{(2,2)} \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \Sigma^{(1,1)} & \rho\Sigma^{(1,2)} \\ \rho\Sigma^{(2,1)} & \Sigma^{(2,2)} \end{pmatrix}$.

The full MCMC algorithm with sampling steps for all model parameters is given in algorithm 8.

Algorithm 8 Structure of the MOC model MCMC algorithm

1: Initialise the chain with values $\gamma^{(0)}$, $\left(f^{(1)}\right)^{(0)}$, $\left(f^{(2)}\right)^{(0)}$, $\rho^{(0)}$ $l^{(0)}$, and $\mathbf{i}^{(0)}$

Repeat the following steps

2: Sample γ from the conditional distribution $\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r}, \chi_\gamma)$ using a Gibbs step

3: Sample $f^{(1)}$ and $f^{(2)}$ using an underrelaxed proposal method for a Metropolis Hastings step

4: Sample ρ using a Metropolis Hasting Random Walk step

5: *Optional: Sample l using a Metropolis Hastings Random Walk step*

6: Sample κ and i_κ using a Metropolis Hastings step

7: Update an infection time

3.5.3 Discrepancy Based Model

In this model, we treat the infection rate function for type one individuals as a baseline, and compute the infection rate function for type two individuals conditioned on the function for type one individuals. However, as this method requires isotopic data, we introduce a pseudo data set $\bar{\mathbf{x}}$, which is constructed from the input sets $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$. Given the training set $\bar{\mathbf{x}}$, we sample both $\bar{f}^{(1)}$ and $\bar{f}^{(2)}$ over $\bar{\mathbf{x}}$. Combining

the training set with the discrepancy assumption gives the prior distributions:

$$\bar{f}^{(1)} \sim \mathcal{GP}(0,\, \Sigma^{(1)}), \quad \Sigma^{(1)}_{jk} = k(\bar{x}_j, \bar{x}_k;\, \alpha, l_1).$$
$$\bar{f}^{(2)} = \bar{f}^{(1)} + u, \quad u \sim \mathcal{GP}(0,\, \Sigma^{(2)}), \quad \Sigma^{(2)}_{jk} = k(\bar{x}_j, \bar{x}_k;\, \alpha, l_2).$$

We then project $\bar{f}^{(1)}$ and $\bar{f}^{(2)}$ onto their respective full data sets using the MPA outlined in equation (3.13). The infection rate functions are given by $\beta^{(j)} = g(f^{(j)})$, where $f^{(j)}$ is the projected function for type j. We can extend the graphical representation of the DBM in figure 3.7 to include the transformation into infection rate functions. A graphical representation is given in figure 3.11.

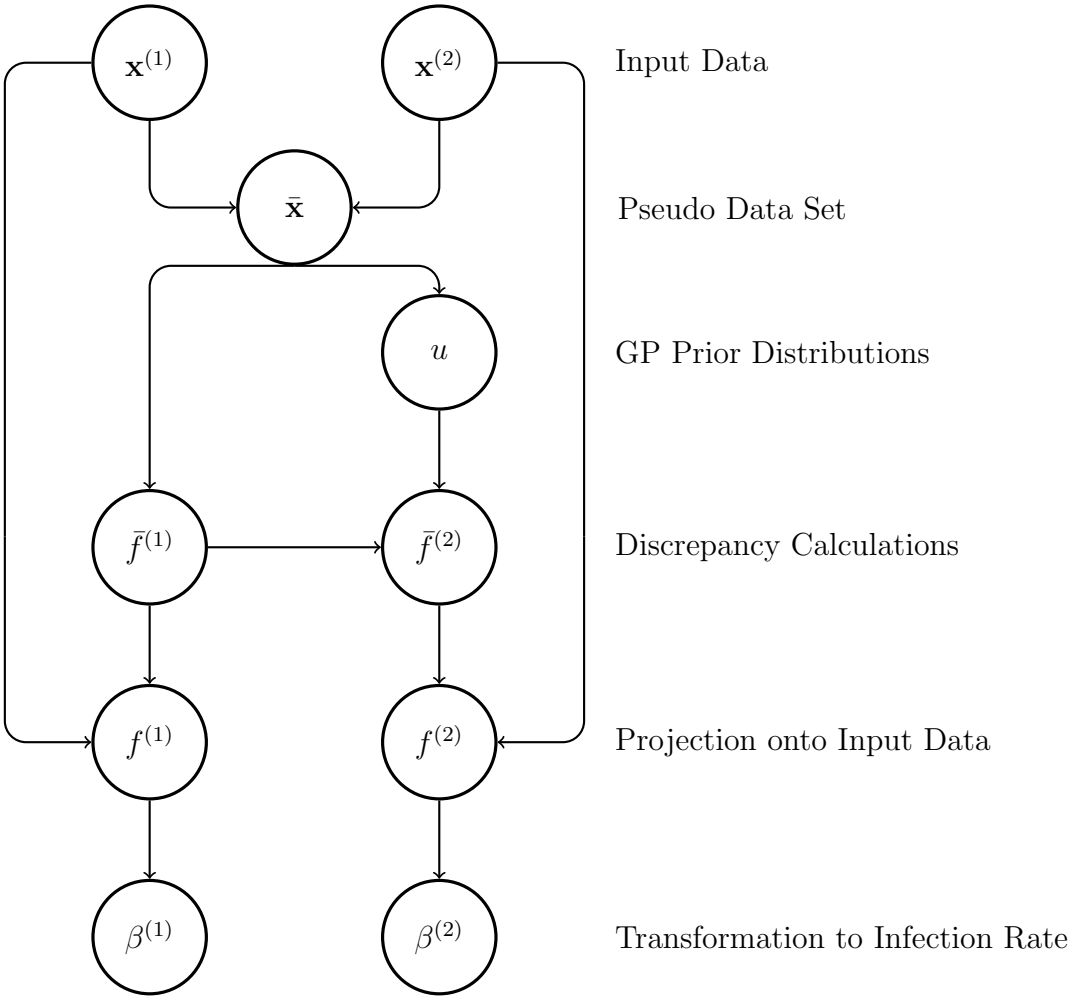


Figure 3.11: The DBM framework for an epidemic model with two types of individuals.

We construct the MCMC algorithm as before and again use a Metropolis-Hastings

algorithm with an underrelaxed proposal mechanism to target the infection rate functions. The full algorithm is shown in algorithm 9.

Algorithm 9 Structure of the DBM MCMC algorithm

1: Initialise the chain with values $\gamma^{(0)}$, $\left(f^{(1)}\right)^{(0)}$, $\left(f^{(2)}\right)^{(0)}$, $l_1^{(0)}$, $l_2^{(0)}$, and $\mathbf{i}^{(0)}$

Repeat the following steps

2: Sample γ from the conditional distribution $\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r})$ using a Gibbs step

3: Sample $f^{(1)}$ using an underrelaxed proposal method for a Metropolis Hastings step

4: Sample $f^{(2)}$ based on $f^{(1)}$ using an underrelaxed proposal method for a Metropolis Hastings step

5: Sample l_1 using a Metropolis Hasting Random Walk step

6: Sample l_2 using a Metropolis Hastings Random Walk step

7: Sample κ and i_κ using a Metropolis Hastings step

8: Update an infection time

3.5.4 MOGP Models for more than two Types

It is possible to use all three methods to model more than two infection rate functions, but some additional modelling choices need to be made. Consider the case where a population consists of three types of individuals, types one, two and three, with respective infection rate functions $\beta^{(1)}$, $\beta^{(2)}$, and $\beta^{(3)}$. The independent GP model is the most simple to construct, we simply place an independent GP prior distribution on the infection rate for each of the types, as follows:

$$\beta^{(j)} = g(f^{(j)}), \quad f^{(j)} \sim \mathcal{GP}(0, \Sigma^{(j)}), \quad \Sigma^{(j)}_{m,n} = k(x^{(j)}_m, x^{(j)}_n; \alpha^2, l^2_j).$$

To implement the MOC model for three types, we can extend the prior distribution given in equation (3.12) to:

$$\begin{pmatrix} f^{(1)} \\ f^{(2)} \\ f^{(3)} \end{pmatrix} \sim \mathcal{GP}\left(0,\ \begin{pmatrix} \Sigma^{(1,1)} & \rho_{1,2}\Sigma^{(1,2)} & \rho_{1,3}\Sigma^{(1,3)} \\ \rho_{1,2}\Sigma^{(2,1)} & \Sigma^{(2,2)} & \rho_{2,3}\Sigma^{(2,3)} \\ \rho_{1,3}\Sigma^{(3,1)} & \rho_{2,3}\Sigma^{(3,2)} & \Sigma^{(3,3)} \end{pmatrix}\right)$$

$$\beta^{(1)} = g(f^{(1)}),\ \beta^{(2)} = g(f^{(2)}),\ \beta^{(3)} = g(f^{(3)}).$$

By including a third type in the population, we have increased the number of correlation parameters from one to three. In general, for a population of m types, we will need $\frac{m(m-1)}{2}$ correlation parameters. Given sufficient data, we can place uniform prior distributions on each of the correlation parameters and infer plausible values. In cases where we have insufficient data, we can make several assumptions. The first is to assume that the correlation parameters are equal, that is $\rho_{i,j} = \rho$ for all i and j. This reduces the number of parameters in the model, while still allowing flexibility and transferring information across the prior distribution. The second assumption we can make is that the infection rate functions for certain types are uncorrelated, that is for some i and j, $\rho_{i,j} = 0$. This assumption reduces the number of parameters in the model, but may make learning the functions more difficult. It is also dependent on the context of the outbreak, where this *a priori* assumption may be suitable, if, for example, we expect the infection rate functions between types i and j not to have any particular relationship.

There are two ways in which we can modify the DBM in equation (3.5.4) to allow for three types. In both cases we treat the infection rate function for type one individuals as a baseline. In the first method, we condition the infection rate function for type two individuals on that of type one individuals and the infection rate function for type three individuals on type two individuals. The prior distributions are given by:

$$\begin{aligned} &\bar{f}^{(1)} \sim \mathcal{GP}(0,\ \Sigma^{(1)}), \quad \Sigma^{(1)}_{j,k} = k(\bar{x}_j, \bar{x}_k;\ \alpha, l_1). \\ &\bar{f}^{(2)} = \bar{f}^{(1)} + u^{(2)}, \quad u^{(2)} \sim \mathcal{GP}(0,\ \Sigma^{(2)}), \quad \Sigma^{(2)}_{j,k} = k(\bar{x}_j, \bar{x}_k;\ \alpha, l_2). \\ &\bar{f}^{(3)} = \bar{f}^{(2)} + u^{(3)}, \quad u^{(3)} \sim \mathcal{GP}(0,\ \Sigma^{(3)}), \quad \Sigma^{(3)}_{j,k} = k(\bar{x}_j, \bar{x}_k;\ \alpha, l_3). \end{aligned}$$

In this nested model, each type builds on all of the previous types. It is suitably appropriate when the types are contained within each other. For example, when measurements are taken over kilometres, metres, and centimetres. In an epidemic setting, an example of this may be when we are modelling the infection rate between adults, children under the age of 15, and children under the age of 5. However, for some contexts this may not be the most suitable model. We can instead choose type one to be a baseline type and model the other types relative to the baseline type. The prior distributions are given by:

$$\begin{aligned}
&\bar{f}^{(1)} \sim \mathcal{GP}(0,\, \Sigma^{(1)}), \quad \Sigma^{(1)}_{j,k} = k(\bar{x}_j, \bar{x}_k;\, \alpha^2, l_1^2).\\
&\bar{f}^{(2)} = \bar{f}^{(1)} + u^{(2)}, \quad u^{(2)} \sim \mathcal{GP}(0,\, \Sigma^{(2)}), \quad \Sigma^{(2)}_{j,k} = k(\bar{x}_j, \bar{x}_k;\, \alpha^2, l_2^2).\\
&\bar{f}^{(3)} = \bar{f}^{(1)} + u^{(3)}, \quad u^{(3)} \sim \mathcal{GP}(0,\, \Sigma^{(3)}), \quad \Sigma^{(3)}_{j,k} = k(\bar{x}_j, \bar{x}_k;\, \alpha^2, l_3^2).
\end{aligned}$$

This model construction bears some similarity to the fixed effects model where we choose a type to be the baseline type and compare all other functions to the baseline function.

There are several ways to construct the GP models for modelling more than two infection rate functions. The construction of the model will ultimately be context dependent, as well as depending on how much data there is. It situations where there is little data, we will need to make stronger assumptions and design the prior distribution accordingly. Depending on how the results will be used, we may need to structure the model so that it is easily interpretable by practitioners and non-mathematicians.

3.5.5 Simulation Studies

We now apply these models to simulated data to see how well we can recover the model parameters and infection times. We generate the positions of 1,000 individuals on a unit square, and choose 500 uniformly at random to be labeled type zero, with the remaining individuals labeled type one. We then simulate 250 outbreaks of an infectious disease, conditioned on the final size being greater than 100, using the infection rate functions shown in equation (3.14) and model parameters shown in

table 3.4.

$$\beta_{i,j} = \begin{cases} \beta_0 \exp\{-\mu_0 d_{i,j}\} & \text{if individual } j \text{ is type 0} \\ \beta_1 \exp\{-\mu_1 d_{i,j}\} & \text{if individual } j \text{ is type 1.} \end{cases} \tag{3.14}$$

These functions have been chosen so that although the functions are different between the two types, they still share some similarities. The parameter values were chosen so that approximately half of the individuals become infected in serious outbreaks. The values for the infectious period distributions give a mean period of two days and a standard deviation of approximately 0.8 days.

3.5.5.1 Independent GP Model

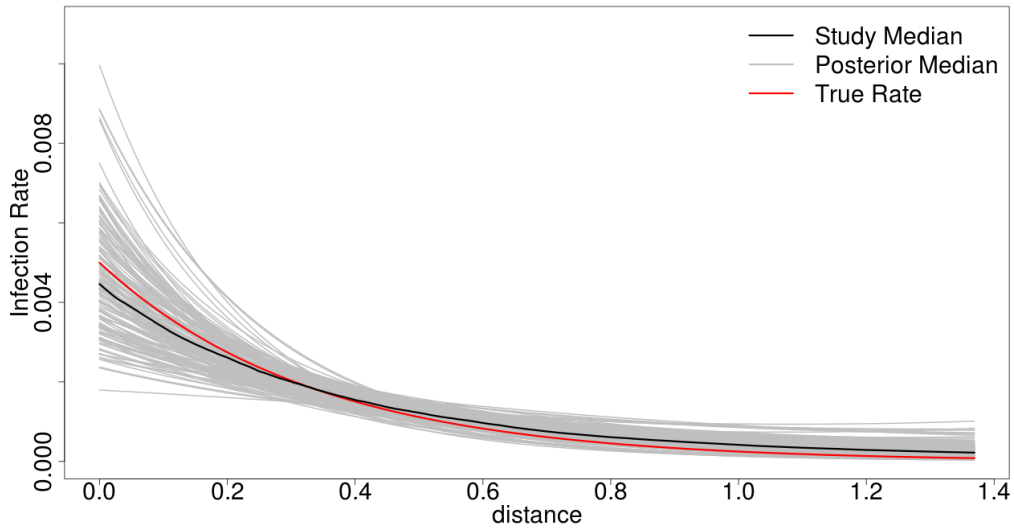
We first implement the Independent GP model. The results for the infection rate functions are shown in figure 3.12 and for the remaining model parameters in figure 3.13. The infection rate functions are inferred well, with the largest errors for the smallest and largest pair-wise distances. The infection rate for type one individuals is inferred more successfully than for type two. This is because type one individuals were five times more susceptible to the disease than type two individuals, therefore more type one individuals were infected. This gives the model more information to learn from on type one individuals. The numerical estimates are shown in table 3.4 and we can see γ is slightly overestimated, but the true value is still contained in the 95% credible interval.

3.5.5.2 Multi-Output Covariance Model

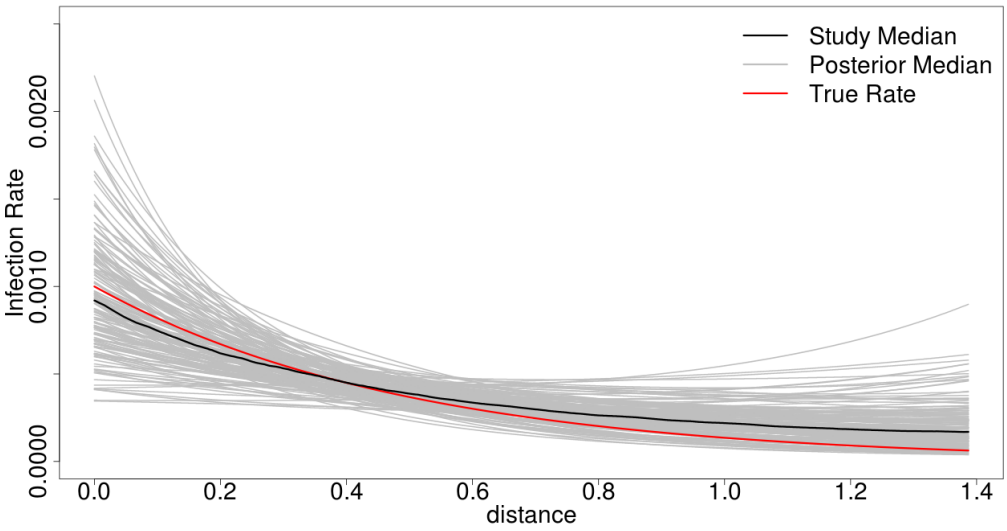
We implement the MOC model MCMC algorithm shown in algorithm 8 and infer the two infection rate functions: the infection period distribution rate parameter and the infection times, given the removal times. For this example we assume the infectious period distribution shape parameter is known, this is so that we can focus on inferring the values of the other parameters. We also fix the GP prior distribution length scale, this is because performing inference for the length scale as well as the covariance parameter can lead to numerical instability in the covariance matrix and its determinant.

Model	Parameter	Study Median	95% Credible Interval
Parameter Values	β_0	0.005	-
	β_1	0.001	-
	α	6	-
	γ	3	-
	$\tilde{i}$	0	-
Independent GPs	β_0	0.00446	(0.00257, 0.00859)
	β_1	0.000920	(0.00415, 0.00180)
	γ	3.13	(2.41, 3.92)
	$\tilde{i}$	-0.0111	(-0.0791 ,0.0470)
MOC Model	β_0	0.00484	(0.00273, 0.00782)
	β_1	0.00113	(0.000644, 0.00191)
	γ	3.07	(2.39, 3.89)
	$\tilde{i}$	-0.00757	(-0.0839, 0.0514)
	ρ	0.762	(0.495, 0.856)
DBM	β_0	0.00430	(0.00223, 0.0808)
	β_1	0.00126	(0.000562, 0.00250)
	γ	3.11	(2.43, 4.02)
	$\tilde{i}$	-0.00989	(-0.102, 0.0505)
	l_1	5.05	(2.49, 10.7)
	l_2	6.87	(2.14, 16.3)

Table 3.4: The median and 95% credible interval for the model parameters using the MOC model and the DBM, compared to the true model parameters.

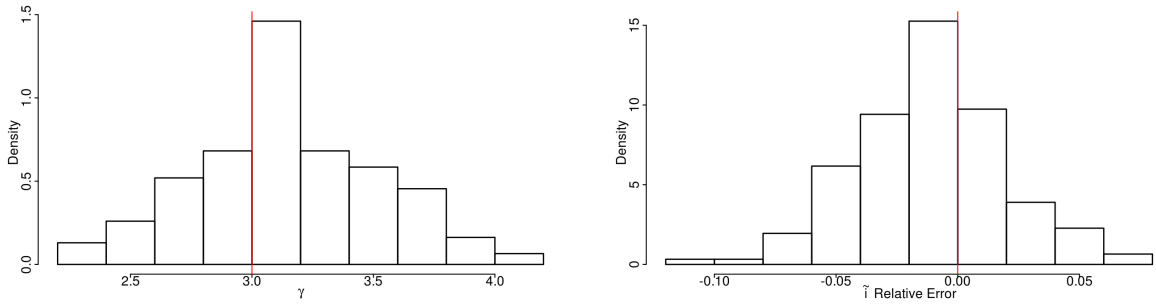


(a) Estimates for the infection rate functions for type 0 individuals.



(b) Estimates for the infection rate functions for type one individuals.

Figure 3.12: The inferred infection rate functions for the independent GP model. The grey lines are the posterior median infection rate functions for each of the 250 data sets. The black line is the median of all 250 infection rate functions and the red line is the true infection rate function.



(a) The distribution of the median estimates for the 250 estimates of γ.

(b) The distribution of the relative error in the sum of the infection times over the 250 data sets.

Figure 3.13: The results of the independent GP model applied to the 250 simulated data sets.

In figure 3.14, we show the posterior median infection rate functions for each of the 250 data sets, the median of the 250 rates, along with the rate used to simulate the data. We can see that the rates for both types are recovered well and almost all of the individual posterior medians are close to the true rate. There are two data sets which yield infection rate functions which are much higher than the true rate, which is due to there being a very high number of infected individuals in these two simulations. In figure 3.15 and table 3.4, we can see that the remaining parameters are inferred well. On average, the relative error in the sum of the infection times is less than 1% and the infectious period distribution rate parameter is inferred almost exactly. The inferred value of the covariance parameter is 0.762 and the 95% credible interval gives some evidence that the functions are similar but not identical.

This simulation shows that the MOC model is suitable for modelling the outbreaks of infectious diseases where their infection rate is both heterogenous and depends on the type of the susceptible individual. The model works well when inferring the infection rate functions alongside other model parameters, such as the infectious period distribution rate parameter, and gives us information on how the infection rate functions are related. Compared to the independent GP model, the variability in the results is less. This is due to an increased number of assumptions in the model, but these are only weak assumptions, and we are able to learn plausible values for

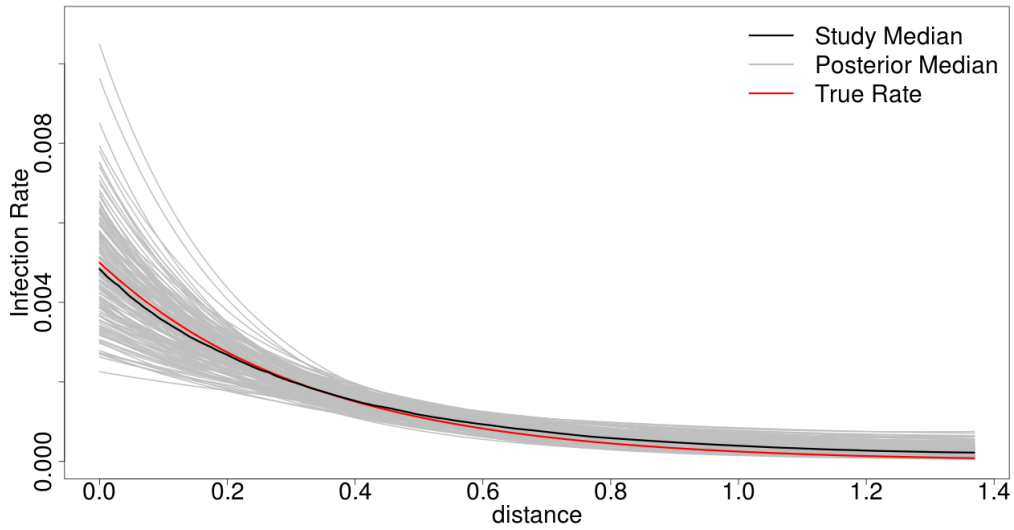
parameters related to these assumptions.

3.5.5.3 Discrepancy Based Model

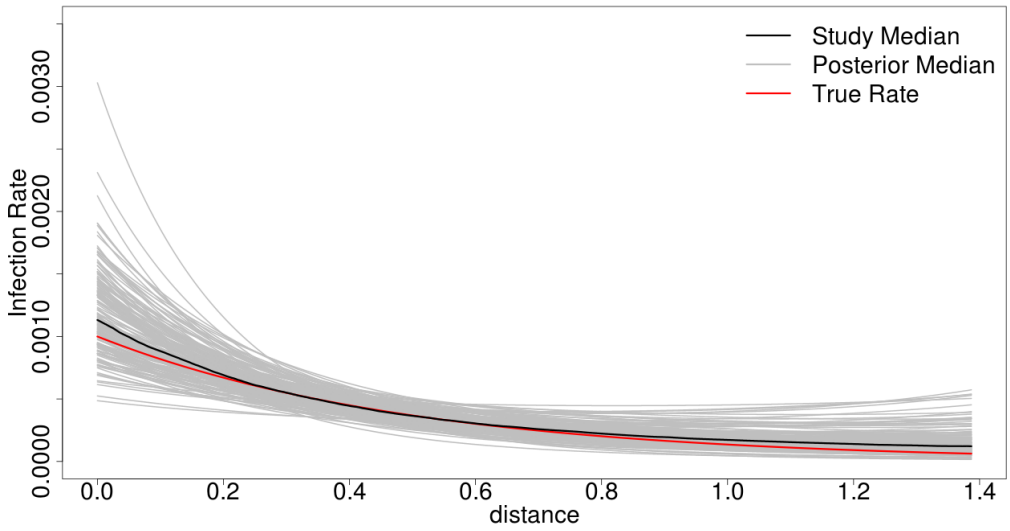
We now repeat the simulation study, except using the DBM outlined in equation (3.5.4). Again, we wish to infer the infection rate functions, infection times and infection period distribution rate parameter. Instead of inferring the covariance between the two rates as in the MOC model, we infer the difference between the two functions.

The results for the infection rate functions are shown in figure 3.16 and we see we can infer both rates well. In a small number of simulations, the estimates for the infection rate functions over short distances are overestimated, and as the infection rate for type one individuals is overestimated, the infection rate for type two is also overestimated. The inference for the infection rate for type two individuals was slightly more accurate the for type one individuals, this is because this function also has information for type one individuals. Figure 3.17 shows the ratio of the functions, which we consider instead of the difference as the GPs are on the log scale. A ratio of 1 implies that the rates are identical and we can see that in only a few simulations is this inferred and this is for large pair-wise distances, so we can correctly conclude that the infection rate functions for the two types are different across the whole range of distances. The study median of all the ratios is decidedly different from the true ratio in the later part of the input region and the study medians are spread across a large range. This is because the infection rate functions asymptotically tend to 0, and a small variation in values near 0 can lead to ratio near 1. For example that relative error in $\beta^{(1)} = 10^{-7}$ and $\beta^{(2)} = 10^{-8}$ is large, but compared to the scale of the functions, the absolute error is small. This plot has a practical use as we can quickly identify that about a third as many type two individuals will be infected per day over short distances as type one individuals, suggesting that control measures should aimed at type one individuals.

With regard to the results of the other model parameters, which are shown in figure 3.18, using the DBM has little impact on the accuracy of the results and we

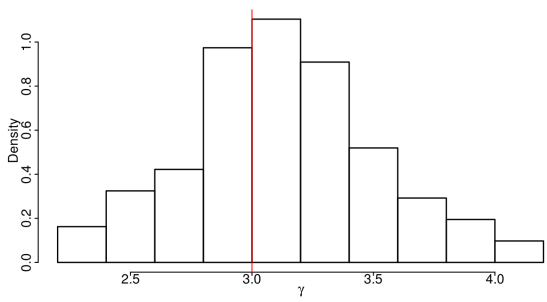


(a) Estimates for the type 0 infection rate function.

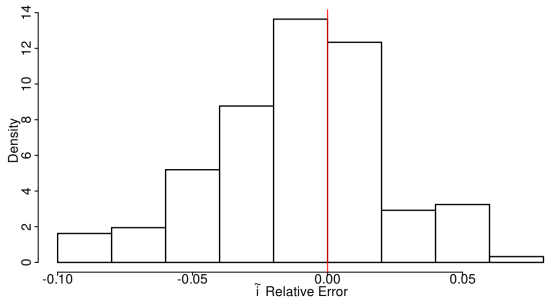


(b) Estimates for the type one infection rate function.

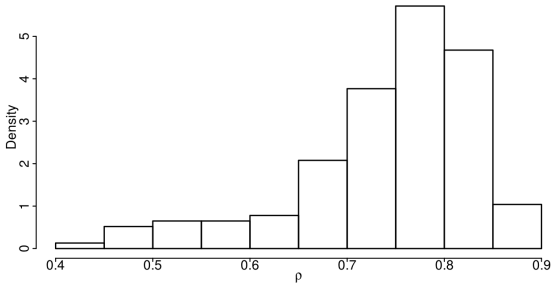
Figure 3.14: The inferred infection rate functions for the MOC model for the simulation study. The grey lines are the posterior median infection rate functions for each of the 250 data sets. The black line is the median of all 250 infection rate functions, and the red line is the true infection rate.



(a) The distribution of the median estimates for the 250 estimates for γ.



(b) The distribution of the relative error in the sum of the infection times over the 250 data sets.



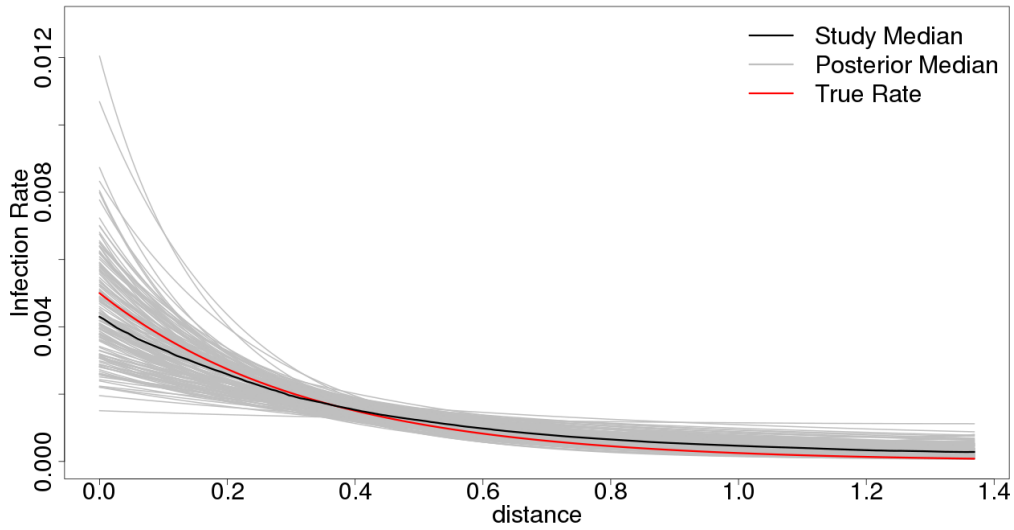
(c) The distribution of the 250 median estimates for the covariance parameter ρ.

Figure 3.15: The results of the MOC model applied to the 250 simulated data sets.

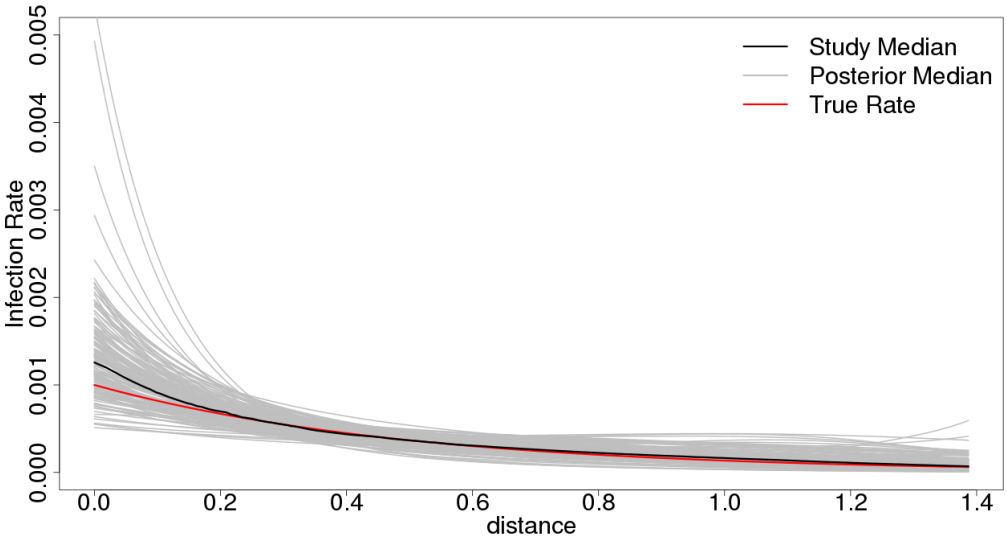
infer the values well and have a similar sized credible interval to that of the MOC model. The numerical results for the parameters are shown in table 3.4.

3.5.5.4 Remarks on the Simulation Studies

We now compare the results of the three simulation studies. The study medians and 95% confidence intervals are shown in figure 3.19. The independent GP model and the DBM are almost identical as they share the same prior structure for the type one infection rate. The MOC model does not give a significantly different result to the other two models and all models give good results across the entire domain. The results for the second infection rate are where we can see larger differences between the models. The 95% credible interval for the DBM is has a higher upper bound than for the other models, which is an artefact of the higher infection rate for type one



(a) Estimates for the type 0 infection rate function.



(b) Estimates for the type one infection rate function.

Figure 3.16: The inferred infection rate functions for the DBM model for the simulation study. The grey lines are the posterior median infection rate functions for each of the 250 data sets. The black line is the median of all 250 infection rate functions, and the red line is the true infection rate.

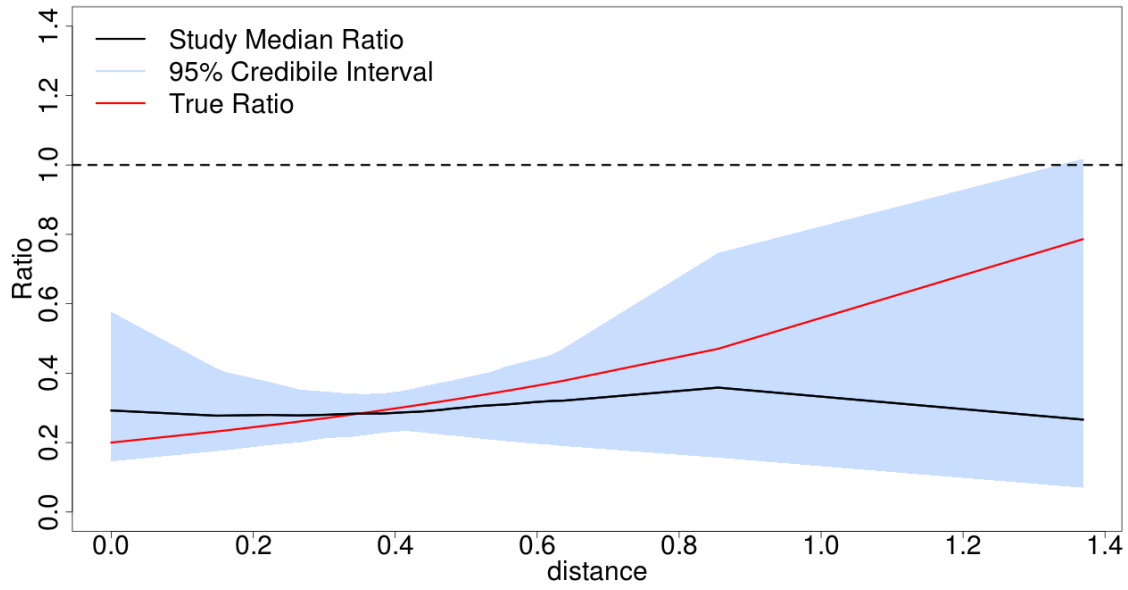


Figure 3.17: The ratio of the two infection rate functions for the DBM.

individuals. However the DBM has a narrower credible interval for large distances, again because it values are based on the infection rate for type one individuals. The MOC model has the smallest credible interval, which is because the information for both types is shared in the GP prior distribution.

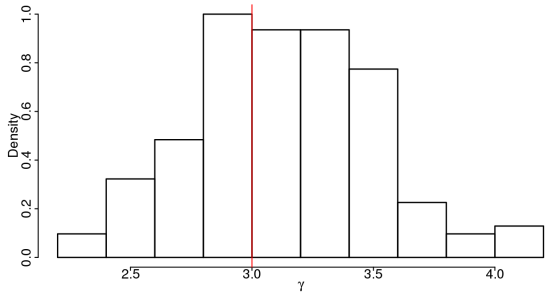
For a numerical comparison, we define the maximum absolute error for individual type t by

$$\mathrm{AE}^{(t)} = \max_{j,k} |\beta^{(\tau)}(x_{j,k}) - \hat{\beta}^{(\tau)}(x_{j,k})|$$

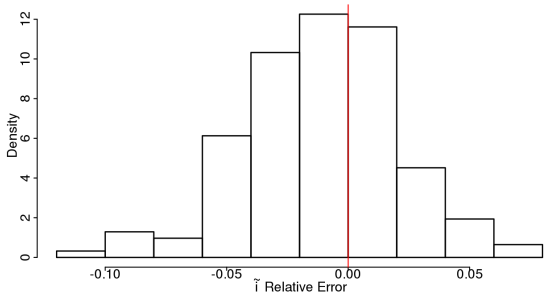
where $\hat{\beta}^{(t)}(x_{j,k})$ is the posterior median of the model in question. These values are given in table 3.5, and show that the MOC model performs best for the type one infection rate and both the MOC and independent GP models perform well for the type two infection rate.

3.6 Conclusion

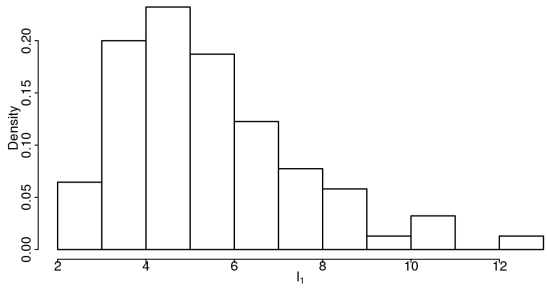
In this chapter we have extended the GP model outlined in chapter two. This has allowed us to include more information in our model and allow for more variation in



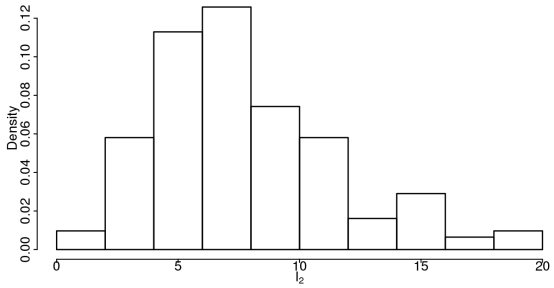
(a) The distribution of the median estimates for γ.



(b) The distribution of the relative error in the sum of the infection times over the 250 data sets.



(c) The distribution of the 250 median estimates for the length scale of function 1.

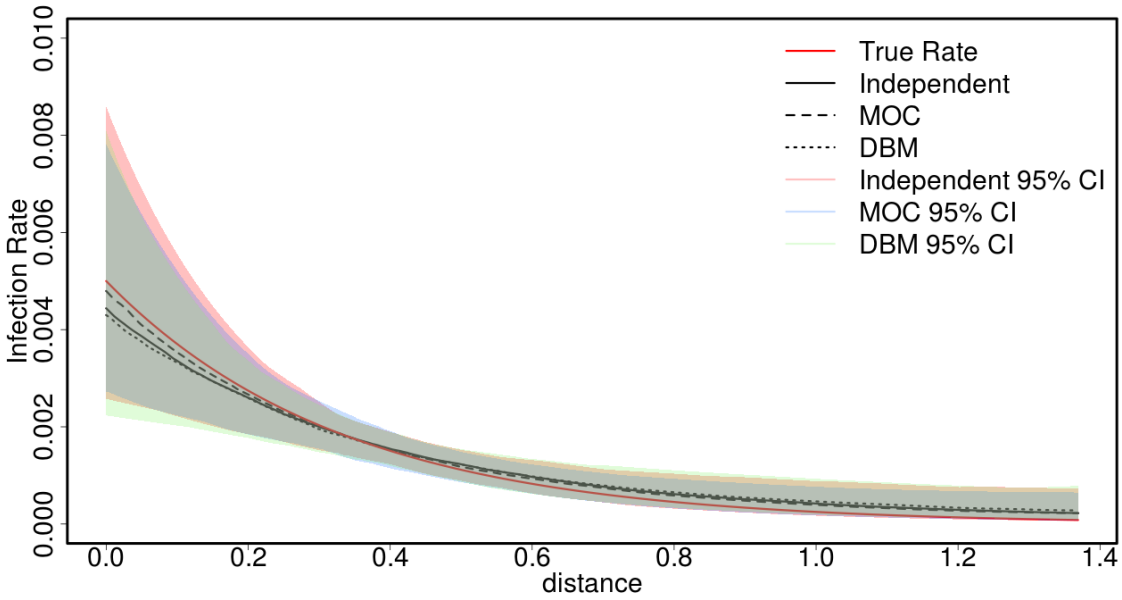


(d) The distribution of the 250 median estimates for the length scale of function 2.

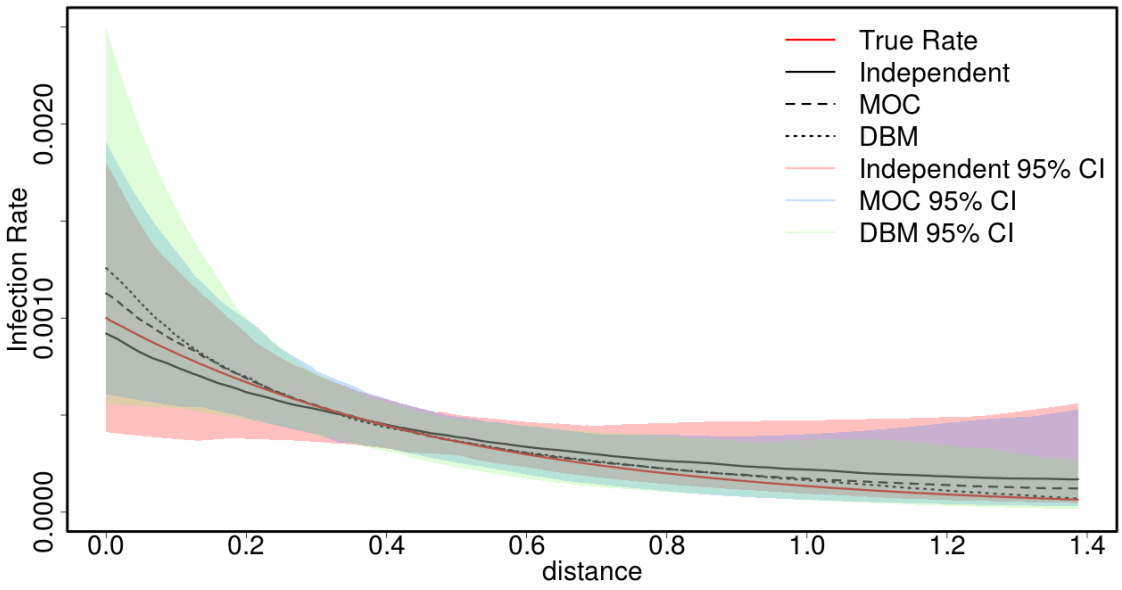
Figure 3.18: The results of the DBM applied to the 250 simulated data sets.

Model	Error for type one ($\times 10^{-4}$)	Error for type two ($\times 10^{-4}$)
Independent GPs	5.4	1.04
MOC	1.86	1.32
DBM	6.94	2.70

Table 3.5: Maximum absolute error for the estimates for type one and type two infection rate functions.



(a) Estimates for the type 0 infection rate function.



(b) Estimates for the type one infection rate function.

Figure 3.19: The median of the 250 estimates for the infection rate functions functions under each model compared to the true rate for the simulated data.

the spread of the disease. The frameworks we have developed maintain our aim of inferring infection rate functions without making strict parametric assumptions, or

choices which lack any real biological or epidemiological justification. The assumptions that we described in this chapter are mild and in many applications the structures we have outlined can be justified. This is in comparison to parametric methods, where the choice of kernel is often arbitrary and difficult to justify. Our methods have been successfully used alongside methods to infer the other model parameters and the missing infection times.

The first extension was to allow for our individual-level model to include information about the type of the individuals. We did this by assuming the infection rate depended on the susceptible individual's type as well as the relationship between the infector and susceptible individual. We outlined four models: a fixed effects model, an independent GP model, a multi-output covariance model, and a discrepancy based model. The fixed effects model assumed the shape of the infection rate was shared between all types, and the infection rate only differed by scale. This model is an extension of the work in the previous chapter and gives good results. This assumption is restrictive as we often have no prior beliefs or evidence that the shape of the infection rate is the same over different types. To allow for this, we introduced the multi-output GPs and the remaining three models. Multi-output GPs allow us to share the learning across types, that is we believe *a priori* that there is some shared structure between the infection rate functions for individuals of different types. This is a less restrictive assumption than for the fixed effects model. The three models differ in how this structure is described. The Multi-Output Covariance model assumes the infection rate functions for individual types are correlated, whereas the discrepancy based model assumes the infection rate for each type is based on the infection rate for the previous type. In the Independent GP model, we assume the infection rate for each type is independent of any other type.

These models share two drawbacks: identifiability and a requirement for a large amount of data. In the fixed effects model, we are unable to fully identify the scale parameters, merely the ratio between them. In the multi-output GP models, we have difficulties identifying the length scale parameters, even in data rich situations. We need to resort to fixing at least one length scale and inferring the remaining parameters.

As all the other models require more than one GP, they require significantly more data. The multi-output covariance model and the discrepancy based model make full use of the data as information can be shared between GPs. In situations where we have a lack of data, we may need to included either weak or strong assumptions in the model. Examples of weak assumptions include the monotonicity framework described in section 2.7.1 and the conditional framework in section 2.6.2. When making stronger assumptions, we would explicitly break the Bayesian nonparametric framework, for example, by including a mean function based on a parametric model and reducing the variance of the GP prior distribution. Despite these potential weaknesses, the simulation studies have shown that these methods yield good results and can be successfully implemented.

The main difference between the MOC model and DBM is their intended audience. From a mathematical viewpoint, being able to characterise the covariance between two functions is useful and the MOC framework allows us to do this. It also allows us to distill the relationship between the types into one parameter, ρ. The correlation between rates however has less practical interpretation. Practitioners are more likely to be interested in direct comparisons of the infection rate functions than correlation parameters, as well as how this can be implemented in diseases control measures. The DBM allows us to directly compare the infection rate functions, and see how different types are affected. However, if we are interested in comparing the rates over a small range, we risk not having enough data to draw significant conclusions. The independent model allows us to do neither of these and does not use the data to the greatest effect. The fixed effects model gives us insight to the data, but we run the risk of making false assumptions about the infection rate. Given that the DBM and MOC model contain the independent GP model and the fixed effects model, these are superior models, and we recommend their use over the independent models.

CHAPTER 4

Bayesian Nonparametric Methods for Heterogeneous Infection Rate Functions with Multiple Covariates

4.1 Introduction

So far, we have used our nonparametric methodology to infer the infection rate functions which depend on a single continuous covariate. We now consider cases where we allow the infection rate from individual i to j to depend on two continuous covariates. Examples of this include an outbreak of Foot and Mouth disease where the infection rate function may depend on the distance between any two farms and the number of animals on them. The infection rate from individual i to j can be written as

$$\beta_{ij} = f(x_{1,i,j}, x_{2,i,j}, \ldots, x_{m,i,j}),$$

where $x_{1,i,j}, x_{2,i,j}, \ldots, x_{m,i,j}$ are covariates describing the relationship between individuals i and j. As in previous chapters, we assume individual i makes infectious contact with individual j at the time points of a Poisson process with rate $\beta_{i,j}$.

Like infection rate functions of one covariate, modelling infection rate functions

with multiple covariates has typically been done parametrically. A modeller will propose a parametric form for f and then estimate the model parameters from the data. In Kypraios (2007), the author proposed a parametric infection rate function for an outbreak of Foot and Mouth disease that depends on the Euclidean distance between farms as well as the number of sheep and cattle on each farm. We now recall this infection rate function from chapter one:

$$\beta_{i,j} = (\beta_1(n_i^c)^{\beta_2} + (n_i^s)^{\beta_2}) \cdot (\beta_3(n_j^c)^{\beta_2} + (n_j^s)^{\beta_2}) \cdot \frac{\beta_4}{d_{i,j}^2 + \beta_5}.$$

Here, n_i^c and n_i^s and the number of cattle and sheep on the i^{th} farm, $d_{i,j}$ is the Euclidean distance between farm i and j, and $\beta_1, \ldots, \beta_5$ are parameters controlling the infection rate. This was then analysed in Jewell et al. (2009) and Stockdale (2019). However, we argue that the choice of parametric form used to model the infection rate is arbitrary and lacks justification from the data. Instead we develop a Bayesian nonparametric framework which allows us to learn the infection rate function from the data.

There are several ways of modelling a function of two variables in a nonparametric way, and this is more challenging than modelling a function over one variable. We will use Multi-Input GP prior distributions to do this. This extends the framework from chapter 2 to allow for multiple inputs. Instead of modelling the infection rate as a function in one dimension, we now consider it a curve in n dimensions, where the number of dimensions is the number of covariates we input into the model. One Multi-input GP method extends the covariance functions to allow for multiple inputs (Rasmussen and Williams, 2006, §4). This is also known as a Gaussian Random Field (Abrahamsen, 1997). A commonly implemented method is the Gaussian Markov Random Field (Rue and Held, 2005) where the prior distribution is placed on a network and a dependence structure is created between the nodes.

We develop two methods for modelling multi-dimensional infection rate functions. The first method is an additive method where we consider the covariates separately and model the infection rate as a sum of n distinct functions. Our second method is a coupled model where we model the infection rate as a function of all the covariates

and use the Multi-Input GP method.

4.1.1 Layout of the Chapter

Before we outline our methods for modelling two-dimensional infection rate functions in epidemic models, we introduce two covariance functions: the multi-dimensional squared exponential covariance function and the linear covariance function. We demonstrate the use of each covariance function with a regression example. We then develop two methods for modelling infection rate functions using Bayesian nonparametric methods based on the covariance functions in section 4.3 before comparing the methods in section 4.4. We finish the chapter by demonstrating our methods in a simulation study in section 4.5.

This chapter extends the methods developed in chapter 2 to allow for multi-covariate functions to be modelled; for example, when the infection rate function depends on size and distance. Given the limited amount of data observed, this is challenging, so we allow for some assumptions to be made to include semi-parametric methods.

4.2 Further Covariance Functions

In previous chapters, we have used the squared exponential function:

$$k(x_i, x_j;\, \alpha, l) = \alpha^2 \exp\Big\{ -\frac{(x_i - x_j)^2}{l^2} \Big\}. \tag{1.2}$$

This can be used to model one dimensional functions which are infinitely differentiable, and in this chapter we will extend this covariance function into n dimensions. We will then introduce the linear covariance function, which is used to model one-dimensional linear functions, full descriptions of which can be found in Rasmussen and Williams (2006, §4). We demonstrate the use of both covariance functions by way of a basic regression example.

4.2.1 Multi-Dimensional Squared Exponential Covariance Function

We now introduce a covariance function which can take a multi-dimensional input. When modelling a function with a single input, the squared exponential covariance function depends on $(x_i - x_j)^2$, the squared distance between the two input points. In multiple dimensions, the equivalent is given by $(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)$. The multi-input squared exponential covariance function is therefore given by:

$$k(\mathbf{x}_i, \mathbf{x}_j; \alpha, l) = \alpha^2 \exp\left\{-(\mathbf{x}_i - \mathbf{x}_j)^T \Lambda (\mathbf{x}_i - \mathbf{x}_j)\right\}, \tag{4.1}$$

where α is the signal variance parameter and Λ is the length scale matrix. There are several options for Λ, the simplest being $\Lambda = l\mathrm{I}$, where I is the identity matrix. In this case we assume the length scale in the same in both dimensions and is given by l. If we expect the length scales to differ, we can construct the vector of length scales $\mathbf{l} = \{l_1, \ldots, l_m\}$ and set $\Lambda = \mathrm{diag}(\mathbf{l})^{-2}$. We can also set Λ to have non-zero off-diagonal elements, which represents an interaction between the terms. When designing Λ, we need to ensure the resulting covariance matrix is a valid symmetric and positive semi-definite covariance matrix.

To choose the most appropriate covariance structure, we consider the effect of the covariance function geometrically. As we are demonstrating this geometrically, we assume the input space is two-dimensional. However this can be extended to any number of dimensions. Consider the point $\mathbf{x}_i = (x_{i,1}, x_{i,2})$. With the length scale matrix $\Lambda = l\mathrm{I}_2$, the covariance is the same between $\mathbf{x}_i$ and any point on the circle centred at $\mathbf{x}_i$ with radius l. For the length scale matrix $\Lambda = \mathrm{diag}(l_1, l_2)^{-2}$, the covariance between $\mathbf{x}_i$ and any point on the ellipse centred at $\mathbf{x}_i$ with semi-major axis l_1 and semi-minor axis l_2 is the same. Allowing for off-diagonal terms rotates the ellipse. In Rasmussen and Williams (2006, §4), the authors consider the matrix $\Lambda = AA^T + \mathrm{diag}(l_1, l_2)^{-2}$, where A is a vector giving the direction of greatest variation.

The choice of covariance structure depends on the modelling scenario. For example, in an outbreak of Avian Influenza we may want to consider the direction of the prevailing wind or the direction in which migratory birds travel. To do this

we would include off-diagonal terms in the length scale matrix, maximising the covariance in the direction of interest. This allows us to include some biological and epidemiological assumptions in the model without the need for strict parametric assumptions.

Figure 4.1 shows two samples from a zero-mean GP using the multi-input covariance function. We used two input sets $\mathbf{x} = \{0, 0.1, 0.2, \ldots, 5\}$, $\mathbf{y} = \{0, 0.1, 0.2, \ldots, 6\}$. We set the signal variance parameter, α, to be 3 and the length scale matrix to be $\Lambda = \text{diag}(l_1, l_2)^{-2}$. In figure 4.1(a) both length scales are the same, whereas in 4.1(b) the length scale for the y dimension is six times as long as that for the x dimension. This results in the function moving much more quickly over the x domain than the y domain. For example, taking either a horizontal or vertical slice in figure 4.1(a), we note the values do not vary considerably. This is the same in figure 4.1(b) for horizontal slices, as the length scale is also 6. However when taking vertical slices, we see much larger variability, as the length scale has been reduced.

4.2.1.1 Two-Dimensional Regression

We now outline the coupled model in a regression setting. We wish to infer the function f in the following model:

$$z_{i,j} = f(x_i, y_j) + \varepsilon_{i,j}, \quad \varepsilon_{i,j} \overset{i.i.d}{\sim} N(0, \sigma^2).$$

We denote $\mathbf{x} = \{x_1, \ldots, x_n\}$, $\mathbf{y} = \{y_1, \ldots, y_m\}$ and $\mathbf{z} = \{z_{1,1}, \ldots, z_{n,m}\}$. The likelihood function for this model is given by:

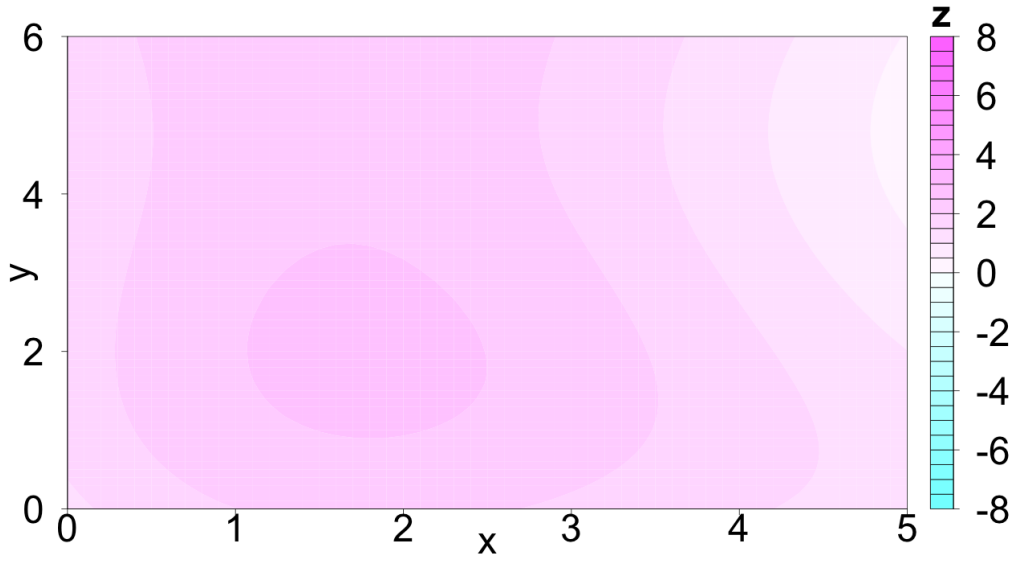
$$\pi(\mathbf{z}|f, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}(z_{i,j} - f(x_i, y_j))^2\right\}.$$

We model the function f by placing a GP prior distribution on it as follows

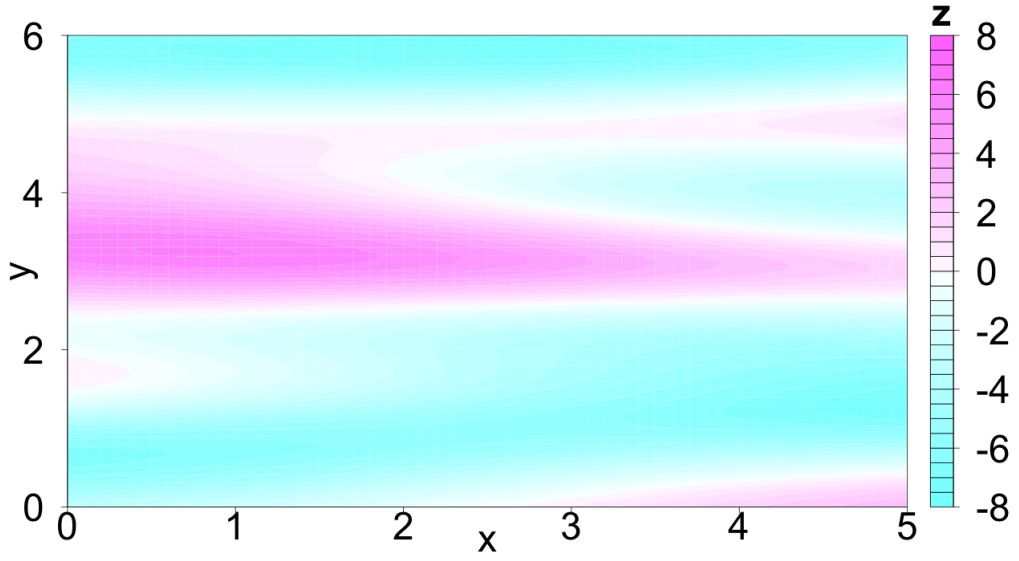
$$f \sim \mathcal{GP}(0,\ \Sigma), \quad \Sigma_{i,j} = k\left((x_i, y_i), (x_j, y_j);\ \alpha, \Lambda\right),$$

where $\Lambda = \text{diag}(l_x, l_y)^{-2}$. As this is a toy example, we assume that σ^2 is known meaning we can write down the posterior distribution as follows:

$$\pi(f|\mathbf{z}, \sigma^2, \alpha, l_x, l_y) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}(z_{i,j} - f(x_i, y_j))^2\right\}\mathcal{GP}(f;\ 0, \Sigma).$$



(a) A sample from a GP with length scales (6, 6).



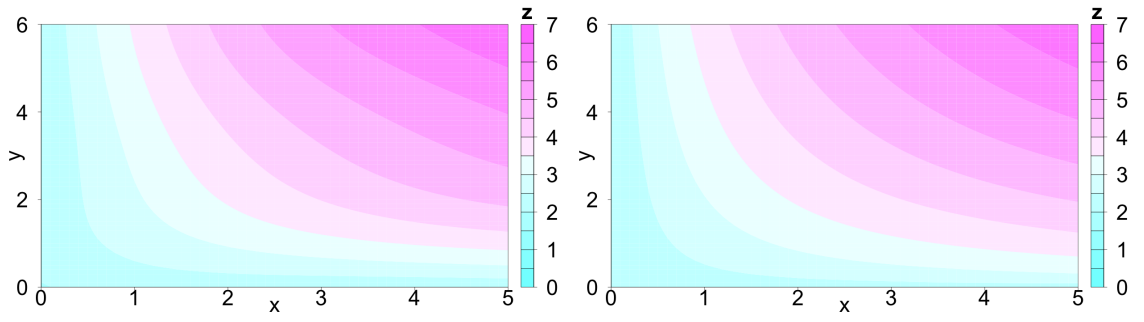
(b) A sample from a GP with length scales (6, 1)

Figure 4.1: Samples from a GP prior distribution with multi-input squared exponential covariance function in equation (4.1) and different length scale values. The signal variance parameter was set to $\alpha^2 = 9$.

To sample from this distribution, we construct an MCMC algorithm. We use the Metropolis-Hastings algorithm with an underrelaxed proposal mechanism to generate samples of the function f. To demonstrate this method, we use the following model:

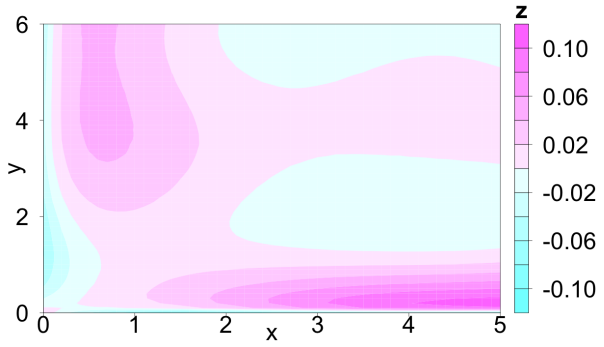
$$z_{i,j} = \log(x_i + 1)\sqrt{y_j} + 2 + \varepsilon_{i,j}, \quad \varepsilon_{i,j} \overset{i.i.d}{\sim} N(0, 0.2^2).$$

We use the input sets $\mathbf{x} = \{0, 0.1, \ldots, 5\}$, $\mathbf{y} = \{0, 0.1, \ldots, 6\}$ and fix the hyperparameters to be $\alpha = 4$, $l_x = 5$ and $l_y = 6$. These values were chosen based on preliminary runs of the algorithm; it is possible to learn plausible values for these parameters, but as this is a toy example, we choose not to. We sample from the posterior density 15,000 times, removing the first 100 samples as a burn-in period. The results are shown in figure 4.2 and show good results for this dataset. The relative largest errors are in the top left and bottom right corners of the surface, where both x and y values are small.



(a) The posterior mean surface.

(b) The true surface, $f(x, y) = \log(x + 1)\sqrt{y} + 2$.



(c) The relative error at each point.

Figure 4.2: The results of the 2D nonparametric regression example in section 4.2.1.1 with multi-input squared exponential covariance function.

Type	Input Scale	Output Scale	Functional Form
Linear	Linear	Linear	$f = px + q$
Exponential	Linear	Logarithmic	$f = Q\exp\{px\}$
Monomial	Logarithmic	Logarithmic	$f = Qx^p$

Table 4.1: The three possible functional forms for samples from a GP prior distribution with a linear covariance function.

4.2.2 The Linear Covariance Function

The linear covariance function takes a one-dimnesnion input and is given by:

$$k_{lin}(x_i, x_j; \alpha_0, \alpha_1, \alpha_2) = \alpha_0^2 + \alpha_1^2(x_i - \alpha_2)(x_j - \alpha_2), \tag{4.2}$$

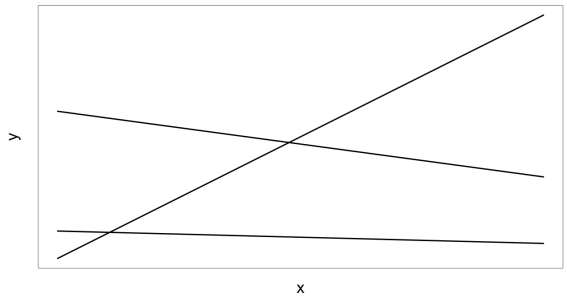
and it is sometimes referred to as the dot product covariance function. Placing a GP prior distribution on a function f using a linear covariance function results in samples taking the following form:

$$f \sim \mathcal{GP}(0, \Sigma), \quad \Sigma_{i,j} = k_{lin}(x_i, x_j; \alpha_0, \alpha_1, \alpha_2), \quad \Longrightarrow f = px + q, \quad p, q \in \mathbb{R}$$

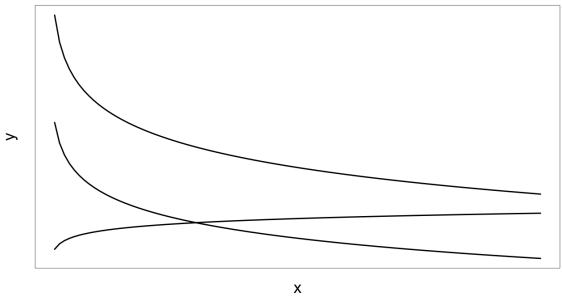
Although samples drawn from a GP prior distribution with this covariance function are linear, we are able to expand the type of functions we can model slightly by taking either the output alone or both the input and the output to be on a log scale. This fits in with our use of the function $g = \exp$, to ensure the function is non-negative. As we are using the exponential function, we can model monomial functions by taking both the input and the output on log scales, which is shown by:

$$\begin{aligned} \log f \sim \mathcal{GP}(0, \Sigma), \quad \Sigma_{i,j} = k_{lin}(\log x_i, \log x_j; \alpha_0, \alpha_1, \alpha_2) &\Longrightarrow \log f = p \log x + q \\ &\Longrightarrow f \propto x^p. \end{aligned}$$

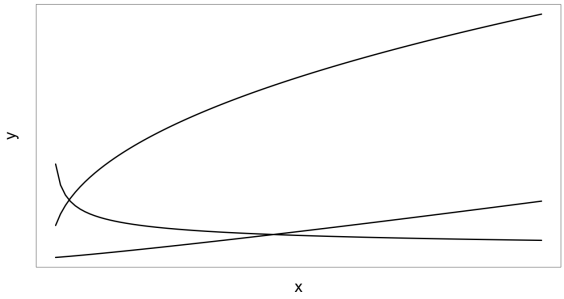
Similarly, by taking the output on a log scale, and the input on a natural scale, we can model exponential functions. Samples from all three combinations are shown in figure 4.3 and table 4.1. Although this covariance function can model three types of function, this choice must be made before the data is observed.



(a) Samples drawn on a linear scale.



(b) Samples drawn on a log-linear scale



(c) Samples drawn on a log-log scale

Figure 4.3: Three samples from a GP with a linear covariance function. The hyperparameters used were $\{\alpha_0, \alpha_1, \alpha_2\} = \{3, 1, 0\}$. Each of the three plots show the effect of different scales on the inputs and outputs.

The linear covariance function has three hyperparameters, $\alpha_0, \alpha_1, \alpha_2$. In this chapter, we will set the location parameter $\alpha_2 = \min \mathbf{x}$, which forces the smallest covariance value to be α_0^2. The final parameter is the scaling parameter α_1, which in part controls the maximum covariance. The linear covariance function is non-stationary and the variance increases as x increases. It is possible to perform inference for these hyperparameters. However, it can be challenging, and they do not have the same physical interpretation as the length scale in the squared exponential function. We recommend setting α_1 and α_2 with respect to input data such that GP prior distribution is vague, yet still contains information about the data points. That is α_1 should be large enough to increase the covariance of data points near zero, but not overpower the covariance of larger points. Similarly, α_2 should be small enough so that the covariance between large points is of a similar order to the covariance

between small points. It should, however, should be large enough to ensure the covariance between points is distinct. These decisions should be made with the input space in mind, which can be challenging.

Instead of using a GP prior distribution with a linear covariance function, we could propose the model $f(x) = px + q$ and place prior distributions on p and q. However, this model has a different posterior distribution to using the GP prior distribution . We have chosen the linear covariance function over the parametric linear model because it is in keeping with our other proposed GP models.

4.2.2.1 Regression with a Linear Covariance Function

We now demonstrate the linear covariance function by applying it to a regression problem. We generate 100 input points, $\mathbf{x} = \{x_1, \ldots, x_{100}\}$, uniformly at random from $\{100, 101, \ldots, 30,000\}$. We generate observations $\mathbf{y} = \{y_1, \ldots, y_{100}\}$ by:

$$y_i = x_i^{0.8} \exp \varepsilon_i, \quad \varepsilon_i \sim N(0, 0.5^2).$$

We therefore consider the model:

$$\log y_i = f(\log x_i) + \varepsilon_i.$$

We place the following GP prior on the function f:

$$f \sim \mathcal{GP}(0,\, \Sigma), \quad \Sigma_{i,j} = k_{lin}(\log x_i, \log x_j).$$

The likelihood function is given by:

$$\pi(\mathbf{y}|f, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{100} (\log y_i - f(\log x_i))^2 \right\},$$

and by Bayes' Theorem, the posterior distribution is:

$$\pi(f|\mathbf{y}, \sigma^2) = \mathcal{GP}(f; 0,\, \Sigma) N(\log \mathbf{y};\, f,\, \sigma^2 I).$$

We sample from this density using an MCMC algorithm with an underrelaxed proposal mechanism, as described in section 2.5, whereby given the current function f, we propose a new vector f' by:

$$f' = \sqrt{1 - \delta^2} f + \delta \nu, \quad \nu \sim \mathcal{GP}(0,\, \Sigma),$$

which is accepted with probability

$$p_{acc} = \frac{\pi(\mathbf{y}|f', \sigma^2)}{\pi(\mathbf{y}|f, \sigma^2)}.$$

The results of 10,000 iterations of the MCMC algorithm are shown in figure 4.4, and we see the method is effective resulting in a good estimate for the true function.

4.3 Nonparametric Methods for Two Covariate Infection Rates

Now we have introduced the linear and two dimensional covariance functions, we extend our method from the previous chapters to model infection rates with two continuous covariates. The first two methods will be additive. We consider the covariates independently of each other. We model the infection rate over the covariates separately and then combine them to compute the pairwise infection rate.The third method is a coupled model, where we consider the two covariates to be dependent and infer the infection rate between each pair of individuals.

We consider two models for the infection rate function. The first is the additive model where we assume the infection rate from individual j to k is given by:

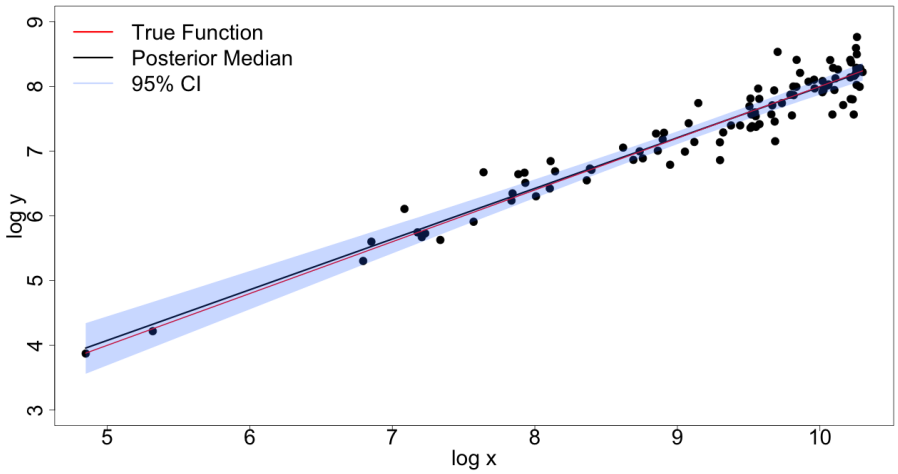
$$\beta_{j,k} = g(f_x(x_{j,k}) + f_y(y_{j,k})).$$

The second model is the coupled model, where we assume the infection rate is given by:

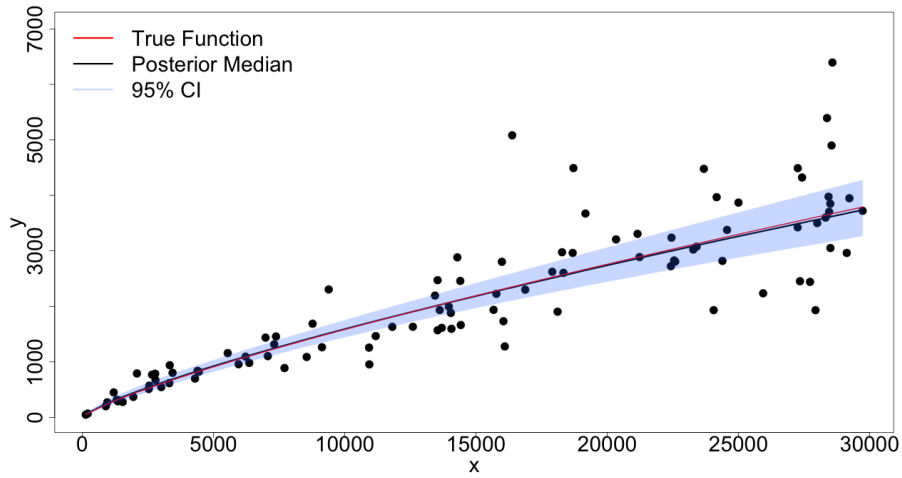
$$\beta_{j,k} = g(f(x_{j,k}, y_{j,k})).$$

Otherwise, the model remains the same and we recall the corresponding likelihood function in equation (2.1)

$$\pi(\mathbf{i}, \mathbf{r}|\beta, \lambda, \gamma, \kappa, i_\kappa) = \exp\left\{-\sum_{j=1}^{n}\sum_{k=1}^{N}\beta_{j,k}\left((r_j \wedge i_k) - (i_j \wedge i_k)\right)\right\}\prod_{\substack{j=1\\ j\neq\kappa}}^{n}\left(\sum_{k\in\mathcal{Y}_j}\beta_{k,j}\right)$$
$$\times\prod_{j=1}^{n} f_{\mathcal{D}}(r_j - i_j|\lambda, \gamma).$$



(a) The results on a log-log scale



(b) The results on the natural scale

Figure 4.4: The posterior median and 95% credible interval for the functions in the regression problem in section 4.2.2.1 with linear covariance functions.

4.3.1 The Additive Model

In the additive method, we assume that infections occur according to two independent Poisson processes underpinning the infection rate, the first according to the covariate x with rate f_x, and the second according to y with rate f_y. We superimpose these processes to model the infection rate function by:

$$\beta_{j,k} = g(f_x(x_{j,k}) + f_y(y_{j,k})),$$

where g is the transformation function to ensure $\beta_{j,k}$ is positive. We place independent GP prior distributions on f_x and f_y, as follows:

$$f_x \sim \mathcal{GP}(0,\, \Sigma_x) \quad (\Sigma_x)_{j,k} = k_x(x_j, x_k)$$
$$f_y \sim \mathcal{GP}(0,\, \Sigma_y) \quad (\Sigma_y)_{j,k} = k_y(y_j, y_k),$$

and a graphical representation of the model is given in figure 4.5.

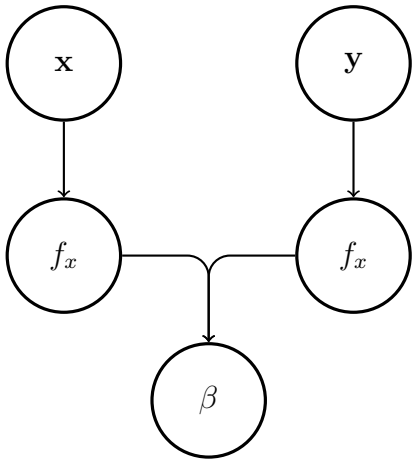


Figure 4.5: The additive GP prior distribution structure for the epidemic model.

For the epidemic model, we must choose suitable covariance functions to construct the covariance matrices for f_x and f_y. The most straightforward option is to continue using the squared exponential covariance function, shown in equation (1.2), as this can model a wide range of smooth functions. As the squared exponential function allows us to model such a large range of functions, it can cause the Markov chain to be slow to converge and this problem is exacerbated in two dimensions. As we typically

only observe removal times of individuals and these are not wholly informative, we may struggle to learn this function well. To overcome this, we suggest using the linear covariance function for one of the GP prior distributions. Using the linear covariance function is incredibly restrictive as we are only able to model functions that are linear or linear on the log or log-log scale. Reducing the number of functions we can model does increase the rate of convergence and improves the mixing of the Markov chain. This contradicts the overarching theme of the thesis somewhat, to avoid making strict parametric assumptions, but given the limited data, we may be forced to implement pragmatic, parametric assumptions.

If our chosen transformation function is $g = \exp$, we have three choices for the covariance functions in the additive model:

1. Both k_x and k_y are the squared exponential function,
2. k_x is the squared exponential covariance function; k_y is the linear covariance function with input on the natural scale, giving an exponential function and
3. k_x is the squared exponential covariance function; k_y is the linear covariance function with input on the log scale, giving a monomial function.

As for the remaining model parameters, we use the following independent prior distributions:

$$l \sim \text{Exp}(\chi_l), \tag{4.3}$$

$$\gamma \sim \text{Exp}(\chi_\gamma), \tag{4.4}$$

$$\kappa \sim \text{U}[1, \ldots, n], \tag{4.5}$$

$$i_\kappa \sim -z, \qquad z \sim \text{Exp}(\chi_\kappa), \tag{4.6}$$

where l is the length of f_x, γ the infectious period distribution rate parameter, κ the label of the first infected individual and i_κ the time at which they were infected. Combining these distributions with likelihood function in equation 2.1 by Bayes'

theorem, the posterior distribution is given by:

$$\begin{aligned}\pi(\beta,\gamma,\kappa,i_\kappa,l_x|\mathbf{r},\lambda,\alpha,l_y) \propto& \pi(\mathbf{i},\mathbf{r}|f_x,f_y\lambda,\gamma,\kappa,i_\kappa)\pi(f_x|l_x)\pi(l_x)\pi(f_y)\pi(\gamma)\pi(\kappa))\\ &\times\pi(i_\kappa|\kappa)\pi(f_y)\pi(\gamma)\pi(\kappa)\pi(i_\kappa|\kappa)\\ &\propto\exp\left\{-\sum_{j=1}^{n}\sum_{k=1}^{N}g(f_x(x_{j,k})+f_y(y_{j,k}))\left((r_j\wedge i_k)-(i_j\wedge i_k)\right)\right\}\\ &\times\prod_{\substack{j=1\\ j\neq\kappa}}^{n}\left(\sum_{k\in\mathcal{Y}_j}g(f_x(x_{k,j})+f_y(y_{k,j}))\right)\prod_{j=1}^{n}f_{\mathcal{D}}(r_j-i_j|\lambda,\gamma)\\ &\times\mathcal{GP}(f_x;0,\Sigma_x)\mathcal{GP}(f_y;0,\Sigma_y)\exp\{-l_x\chi_l\}\exp\{-\gamma\chi_\gamma\}\\ &\times\exp\{i_\kappa\chi_\kappa\}.\end{aligned}$$

We have replaced $\beta_{j,k}$ by its nonparametric model $g(f_x(x_{j,k})+f_y(y_{j,k}))$. We only need to make minor adjustments to the overall MCMC algorithm for this model. When sampling the functions f_x and f_y, we again use the MPA to allow for large data sets and this methods works with both the squared exponential and linear covariance functions. To use the MPA, we first construct training input sets for the GP prior distributions and we denote them by $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$. We then propose new functions $f'_{\bar{x}}$ and $f'_{\bar{y}}$ over the training sets, and compute the proposed value of β by:

$$\beta'=g(\Sigma_{\mathbf{x},\bar{\mathbf{x}}}\Sigma^{-1}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}f'_{\bar{x}}+\Sigma_{\mathbf{y},\bar{\mathbf{y}}}\Sigma^{-1}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}f'_{\bar{y}}),$$

where $\Sigma_{\mathbf{x},\bar{\mathbf{x}}}$ is the covariance function for x applied to the full data set $\mathbf{x}$ and the training set $\bar{\mathbf{x}}$, and similarly for $\Sigma_{\mathbf{y},\bar{\mathbf{y}}}$. The function g is our chosen transformation function to ensure the infection rate is positive; letting g be the exponential function and denoting the projected GP samples by f'_x and f'_y gives

$$\beta'=e^{f'_x}e^{f'_y}.$$

As the functions f_x and f_y are independent of each other we store their covariance matrices separately; it may be possible to use a block covariance matrix, but this would be prohibitively large to store and the off-diagonal blocks would be 0. The full MCMC algorithm is given in algorithm 10.

Algorithm 10 Structure of the Additive Model MCMC algorithm

1: Initialise the chain with values $\gamma^{(0)}$, $f_x^{(0)}$, $f_y^{(0)}$,$l_x^{(0)}$, and $\mathbf{i}^{(0)}$

Repeat the following steps

2: Sample γ from the conditional distribution $\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r})$ using a Gibbs step

3: Sample f_x and f_y individually using an underrelaxed proposal mechanism for a Metropolis Hastings step

4: Sample the appropriate GP hyperparameters

5: Sample κ and i_κ using a Metropolis Hastings step

6: Update an infection time

4.3.2 The Coupled Model

We now propose a model where the dependency on x and y cannot be separated and we model the infection rate function by:

$$\beta_{j,k} = g(f(x_{j,k}, y_{j,k})).$$

A graphical representation of the structure of this model is given in figure 4.6.

To apply this model to an epidemic setting, we first construct the covariance matrix using the 2D squared exponential covariance matrix. The prior distribution for β is therefore given by

$$\beta = g(f), \quad f \sim \mathcal{GP}(0,\, \Sigma), \quad \Sigma_{i,j} = k((x_i, y_i), (x_j, y_j);\, \alpha, \Lambda). \tag{4.7}$$

For this model, we need to infer the infection rate $\beta = g(f)$, the GP length scales, l_x and l_y, as well as the infectious period distribution parameter λ and the infection times $\mathbf{i}$. We have found it challenging to infer both GP length scales simultaneously and have encountered issues distinctly identifying them. This is because there is only one covariance value for each pair of individuals, and we are inferring two length scale parameters from this one covariance value. We recommend fixing either length scale, say l_y, and inferring the other. Multiplying the likelihood function in equation (2.1) by the multi-input GP prior in equation (4.7) and the prior distributions in

equations (4.3) – (4.6) gives the following posterior distribution:

$$\begin{aligned}\pi(\beta,\gamma,\kappa,i_\kappa,l_x|\mathbf{r},\lambda,l_y,\alpha) &\propto \pi(\mathbf{i},\mathbf{r}|f,\lambda,\gamma,\kappa,i_\kappa)\pi(f|l_x,l_y)\pi(l_x)\pi(\gamma)\pi(\kappa)\pi(i_\kappa|\kappa)\\ &\propto \exp\left\{-\sum_{j=1}^{n}\sum_{k=1}^{N} g\left(f(x_{j,k},y_{j,k})\right)\left((r_j\wedge i_k)-(i_j\wedge i_k)\right)\right\}\\ &\times\prod_{\substack{j=1\\ j\neq\kappa}}^{n}\left(\sum_{k\in\mathcal{Y}_j} g\left(f(x_{k,j},y_{k,j})\right)\right)\prod_{j=1}^{n} f_{\mathcal{D}}(r_j-i_j|\lambda,\gamma)\mathcal{GP}(f;0,\Sigma)\\ &\times\exp\{-l_x\chi_l\}\exp\{-\gamma\chi_\gamma\}\exp\{-i_\kappa\chi_\kappa\}.\end{aligned}$$

The full MCMC algorithm is shown in algorithm 11.

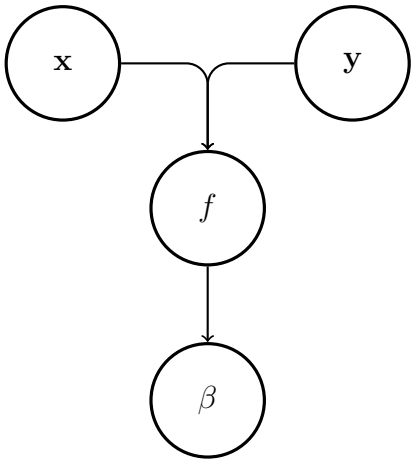


Figure 4.6: The coupled GP prior distribution structure for the epidemic model.

Algorithm 11 Structure of the Coupled Model MCMC algorithm

1: Initialise the chain with estimates $\gamma^{(0)}$, $f^{(0)}$, $l_x^{(0)}$, and $\mathbf{i}^{(0)}$

Repeat the following steps

2: Sample γ from the conditional distribution $\pi(\gamma|\lambda,\mathbf{i},\mathbf{r},\chi_\gamma)$ using a Gibbs step

3: Sample f using an underrelaxed proposal mechanism for a Metropolis Hastings step

4: Sample l_x using a Metropolis Hastings Random Walk step

5: Sample κ and i_κ using a Metropolis Hastings step

6: Update an infection time

4.4 Comparison of the Models

In this section we have constructed two models: the additive model and the coupled model. There are two types of additive models, one where both functions are modelled using a GP with a squared exponential kernel, and one where we assume one of the functions to be linear on either a natural or log scale. These models allow us to consider the same problem in different ways. As in the previous chapter, the models place emphasis on different outputs and serve different audiences. They also have different approaches to large data sets or large amounts of missing data.

The additive model allows us to model the effect of the different covariates and see how they differ. This is of interest to practitioners. For example when developing control measures for an outbreak of Foot and Mouth disease, it is useful to know whether to place high priority on preventing infections between nearby farms or from farms that have a large number of animals present. The coupled method, however, gives the infection rate between each pair of farms and as such is more suited to a mathematical analysis with the posterior and posterior predictive distributions. This model is therefore well suited to statistical analysis of the outbreak and making predictions about outbreaks on the data set.

The models also differ in the assumptions we need to make when dealing with limited data. The additive model may require strict parametric assumptions about the shape of the function of one of the covariates. When using a Bayesian nonparametric framework, this is undesirable. It is nevertheless a pragmatic approach. The coupled approach does not require such strong assumptions as the coupling allows information to be shared between covariates.

4.4.1 Extending the Models for more than two Covariates

In this chapter, we have considered outbreaks where the infection rate depends on two covariates only. The framework can be extended to include more covariates. For the additive model, we can place a GP prior distribution on the function for each covariate and their sum gives the log infection rate functions. For the coupled model,

we can extend the covariance function into any finite number of dimensions. However, this becomes increasingly difficult as the amount of information contained in the data is limited and we may not be able to learn about all of the covariates from the data.

To overcome this lack of information in the data, we can make a series of increasingly strict assumptions. The first is to use the additive linear model, where we assume any number of the functions for individual covariates are linear or monomial. We can also form a more traditional semi-parametric model where we use GP prior distributions for some covariates and propose parametric forms for others. Another method is to strengthen the assumptions we make about the infection times, for example assuming a constant infectious period.

4.5 Simulation Studies

We now examine the success of the models using simulated data. We simulate 250 outbreaks of a disease among 1,000 individuals, where the infection rate from individual j to individual k is given by:

$$\beta_{j,k} = \frac{1}{1,000} \exp\{-3d_{j,k}\} \sqrt{\frac{w_k}{100}},$$

where $d_{j,k}$ is the Euclidean distance between them and w_k is the weight of individual k, the individual being infected. The weights were drawn uniformly at random from the interval $[100, 1000]$ and we label the individuals by weight, from smallest to largest. We generate the positions of the individuals by:

$$\begin{aligned}
x_{i,1} &\overset{\text{i.i.d}}{\sim} N(0.5, 0.1^2), \quad i = 1, \ldots, 500, \\
y_{i,1} &\overset{\text{i.i.d}}{\sim} N(0.5, 0.1^2), \quad i = 1, \ldots, 500, \\
x_{i,1} &\overset{\text{i.i.d}}{\sim} N(1.5, 0.3^2), \quad i = 501, \ldots, 1000, \\
y_{i,1} &\overset{\text{i.i.d}}{\sim} N(1.5, 0.3^2), \quad i = 501, \ldots, 1000,
\end{aligned}$$

so that there are two clusters of individuals. We choose to have two clusters for similar reasons to those in chapter 2 – to test the model's ability to overcome few intracluster transmissions. Our justification for the weights are related to the Avian

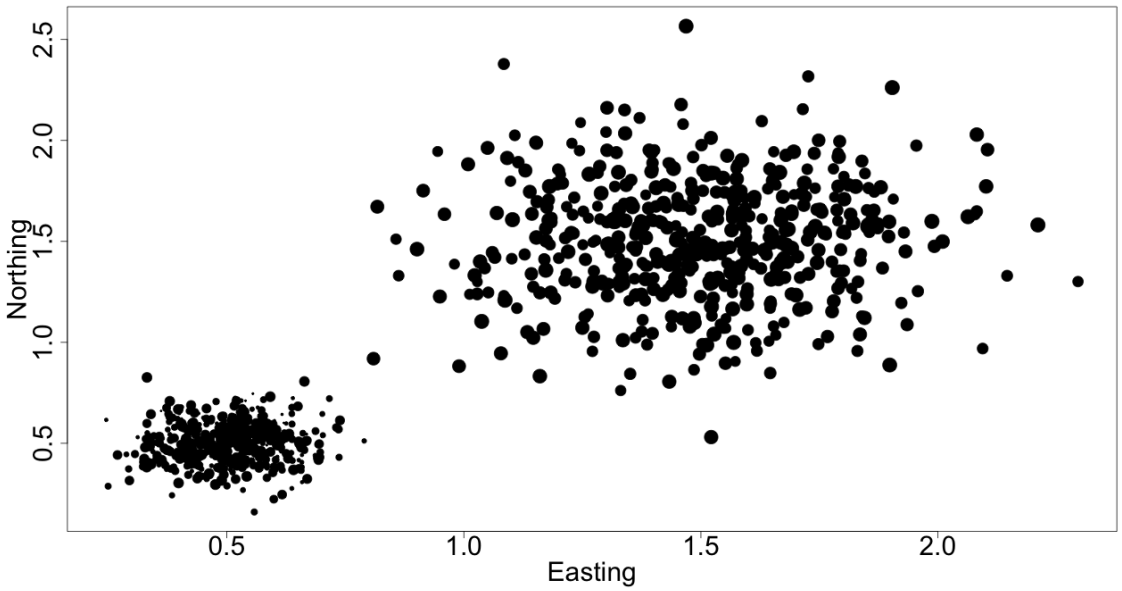


Figure 4.7: The population used for the simulation studies in section 4.5. The coordinates of the individuals are shown with the size of the individuals representing the weight.

Influenza data set we analyse in chapter 5. In outbreaks among livestock hobby farmers are often affected. These are typically households in towns or villages who keep a handful of animals in their back garden, and as such there will be a large number of farms with a small number of animals in the same cluster. Figure 4.7 shows the population structure.

4.5.1 The Additive Squared Exponential Model

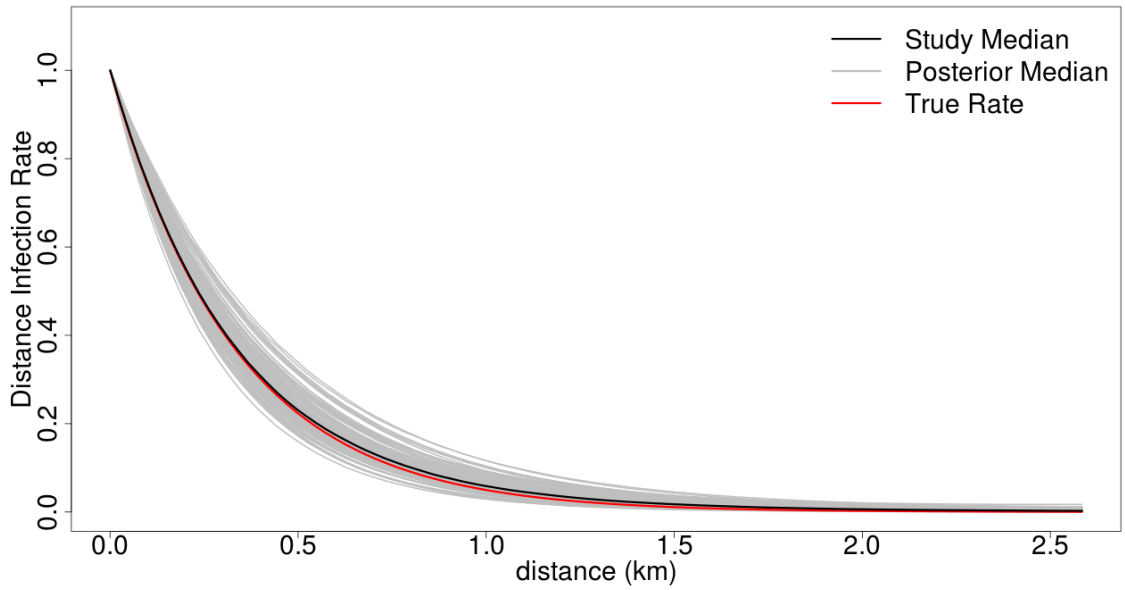
For the squared exponential model, we place independent GP prior distributions on f_x and f_y, the values for the infection rate functions for distance and weight respectively. For both prior distributions we set $\alpha = 10$, which results in a vague distribution. For these data sets and combined with the squared exponential covariance functions, we found it challenging to infer both length scales, so we chose to set $l_x = 4$km and $l_y = 15,000$. These values were chosen by generating samples from the prior distributions with various length scales.

Figure 4.8 shows the results of the simulation study. We normalise the results such that the spatial component for immediate neighbours is 1, i.e., $f_x(0) = 1$. This allows us to compare the results from each of the 250 datasets. The results for the distance function closely match the true fucntion. However the variance for these functions is much higher than for the one covariate simulation study. For the multi-covariate study, we do not see functions increasing at the furthest pair-wise distances as we did in the single covariate case (for example see figure 2.12). Despite there being more information in these specific cases, we require the model to learn more than in the single covariate case and the model struggles to capture the effects of both covariates. The results for the weight covariate are substantially worse than for the distance covariate. The model had difficulties when inferring this rate. The first reason for this is there being fewer data points for this covariate; for a population of size N, we observe N weights, however we observe $\frac{N(N-1)}{2}$ pair-wise distances. The second reason is that the data was less informative about this covariate than the spatial element. It still, however, captures the overall shape. This poor performance is one motivation for implementing the linear model.

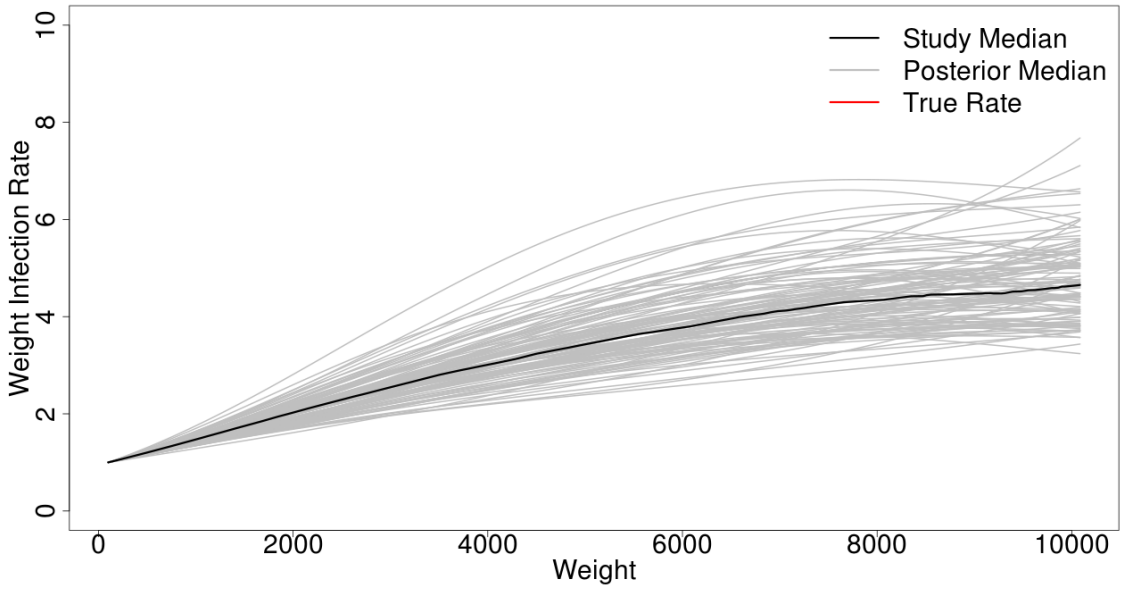
4.5.2 The Additive Linear Model

We now turn our attention to the additive linear model. We place a GP prior distribution with a squared exponential covariance function on f_x, the function used to model the distance element of the infection rate, and a GP prior distribution with a linear covariance function on f_y, the function modelling the size-based function on the infection rate. We place the prior distribution of f_y over the logarithm on the size space and, as placing the prior distributions on $\log \beta$, f_y will model monomial functions. To allow for uncertainty between pairs of individuals with small sizes, in the linear covariance function, we set $\alpha_0 = 5$, and to reduce the impact of very large individuals, we set $\alpha_1 = 0.01$. In the squared exponential function, we set $\alpha_0 = 10$ and placed a vague exponential prior on the length scale to infer this parameter.

The results of functions f_x and f_y are shown in figure 4.9, as there are identifiability issues for this model, we scaled the results so $\log f_x(0) = 0$ and $\log f_y(0) = 0$. Theses



(a) The inferred functions for the distance part of the infection rate.



(b) The inferred functions for the size part of the infection rate.

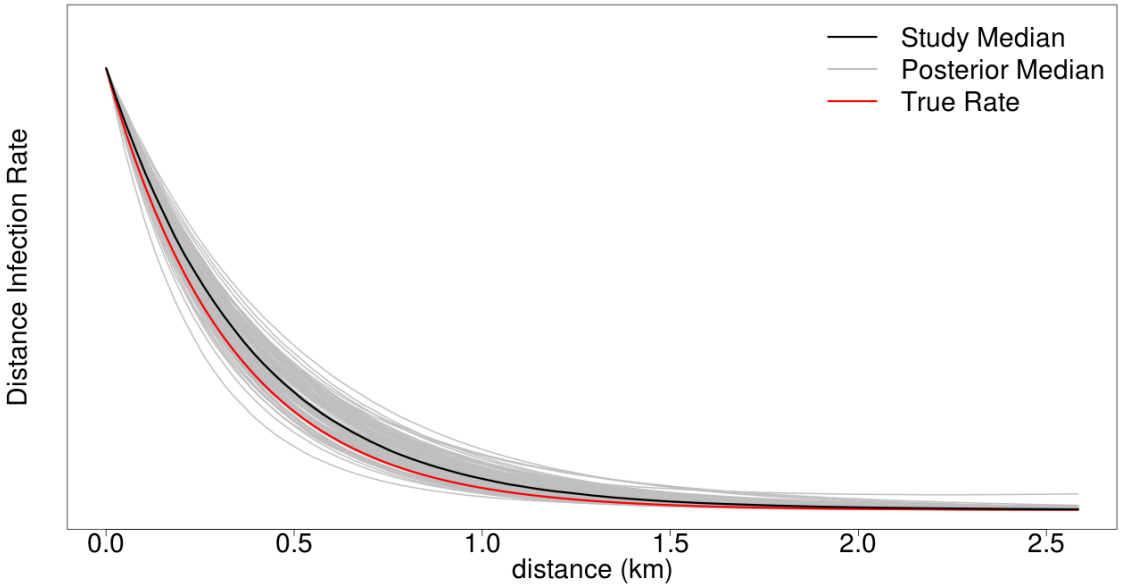
Figure 4.8: The inferred infection rate functions for the additive squared exponential model simulation study in section 4.5.1. The grey lines are the posterior median infection rate functions for each of the 250 data sets. The black line is the median of all 250 infection rate functions and the red line is the true infection rate function.

issues stem from us treating the single function of interest β as the sum of two functions. We can see that we infer the shape of both functions well. However, we overestimate both functions slightly on average. There is much larger uncertainty around the estimates for the size function, this is because there is less data available, and this covariate only depends on the infected individual, not the relationship between both the infector and infected individual.

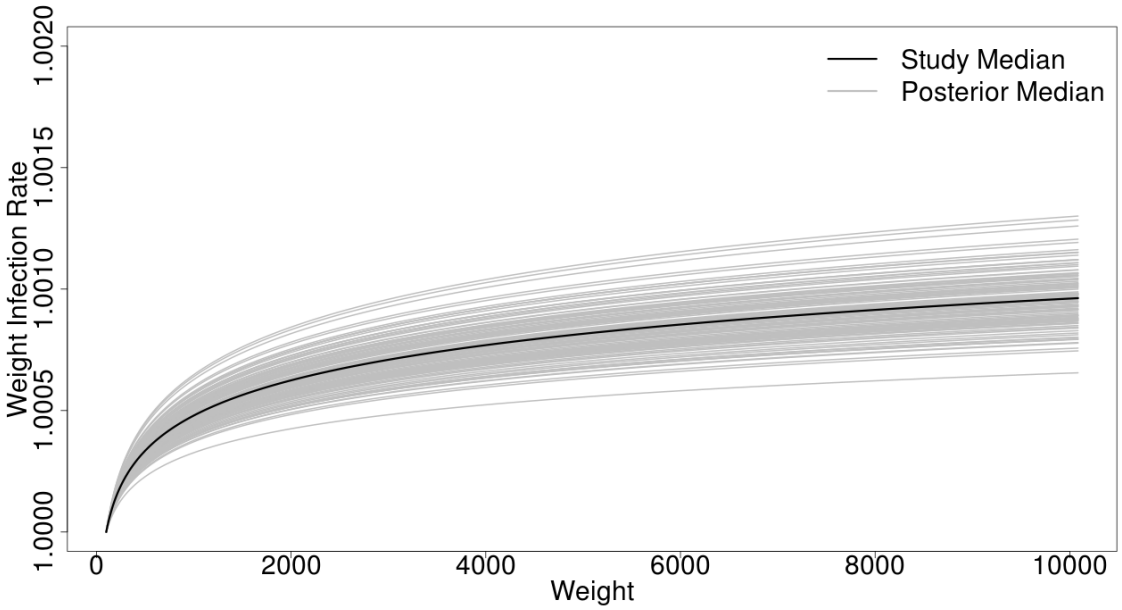
4.5.3 The Coupled Model

We now repeat the inference using the coupled model. Figure 4.10(a-b) shows the true pair-wise infection matrix, our posterior mean pair-wise infection matrix, on a log scale and the relative error for each pairwise rate. These figures display the posterior median for each pair-wise infection rate $\beta_{i,j}$. Figure 4.10(c) is the most revealing, showing that the error in the model is relatively small for individuals infecting other individuals in the same cluster, but outside of the cluster we tend to overestimate the infection rate. We also see that the model overestimates the infection rate for individuals with a small weight. Due to the size of the population, figure 4.10 is not suited to understanding the results of the model, and so we instead consider the posterior median for the infection rate function f.

The median for the pseudo-function $\bar{f}$ is shown in figure 4.12 and we see the study median is close to the true rate. In order the visualise this function, we plot the posterior median of the pseudo-function, which is over a smaller pseudo dataset and on the log scale. We choose to do this as plotting the full function for each outbreak was not feasible. We do see that we overestimate the infection rate for the first 12 individuals, which all have small weights. This figure also shows the uncertainty is much larger when the distance between individuals is larger, this corresponds to the log infection rate functions between -12 and -10. The reason for this uncertainty is that we see many intra-cluster infections, but fewer inter-cluster infections, meaning the model has less data to learn from.

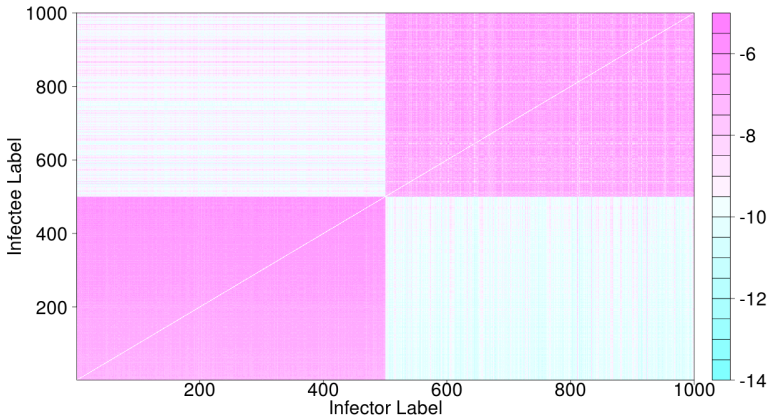


(a) The inferred functions for the distance part of the infection rate.

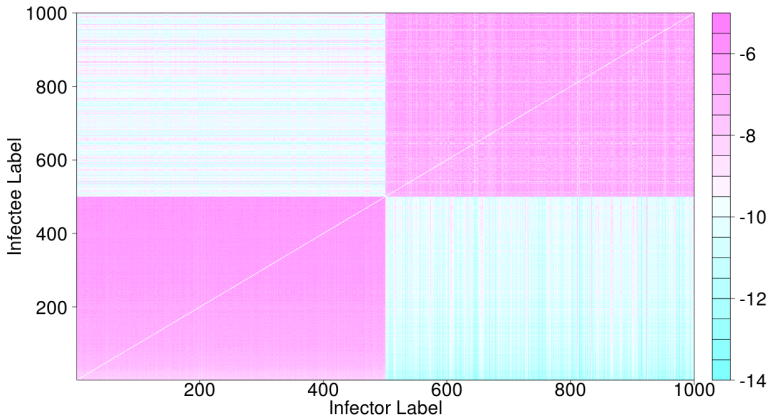


(b) The inferred functions for the size part of the infection rate.

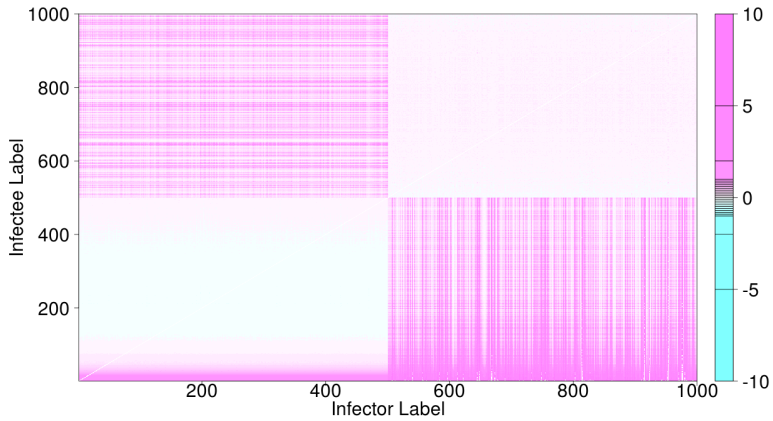
Figure 4.9: The inferred infection rate functions for the additive linear simulation study in section 4.5.2. The grey lines are the posterior median infection rate functions for each of the 250 data sets. The black line is the median of all 250 infection rate functions and the red line is the true infection rate function.



(a) The posterior mean pairwise infection matrix



(b) The true pairwise infection matrix



(c) The relative error for pairwise infection rates

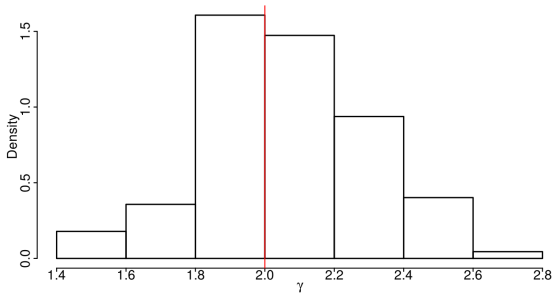
Figure 4.10: The pair-wise log infection rates for each pair of individuals in the coupled simulation study in section 4.5.3. The posterior median value for $\beta_{i,j}$ is given in row i column j. We recall that there are two clusters of individuals and the individuals are ordered from smallest to largest weight.

The results for the remaining parameters are shown in figure 4.11. Both the infectious period distribution rate parameter and the sum of the infection times are inferred well. We can also infer the length scale of the GP for the weight component.

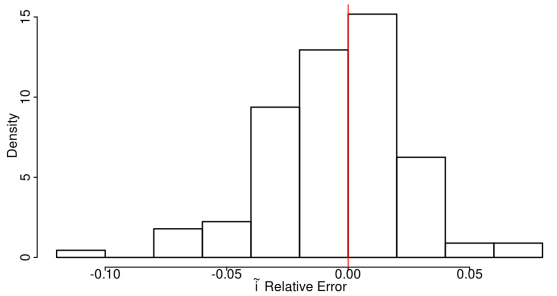
4.5.4 Remarks on the Simulation Studies

In figure 4.13 we directly compare the three models, by computing the relative error in the posterior means for $\beta_{600,j}$. We see the errors in both the size and distance parts. We recall that the individuals are labelled by weight, so the individuals with the smallest weights are first. Considering the first 100 individuals, we see that all models overestimate the infection rate to these individuals, the coupled model in particular. By introducing the log-linear assumption, we are able to reduce this bias, as we are forcing the model to take smaller values in this region. The other systematic bias is in the distance function, and we recall the individuals are divided into two clusters, with individuals 1,...,500 in the first and the remaining in the second cluster. There is no systematic bias when estimating the infection rate to individuals in the same cluster as individual 600, but infection rates in the other cluster are consistently overestimated by all models. This is because the individuals are further away and we observe fewer infections between clusters. Therefore the model has fewer observations from which to learn.

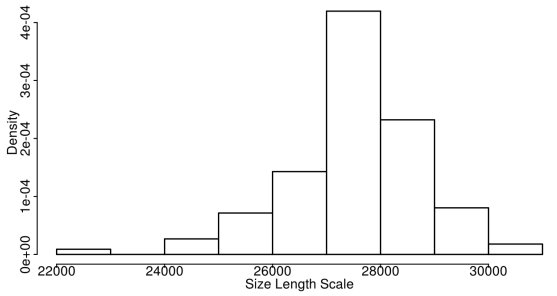
Despite there being systematic biases in the results caused by the lack of data and few assumptions in the model, the simulation studies have shown our methodology can be effective. Alongside the infection rate function, we were also able to infer model hyperparameters and the infection times. The best results come from the model with the linear kernel, which is because we are including an additional assumption, which is true, into the model. As we need to include additional assumptions in the model, and considering our aim was to reduce the number of assumptions about the infection rate function, this suggests we are reaching on the borderline with what is possible with this methodology with this amount of data. To improve results and extend the models further, we would either need to observe a large amount of data, or observe infection times.



(a) The distribution of the 250 estimates for γ.



(b) The distribution of the 250 estimates for the relative error in the sum of the infection times



(c) The distribution of the 250 estimates for l, the weight length scale

Figure 4.11: Results for the coupled model simulation study in section 4.5.3 for the infectious period distribution and length scale parameters.

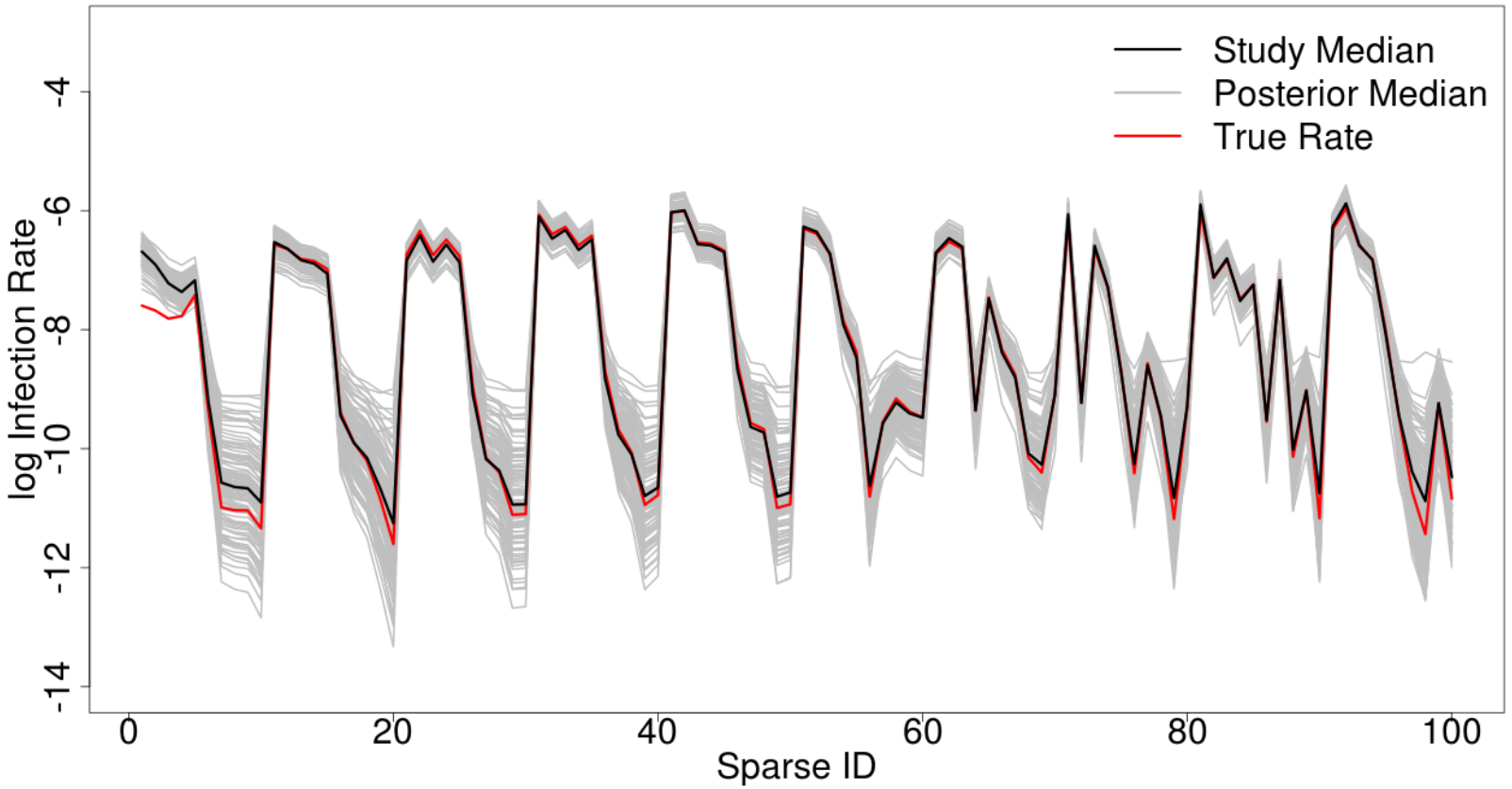


Figure 4.12: The infection rate function before projection onto the full dataset and transformation to a positive function. The sparse ID refers to the MPA and the pair of observations in the pseudo data set.

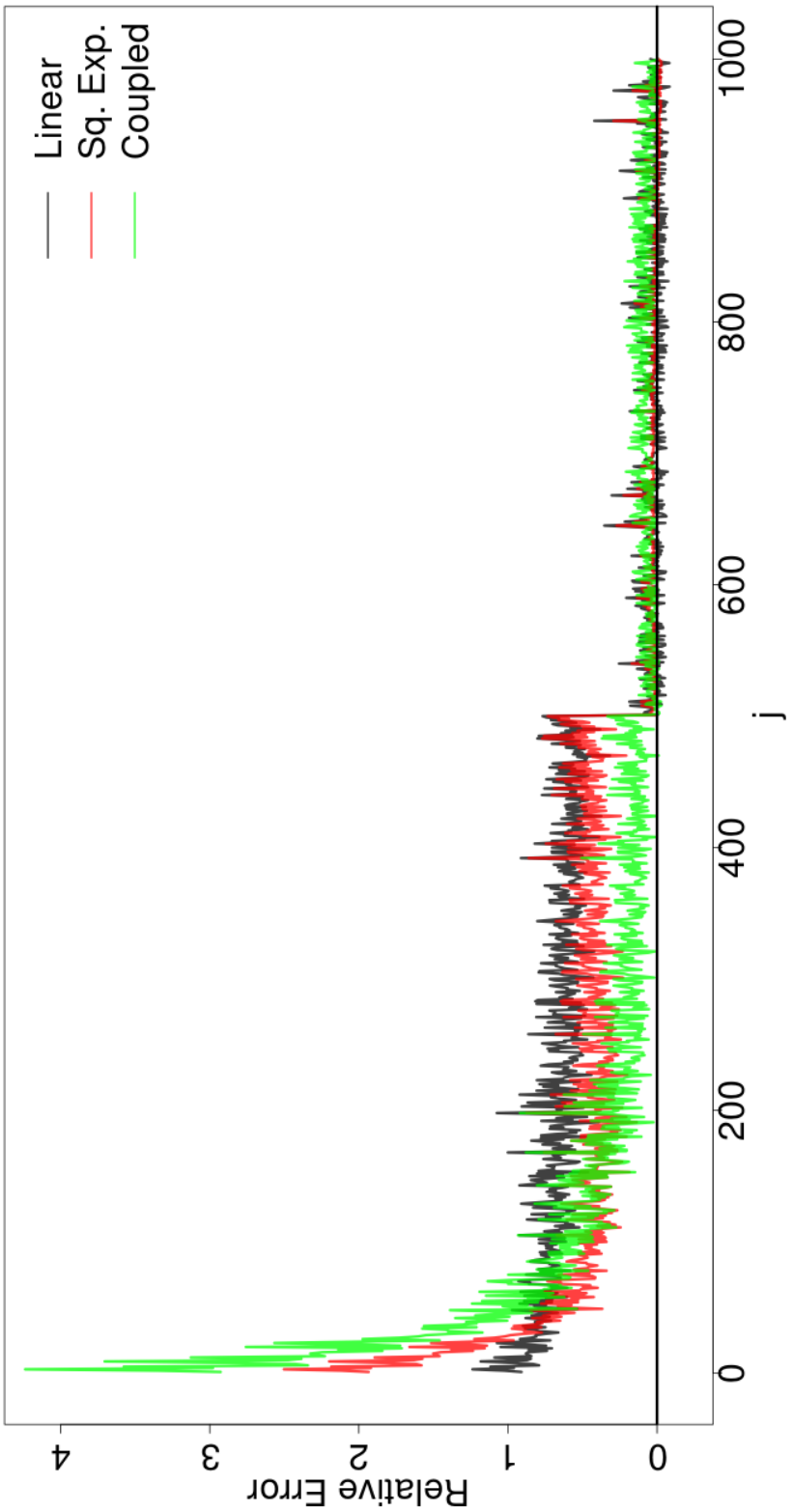


Figure 4.13: The relative error in the infection rate from individual 600 to each other individual, $\beta_{600,j}$.

4.6 Conclusion

In this chapter, we have introduced new models allowing infection rates to depend on two relationships between the infected and susceptible individuals. For example the distance between them and the size of the susceptible individual. We proposed three ways of modelling this problem: an additive model, an additive model with linear assumptions, and a coupled model. In the additive models, we assumed that the two relationships have no influence on one another and they could be considered separately. In the coupled model, we considered the covariates jointly, assuming they cannot be considered separately. The structure of the model is ultimately a modelling choice, all constructions having advantages and being intended for different audiences and purposes. The additive models allowed us to easily see the shape of the functions for each covariate, which may be useful when developing disease control strategies. Conversely, the construction of the coupled model allowed us investigate the pair-wise infection rates and make predictions via the posterior predictive distribution.

These methods require a large amount of data to return good results without systematic bias. In our simulation studies, we have shown that even with 1,000 individuals in the population, there are still irregularities in the results, particularly with individuals with a smaller weight. Another weakness was that we overestimated transmissions between clusters as we do not observe many intra-cluster infections. However, given that we have made much fewer assumptions than in a a comparable parametric framework, we are still able to infer the infection rate functions well. This method can be successfully combined with existing methods for inferring the other model parameters and unobserved infection times. Observing infection times would greatly improve the accuracy of the model, and improve results.

We have developed methods for extending our models for the infection rate functions which have more than two inputs. As the number of inputs increases, the amount of data or strength of assumptions about the function must also increase. Despite this problem, we were still able to infer the infection rate functions alongside the other model parameters and estimate the times individuals were infected.

CHAPTER 5

Bayesian Nonparametric Methods for Individual-Level Stochastic Epidemic Models in Practice

5.1 Introduction

The spread of diseases on farms can pose a high risk to public health and food chains. Governments have a number of powers to control the spread of diseases, including culling animals that are already or at risk of being infected. Developing an effective culling strategy can be challenging as the government has to ensure enough animals are culled to eliminate the disease, but minimise the amount of money paid in compensation to farmers, as well as minimising the economic impact on the wider agriculture industry.

If we are able to estimate the pair-wise infection rate between farms, we can improve the culling strategy in future outbreaks of the disease. Once we have inferred the infection rate functions, we can use predictive simulation techniques to simulate outbreaks of a disease. We can implement culling strategies in the simulation to see their effect. For example, in Backer et al. (2015), the authors fit a model to outbreak data of Avian Influenza in the Netherlands and estimate the economic impact of

various culling and vaccination strategies in controlling an outbreak. In Probert et al. (2018), the authors analyse the effect of culling farms in two outbreaks of Foot and Mouth disease, one in the UK and one in Japan, and develop a method for making decisions about control strategies while the outbreak is ongoing.

In this chapter, we will analyse outbreaks of two diseases, Avian Influenza and Foot and Mouth disease, with the aim of constructing an accurate model of the outbreak and using this to investigate disease control strategies. Both Avian Influenza and Foot and Mouth disease have been widely studied by infectious disease modellers. For example, outbreaks of Avian Influenza have been studied in the Netherlands (Elbers et al., 2004; Boender et al., 2007) and Thailand (Retkute et al., 2018). The outbreak of Foot and Mouth disease we analyse has been studied extensively (see e.g. Keeling, 2001; Lawson and Zhou, 2005; Diggle, 2006; Kypraios, 2007; Jewell et al., 2009; Stockdale et al., 2018; Probert et al., 2018).

We will first analyse an outbreak of Avian Influenza. In 2003, there was a large outbreak of Avian Influenza in the Netherlands in which over 300 million birds were culled, 90 people were infected with the virus and one person died as a consequence (Stegeman et al., 2004). This dataset poses interesting problems as there is a clear spatial element to the spread of the disease (see figure 5.1), which we will quantify. But also, as the authorities pre-emptively culled farms to control the spread of the disease, we have a large number of farms whose disease status in unknown. The second outbreak we will analyse is the 2001 outbreak of Foot and Mouth disease in the UK. During this outbreak, livestock on more than 2,000 farms were culled (Stockdale et al., 2019) and the had an economic impact of over £8 million (Jewell et al., 2009). There is evidence that different animals are susceptible to the disease in different ways (Alexandersen et al., 2003) and so we must consider the type of animals on each farm. To model this outbreak, we will implement our methods for multi-type models and analyse how susceptible different types of farms are to the virus.

Having developed Bayesian nonparametric methodology for modelling the outbreak of infectious diseases, we now apply this to two data sets mentioned above. We will

model the infection rate between two farms without making strict assumptions about the parametric form or make any assumption which lack any biological justification. As well as inferring the infection rate functions, we will infer other parameters such as the times farms were infected and the infectious period distribution rate parameter.

5.1.1 Layout of Chapter

We first analyse an outbreak of Avian Influenza in the Netherlands. As well as inferring the infection rate nonparametrically, we use the posterior predictive distribution to investigate the effects of various culling strategies on future outbreaks. We then investigate an outbreak of Foot and Mouth in Disease in the UK, where we use our MOGP method to quantify the difference between cattle, sheep, and cattle and sheep farms.

In this chapter, we are able to use our Bayesian nonparametric methods to estimate the infection rate functions in both outbreaks. In the Avian Influenza, we identify which pre-emptively culled farms were likely to have been infected and analyse possible culling strategies using the posterior predictive distribution, which has not been done before. As we are using Bayesian nonparametric methods, we are better able to quantify the uncertainty in the data, and take this uncertainty into account when making predictions by using the posterior predictive distribution. In the outbreak of Foot and Mouth disease, we are able to better analyse how the infection rate between farms with two types on animals and one type of animals are related, using both the Multi-Output Covariance model and the Discrepancy Model.

5.2 Avian Influenza

Highly Pathogenic Avian Influenza is an illness caused by a virus. It is most common in birds, but can also be transmitted from birds to humans and, for some strains, between humans. Highly Pathogenic Avian Influenza can cause severe problems for poultry farmers. The symptoms for birds include: lack of appetite, reduced mobility, respiratory problems and ultimately death (Velkers et al., 2006). In countries with

well developed control strategies, if the poultry on a farm are infected with a highly pathogenic strain of the virus, the birds on the farm are culled and the farmer is compensated for the loss of livestock (Backer et al., 2015). The virus is also fatal to humans, though the transmission rate between humans is typically much lower than between birds (Stegeman et al., 2004).

In this section, we begin by giving an overview of a large-scale epidemic of Avian Influenza in the Netherlands from 2003. We review the current literature on this outbreak, examining the model that has been used on this data set previously. We then apply our new nonparametric methods to infer the infection rate and we compare our results to analyses using parametric models. Finally, we investigate various culling strategies and their economic impact, which has not been done in a Bayesian way.

5.2.1 An Epidemiological Overview

In the Netherlands in 2003, a large outbreak of the highly pathogenic Avian Influenza A/*H7N7* virus occurred, which resulted in the death of one veterinarian, the non-fatal infection of 89 others, and the culling of over 30 million birds. Of the 5,397 bird farms in the Netherlands, 233 were confirmed to have been infected immediately after the outbreak was finished, and over 1,200 farms with an unknown infection status were pre-emptively culled (Boender et al., 2007). The locations and status of the farms are shown in figure 5.1. Several poultry farms in Belgium and Germany were also infected and control measures were introduced in these countries as well. Before applying our nonparametric methods, we first outline the timeline of the outbreak and how the infection spread.

5.2.1.1 A Brief Timeline

The European Commission has made available a complete chronology of this outbreak covering all infections, directives implemented, and safeguarding measures used in the Netherlands, Germany, Belgium, and Luxembourg (Director Generate Health and Consumers, 2003). We now summarise the key events in the outbreak in the Netherlands.

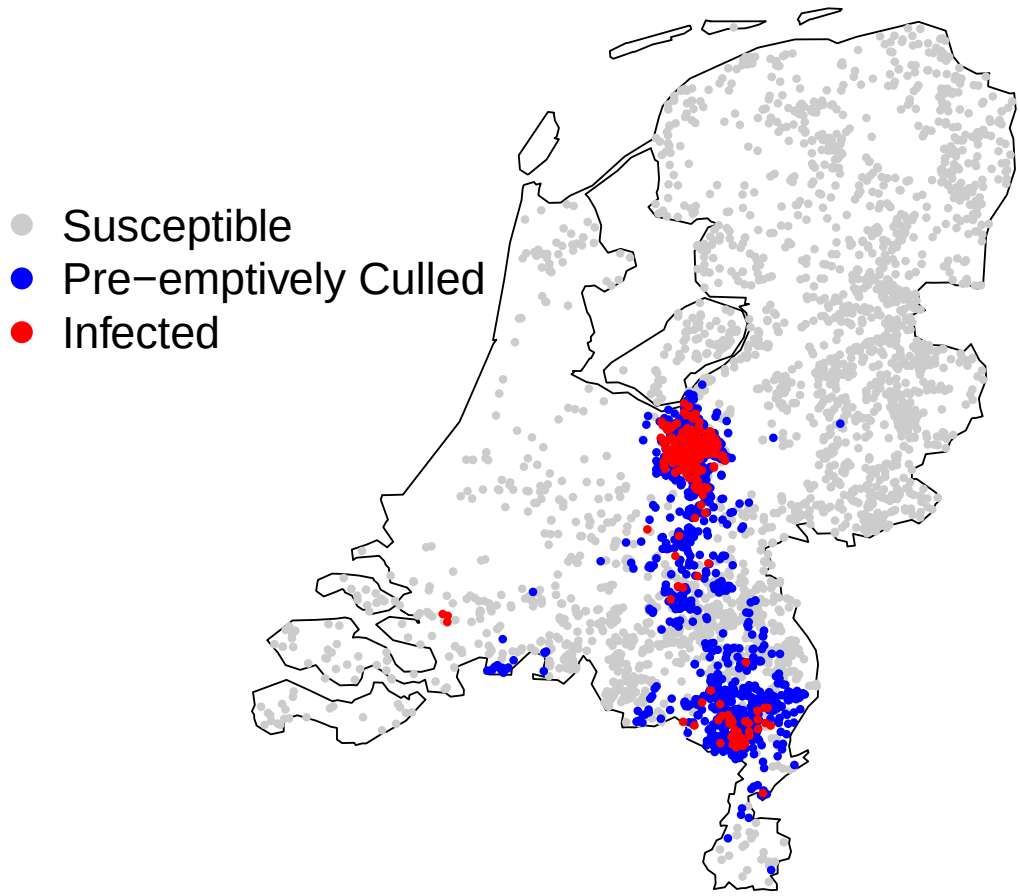


Figure 5.1: The locations and statuses of poultry farms after the outbreak of Avian Influenza A in the Netherlands in 2003.

The first recorded case of Avian Influenza in this outbreak occurred on the 28^{th} February 2003 on a Layer Farm in a bird farming area in the central Netherlands. The cause of this infection is suspected to be wild fowl infecting birds on this farm with a low-pathogenic strain of the H5 or H7 viruses. This strain then mutated into a highly pathogenic strain and spread among one layer flock on the farm.

The following day, all movements of poultry and eggs were banned both nationally and internationally. The government of the Netherlands implemented an EU directive imposing protection zones around infected farms, introducing disinfection points on farms and ensuring poultry on infected farms were kept indoors. On the 3^{rd} March, at least 17 farms were confirmed to have been infected and the government started to cull all infected farms. On 5^{th} March, the culling was expanded to include any farm within a 1 km radius of a farm which was infected.

The following week, nationwide testing for the virus began. The testing suggests the infection had now spread to a region much further south of the first cases, as well as farms on the border with Belgium. By this point over 2.4 million birds had been culled, and as a result of the new areas being infected, the government was emptying farms around the infected regions to create buffer zones.

At the start of April, 177 farms were confirmed to be infected, including 5 turkey farms, the first of the species to be infected. As a result, the culling radius was increased to 3 km of a suspected infected farm, and the culling of all turkeys within a 10 km radius of an infected turkey farm was implemented. Disinfection points were introduced on the German, Belgian and Luxembourgish borders with the Netherlands, and farms in these countries within 3 km of the borders were culled to reduce the chance of the epidemics spreading to these countries.

The only human fatality was announced on 19^{th} April – a veterinarian who died of pneumonia, with the Avian Influenza virus found in his lungs. As a result of this, the Dutch government implemented new safety strategies for individuals who were working with the virus, began testing for the virus in pigs, and increased the size of buffer zones on the borders. By this point, the number of birds being culled reached 750,000 birds a day (Stegeman et al., 2004), and over 20 million birds had been culled.

The last confirmed case of the virus in the Netherlands was on 16^{th} May. Over the following months, the restrictions put in place were slowly lifted, with the final safeguarding measure being removed on 22^{nd} August.

5.2.1.2 The Transmission Mechanisms

So far we have only viewed the infection process as the result of an inhomogeneous Poisson process. To aid our understanding of the model, we consider how livestock on farms are infected. The data detailed above shows there are several infection mechanisms which spread the virus in different ways.

The first mechanism was through wild waterfowl, who carried a strain of low pathogenic Avian Influenza. Wildfowl harbour a low pathogen Avian Influenza virus which can be transmitted to free-range birds, and the virus can then mutate into a highly pathogenic variety (Boender et al., 2007). In both Elbers et al. (2004) and Fouchier et al. (2004), the authors suggest that the first confirmed infection was caused by wildfowl at a lake less than 1 km from the farm. In Boender et al. (2007) the authors also noted that the fifth confirmed farm to be infected was located on the opposite side of the same road as the first case, and in Fouchier et al. (2004) the high number of Avian Influenza viruses that circulated among ducks, geese and migratory birds during the same time period are cited. This suggests that wildfowl were likely to be a cause of spatial transmission, spreading the virus to farms in a small area.

The second mechanism was the movements of livestock and eggs between farms. In Boender et al. (2007), the authors suggest that the disease could have been detected earlier had warning systems been put in place and followed. Only six farms were suspected to be infected when farm movements were halted, it is likely that infected birds were traded before this, allowing the virus to be spread at a national and international level.

The third mechanism was human transmission, where humans acted as vectors. In Fouchier et al. (2004) the 89 individuals who were infected with Avian Influenza are documented. 86 of these individuals had come into contact with infected poultry, most of them after having visited several farms. It is possible for the virus to be

carried between farms by humans on their clothing or on machinery.

5.2.2 Literature Review

This outbreak has been studied extensively, with many studies carried out into the medical aspect of the virus. The outbreak posed several interesting factors from a medical standpoint. In Fouchier et al. (2004) and Koopmans et al. (2004) human-human transmission is documented, something which had not been observed in similar H7N1 and H7N3 outbreaks. A single fatality occurred during the outbreak, which is identified as unusually many in Fouchier et al. (2004). In Elbers et al. (2004), the authors discuss the outbreak from a veterinary standpoint, investigating the events that occurred on infected farms leading up to the outbreak and developing more effective disease control measures. In Bataille et al. (2011), the authors analyse virus sequence data from the outbreak.

In Stegeman et al. (2004), the authors were the first to look at the outbreak from a modelling standpoint. They assumed that the infection rate was constant between all farms and that all farms were infected for a fixed time period T. Using a fixed time period for the infectious period distribution means the authors assume the infection times are known. They used these assumptions to implement a generalised linear model proposed in Becker (1989)[pp. 108-110], where the probability a susceptible farm avoids infection on day t is given by:

$$p(t) = \exp\Big\{ -\beta \frac{I(t)}{N} \Big\}.$$

Using this probability, the number of cases of Avian Influenza arising day t, $C(t)$, can be modelled by the following binomial distribution:

$$C(t) \sim \text{Bin}\big(S(t),\, 1 - p(t)\big).$$

Combining this with the complementary log-log link function, the infection rate can be estimated using the formula:

$$\log(-\log(1 - p(t)) = \log \frac{I(t)}{N} + \log \beta.$$

In Boender et al. (2007), the authors built on this model by including a spatial element, replacing the constant infection rate with a heterogenous rate given by $\beta_{i,j}$. In the model, farms are either susceptible, infected but not infectious, infected, or removed, and they define the infection force being applied to a susceptible farm i up to day t as:

$$\lambda_i(t) = \sum_{j \neq i} \beta_{i,j} 1_{[j \text{ is infectious at time } t]},$$

which gives the probability farm i avoids infection up to day t as:

$$p_i(t) = \exp\{-\lambda_i(t)\}.$$

Their likelihood function is split into three parts: the farms that remained susceptible throughout the entire outbreak, the farms that were susceptible but were culled as a disease control measure, and the farms that were infected. The likelihood function is therefore written as:

$$\pi(\mathbf{i}, \mathbf{r}|\beta) = \prod_{k \in K} p_k(t_{max}) \prod_{l \in \Lambda} p_l(t_{cull,l}) \prod_{m \in M} p_m(t_{inf,m})(1 - p_m(t_{inf,m})),$$

where K is the set of farms which remained susceptible and t_{max} is the final day of the outbreak, Λ is the set of farms which were preemptively culled and $t_{cull,l}$ is the day farm l was culled, and M is the set of infected farms and $t_{inf,m}$ is the day farm m was infected. They propose the spatial, parametric infection rates shown in table 5.2. They numerically compute MLEs for the model parameters and use Akaike Information Criterion to choose the best of the proposed models.

Although the latter method takes spatial variation into account, the parametric forms of the infection rate can be restrictive. Using nonparametric inference would allow us to be flexible and learn the infection rate directly from the data. Both methods assume a constant infectious period, whereas using data augmentation within a Bayesian framework will again allow us to be more flexible, and give better estimates of the infectious period distribution and infection dates for individual farms. Data augmentation will allow us to estimate which of the preemptively culled farms were likely to have been infected, something that has not yet been considered.

5.2.3 Data

The data set contains the locations of all 5,397 poultry farms in the Netherlands at the time of the outbreak. We discarded all 931 hobby farms from the data set, as they had fewer than 500 poultry. For each farm, we have its status at the end of the outbreak, describing whether it had remained susceptible throughout the course of the outbreak, had been culled due to infection, or had been culled pre-emptively. These were removed as they are civilians who keep a small number of chickens in their back gardens, and so the strict bio-securtiy measures were not applied to these farms. For farms which were culled, we have the date on which this occurred. At the end of the outbreak, 233 farms had been culled because they were confirmed to be infected with the virus and 1,232 were culled as a precaution. The positions of the farms are shown in figure 5.1.

5.2.4 Stochastic Epidemic Model

We construct our model based on the standard epidemic model in continuous time (see e.g. Bailey (1975); Andersson and Britton (2000)). We begin by assuming that all farms are initially disease free but the animals on one farm become infected due to an external source. At any time t, a farm is either susceptible to the disease, infected with the disease and infectious, or removed as the animals on the farm have been culled. The model can be separated into two processes: the infection process and the removal process. The infection process is governed by a rate function $\beta = \beta(d)$, where d denotes the Euclidean distance between two farms.

We assume an infectious farm infects a given susceptible farm that is $d$km away according to a Poisson process with rate $\beta(d)$. The processes governing different pairs of farms are assumed to be independent. For the removal process, once a farm is infected it is infectious for a time period given by a random variable distributed according to a $\Gamma(\lambda, \gamma)$ distribution, which has mean λ/γ and variance λ/γ^2. Once this time has elapsed the birds on this farm are culled. We model the culling strategy by assuming the animals on all farms within an $r$km radius are also culled at this

time.

We now derive the likelihood function for this model. This is similar to the likelihood function derived in equation (2.1), but is adjusted to allow for farms to be culled pre-emptively. Let N denote the total number of poultry farms in the Netherlands and n the number of confirmed infected farms. We label the infected farms from $1, \ldots, n$ by their culling date and the remaining farms $n+1, \ldots, N$ arbitrarily. We denote the infection and culling times for farm j by i_j and r_j respectively and we define κ to be the label of the initial infected farm. We define $\mathbf{i} = \{i_1, \ldots, i_{\kappa-1}, i_{\kappa+1}, \ldots, i_N\}$ to be the set of infection times excluding the initial infected and $\mathbf{r} = \{r_1, \ldots, r_N\}$ to be the set of removal times, and we centre the times such that $r_1 = 0$. For farms which were not infected, their infection times are set to be $i_j = \infty$ and for farms which were not culled, we set $r_j = \infty$.

We construct the following four sets based on the final infection status of the farms:

- set A consists of the farms that remained susceptible to the disease throughout the course of the epidemic and were not culled,
- set B is the set of farms that were infected with the virus and culled as a consequence,
- set C is the set of farms that were infected but were culled pre-emptively due to a nearby farm being infected,
- and set D consists of the farms that were not infected but pre-emptively culled due a nearby farm being infected.

These sets are shown in table 5.1. From the data, we are unable to distinguish between farms in sets C and D.

The likelihood function consists of three parts: the contribution from farms avoiding infection, the contribution from farms being infected, and the contribution for farms remaining infectious for a given length of time. For a farm k in either set A, B or C, the contribution from the farm avoiding becoming infected by infectious

Set	Infected	Culled	Pre-emptively Culled
A	$\times$	$\times$	$\times$
B	✓	✓	$\times$
C	✓	✓	✓
D	$\times$	✓	✓

Table 5.1: The different sets of farms in the Avian Flu data set.

farm j is:

$$\psi_{j,k} = \exp\{-\beta(d_{j,k})((r_j \wedge i_k) - (i_j \wedge i_k))\},$$

where $\beta(d_{j,k})$ is the infection rate for a pair of farms that are $d_{j,k}$km apart, and $a \wedge b = \min\{a, b\}$. The difference in minimum times is the amount of time during which farm j exerted pressure on farm k. If the farm k is in set D we must take into account its pre-emptive culling time and the contribution from it avoiding infection is given by:

$$\psi_{j,k} = \exp\{-\beta(d_{j,k})((r_j \wedge r_k) - (i_j \wedge r_k))\}.$$

When farm j is infected, we consider the set of farms that were infectious immediately before j was infected. This is given by:

$$\mathcal{Y}_j = \{k : i_k < i_j < r_k\},$$

and the event that j is infected contributes to the likelihood function through the overall hazard rate of the infection:

$$\phi_j = \sum_{k \in \mathcal{Y}_j} \beta(d_{k,j}).$$

For the removal process, the likelihood is given by:

$$L_{rem} = \prod_{j \in B} h(r_j - i_j|\lambda, \gamma) \prod_{j \in C} S(r_j - i_j|\lambda, \gamma),$$

where $h(x|\lambda, \gamma)$ is the probability density function of the $\Gamma(\lambda, \gamma)$ distribution evaluated at x and $S(x|\lambda, \gamma)$ is the survivor function

$$S(x|\lambda, \gamma) = \int_x^\infty h(u|\lambda, \gamma)du.$$

Farms in set B, that were infected and culled at the end of their infectious period, contribute to the likelihood function through the total time they were infectious with respect to the infectious period distribution. For those in set C, who were infected but culled pre-emptively, we consider their removal time as a censoring time, and compute the probability they would have remained infectious longer given their culling time. Combining the infection and removal processes gives the following likelihood function:

$$\begin{aligned}\pi(\mathbf{i}, \mathbf{r}, B, C, D|\beta, \lambda, \gamma, \kappa, i_\kappa) &= \left(\prod_{j\in\mathcal{B}\cup\mathcal{C}}\prod_{k=1}^{N}\psi_{j,k}\right)\left(\prod_{\substack{j\in\mathcal{B}\cup\mathcal{C}\\ j\neq\kappa}}^{n}\kappa_j\right)\prod_{j\in B}h(r_j - i_j|\lambda,\gamma)\\ &\quad\times\prod_{j\in C}S(r_j - i_j|\lambda,\gamma) \qquad (5.1)\\ &= \exp\{-\Psi\}\prod_{\substack{j\in\mathcal{B}\cup\mathcal{C}\\ j\neq\kappa}}\left(\sum_{k\in\mathcal{Y}_j}\beta(d_{k,j})\right)\prod_{j\in B}h(r_j - i_j|\lambda,\gamma)\\ &\quad\times\prod_{j\in C}S(r_j - i_j|\lambda,\gamma),\end{aligned}$$

where:

$$\Psi = \sum_{j\in\mathcal{B}\cup\mathcal{C}}^{n}\left[\sum_{k\in A\cup B\cup C}\beta(d_{j,k})\left((r_j\wedge i_k) - (i_j\wedge i_k)\right) + \sum_{k\in D}\beta(d_{j,k})\left((r_j\wedge r_k) - (i_j\wedge r_k)\right)\right]. \qquad (5.2)$$

We note that the removal times determine the set A, which is why the set A does not appear in the left hand side of equation (5.1).

5.2.5 Fixed Infectious Period

We compare the results of parametric and Bayesian nonparametric methods. In order to compare our results to those presented in Boender et al. (2007), we assume the infectious to be constant and last $7\frac{1}{2}$ days for each farm. We follow Jewell et al. (2009) and Stockdale et al. (2019) for the infectious period distribution and we set $\lambda = 4$.

Rate	Kernel
1	$\beta_{j,k} = \beta_0$
2	$\beta_{j,k} = \frac{\beta_0}{1+d_{j,k}}$
3	$\beta_{j,k} = \frac{\beta_0}{1+d_{j,k}^2}$
4	$\beta_{j,k} = \frac{\beta_0}{1+d_{j,k}^{\beta_1}}$
5	$\beta_{j,k} = \frac{\beta_0}{1+(d_{j,k}/\beta_2)^{\beta_1}}$

Table 5.2: The proposed parametric pair-wise infection rates for the Avian Influenza data set.

5.2.5.1 Parametric Methods

We follow Boender et al. (2007) and propose the infection rate kernels in table 5.2. The first proposed kernel is a homogeneously mixing model, where the distance between farms is not taken into account and the remaining models are all variants on the logistic function. We are interested in the infection rate function parameters β_0, β_1, β_2. We assume the infectious period distribution lasts for $7\frac{1}{2}$ days for each farm (Elbers et al., 2004). The posterior distribution is given by:

$$\pi(\beta_0, \beta_1, \beta_2, |\mathbf{i}, \mathbf{r}, A, B, C, D, \lambda, \kappa, i_\kappa) \propto \pi(\mathbf{i}, \mathbf{r}, B, C, D|\beta_0, \beta_1, \beta_2, \lambda, \kappa, i_\kappa) \times \pi(\beta_0)\pi(\beta_1)\pi(\beta_2).$$

As for this analysis, we assume the infectious period is fixed, the label and time of the first farm to be infected, κ and i_κ are known. We place the following independent prior distributions on the model parameters:

$$\beta_0 \sim \text{Exp}(0.01),$$
$$\beta_1 \sim \text{Exp}(0.01),$$
$$\beta_2 \sim \text{Exp}(0.01),$$

In section 5.2.5.3, we report on the implementation of the MCMC algorithm for parametric infection rates in chapter 2 (algorithm 4).

5.2.5.2 Nonparametric Methods

We now repeat the inference for the model. However, instead of proposing a parametric form for the function β, we model it nonparametrically. The posterior distribution is given by:

$$\pi(\beta, \gamma|\mathbf{i}, \mathbf{r}, A, B, C, D, \lambda, \kappa, i_\kappa) \propto \pi(\mathbf{i}, \mathbf{r}, B, C, D|\beta, \lambda, \gamma, \kappa, i_\kappa) \times \pi(\beta)\pi(\gamma).$$

We place a GP prior distribution on the dummy function f and use the exponential function to transform it into the non-negative rate function β. The prior distribution is:

$$\beta = \exp(f), \quad f \sim \mathcal{GP}(0,\ \Sigma), \quad \Sigma_{jk} = k(d_j, d_k; \alpha, l),$$

where d_j is the distance between the j^{th} pair of farms. As there are almost 10 million pairs of farms in this dataset, we use the MPA method (see section 2.6). We set the pseudo-dataset to be $\bar{\mathbf{d}} = \{0, 0.5, 1, \ldots, 24.5, 25, 35, 45, \ldots, 425\}$. This allows us a finer level of detail over the first 25 km, as this is the likely range where the infection rate function will be non-zero. We use the squared-exponential covariance function and fix α to 4 as this gives a distribution that is sufficiently vague. We place a vague exponential prior distribution on the length scale parameter l, such that $l \sim \text{Exp}(0.01)$. We place a vague, conjugate prior distribution on γ such that $\gamma \sim \text{Exp}(0.01)$. We generate samples from the posterior distribution using algorithm 5 in Chapter 2.

5.2.5.3 Comparison of the Methods

For the parametric model, we run the MCMC algorithm for a burn-in period of 500 iterations, followed by another 10,000 iterations. The posterior medians for the model parameters are given in table 5.3 alongside the Akaike Information Criterion (AIC) for the proposed form. We use AIC to directly compare our findings to those in Elbers et al. (2004); the AIC for kernel m is defined by:

$$\text{AIC}_m = 2k - 2\log\pi(\beta_{j,k}^{(m)}|\mathbf{i}, \mathbf{r}),$$

Infection Rate	Parameter	Posterior Median	95% Cred. Int.	Δ AIC
1	β_0	2.88×10^{-5}	$(2.53\times10^{-5}, 3.25\times10^{-5})$	860
2	β_0	9.27×10^{-4}	$(8.14\times10^{-4}, 10\times10^{-4})$	208
3	β_0	5.37×10^{-3}	$(4.71\times10^{-3}, 6.07\times10^{-3})$	0
4	β_0	4.96×10^{-3}	$(3.96\times10^{-3}, 6.04\times10^{-3})$	0.778
	β_1	1.93	(1.79, 2.07)	
5	β_0	3.54×10^{-3}	$(2.18\times10^{-3}, 6.06\times10^{-3})$	3.98
	β_1	2.05	(1.81, 2.30)	
	β_2	1.41	(0.842, 1.94)	

Table 5.3: Posterior medians and 95% credible intervals for the parametric infection rates, alongside the ΔAIC values.

where k is the number of model parameters and $\beta_{j,k}^{(m)}$ is the kernel for infection rate m. It attempts to balance model fit with parsimony, and we choose the best fitting model to be that with the lowest AIC value. We define ΔAIC_m to be:

$$\Delta\text{AIC}_m = \text{AIC}_m - \min_j \text{AIC}_j,$$

which shifts all the AIC values, such that the best of the proposed models has an AIC value of 0. We use this to easily identify the differences between the AIC values. Table 5.3 shows the AIC values, and infection rate function 3 is chosen as the most suitable of the proposed forms, however models 4 and 5 are also suitable.

For the nonparametric method, we run the MCMC algorithm for 10,000 iterations, removing the first 1,000 as a burn-in period. Figure 5.2 shows the results of the parametric and nonparametric methods when the infectious period is fixed and constant. We can see the methods broadly agree with each other, with the main difference being the uncertainty. The large uncertainty in the nonparametric methods is the price paid for a more flexible model. As well as a more flexible model, the nonparametric method allows us to infer plausible values of the length scale, with our posterior median being 2.75 km (95% CI: (2.55, 3.01)). This can help inform judgements about control strategies.

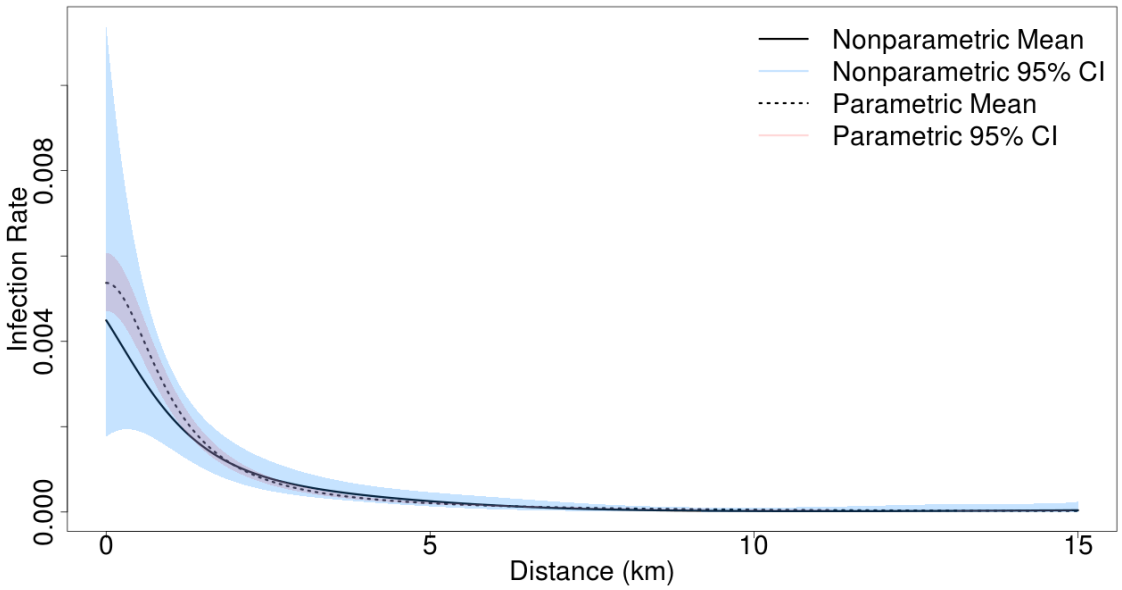


Figure 5.2: The results of the parametric (dotted and red) with the nonparametric (solid and red) methods for the infection rate function with fixed infectious period.

Another noticeable difference between the two methods is the computational time required. The parametric method took approximately 2 hours to run 10,000 iterations of an MCMC algorithm, whereas the nonparametric method took approximately 100 hours to run the same number of iterations. Both methods were written in the C programming language, however the nonparametric method also used OpenMP software when calculating the likelihood function. The difference in speed is down to the computation of the likelihood function, constructing the infection matrix, $\beta_{i,j}$, but mainly inverting and decomposing the covariance matrix when proposing new values of the length scale. Some improvement may be made by using a different C compiler, as this was performed using the GCC compiler and GSL BLAS libraries.

We also note the difference in the fitted values compared to Elbers et al. (2004), in which this authors estimate for an immediate neighbour ($d_{i,j} = 0$) the infection rate is 0.002 farms per day. Our estimates are over twice as high, 0.042 farms per day for the nonparametric and 0.00537 farms per day for the parametric model. This difference is largely caused by two factors: the use of a different compartmental

models and the different likelihood functions.

5.2.6 Unknown Infection Times

We now repeat the nonparametric method, except with unknown infection times. As in the inference for the parametric model, we assume the infectious period distribution is given by a $\Gamma(4, \gamma)$ distribution, where the shape parameter γ is to be estimated. We also wish to estimate the infection times and statuses of the pre-emptively culled farms, that is whether they were in set C or D, and we do this using data augmentation techniques.

By Bayes' theorem, the target posterior density is given by:

$$\pi(\beta, \gamma, \kappa, i_\kappa, \mathbf{i}, C, D|\mathbf{r}, B, \lambda) \propto \pi(\mathbf{i}, \mathbf{r}, B, C, D|\beta, \lambda, \gamma, \kappa, i_\kappa)\pi(\beta)\pi(\lambda)\pi(i_\kappa|\kappa)\pi(\kappa).$$

We implement the same prior distribution of the infection rate function as the fixed infection time case and place a GP prior distribution on it as follows:

$$\beta = \exp\{f\}, \quad f \sim \mathcal{GP}(0, \Sigma), \quad \Sigma_{jk} = k(d_j, d_k;\, \alpha, l).$$

As inferring the infection times and statuses is computationally expensive, we fix the GP prior distribution length scale parameter to be 3 km. This avoids us having to repeatedly invert and decompose the covariance matrix.

We now describe the prior distributions we place on the remaining model parameters. We first consider the infectious period distribution, which is given by the $\Gamma(\lambda, \gamma)$ distribution. We place a vague conjugate exponential prior on γ such that $\gamma \sim \text{Exp}(0.01)$. We follow Jewell et al. (2009) and assume λ to be fixed and greater than 1 as this gives a bell-shaped distribution with mean that only depends on γ. This has the advantage of improving the mixing of the resulting Markov chain in the MCMC algorithm. With regards to the infection times, we place a uniform prior on κ, the label of the initial infected. As we assume the first removal to be on day 0, the prior on the infection time of κ is given by :

$$i_\kappa|\kappa = -z, \quad z \sim \text{Exp}(0.01).$$

The posterior distribution is given by:

$$\begin{aligned}\pi(\beta,\gamma|\mathbf{r},B,C,D,\lambda,\kappa,i_\kappa,\mathbf{i}) &\propto \exp\{-\Psi\}\prod_{\substack{j=1\\ j\neq\kappa}}^{n}\left(\sum_{k\in\mathcal{Y}_j}\exp\{f(d_{k,j})\}\right)\\ &\propto \prod_{j\in B} f(r_j-i_j|\lambda,\gamma)\prod_{j\in C} S(r_j-i_j|\lambda,\gamma) \qquad (5.3)\\ &\quad\times \mathcal{GP}(f;\,0,\Sigma)\exp\{-0.01\gamma\}.\end{aligned}$$

The likelihood contribution to the posterior distribution is the same as in equation (5.1) and Ψ is the same as in equation (5.2), but β is replaced by the inferred function $\exp(f)$.

As the infection times are unknown, we need to infer these times as well as whether pre-emptively culled farms were infected. To implement this we develop the MCMC algorithm for inferring the infection rate function and infection times (algorithm 5) to allow us to infer the infection status of pre-emptively culled farms. This is given in algorithm 12.

Algorithm 12 Structure of the MCMC algorithm

1: Initialise the chain with values $f^{(0)}$, $\gamma^{(0)}$, $l^{(0)}$, and $\mathbf{i}^{(0)}$

Repeat the following steps

2: Sample f using an underrelaxed proposal mechanism for a Metropolis-Hastings step

3: Sample l using a Metropolis-Hastings random walk step

4: Sample γ using a Metropolis-Hastings random walk step

5: Sample κ using a Metropolis-Hastings random walk step

6: Sample i_κ using a Metropolis-Hastings step

7: Choose one of the following steps with equal probability:

- Update an infection time
- Remove an infection time for a pre-emptively culled farm
- Add an infection time for a pre-emptively culled farm

To infer the infection status of pre-emptively culled farms in the MCMC algorithm,

we follow Jewell et al. (2009), where the authors proposed a method where the individuals were removed without their infection status being known. This method allows us to propose inserting or deleting infection times for individuals with an unknown infection status. Suppose there are N individuals in the population, n of which were infected, and where m were removed without their infection status being known. We suppose that at each iteration of the algorithm, $\tilde{m}$ of the individuals who were pre-emptively removed have had infection times added by the algorithm. For each iteration of the algorithm we choose one of three events: moving, inserting or deleting an infection time. We choose which of the three events occurs with equal probability. For epidemics where the infection statuses are completely known, we only choose to update infection times.

5.2.6.1 Moving Infection Times

We randomly choose an individual j with an infection time and propose a new infection time by $i'_j = r_j - t_j$, where $t_j \sim \Gamma(\lambda, \gamma)$. We accept the proposal with probability

$$p_{acc} = \frac{h(r_j - i_j|\lambda,\gamma)}{h(r_j - i'_j|\lambda,\gamma)} \frac{\pi(\mathbf{i} - i_j + i'_j, \mathbf{r}|\beta,\lambda,\gamma)}{\pi(\mathbf{i},\mathbf{r}|\beta,\lambda,\gamma)} \wedge 1.$$

5.2.6.2 Inserting Infection Times

We randomly choose one of the $m - \tilde{m}$ pre-emptively removed individuals with no infection time and propose infecting them. Should $m = \tilde{m}$ the step is abandoned. We propose an infection time as above and accept it with probability

$$\begin{aligned} p_{acc} &= \frac{1/(\tilde{m}+1)}{(1/(m-\tilde{m}))h(r_j - i'_j|\lambda,\gamma)} \frac{\pi(\mathbf{i} + i_j, \mathbf{r}|\beta,\lambda,\gamma)}{\pi(\mathbf{i},\mathbf{r}|\beta,\lambda,\gamma)} \wedge 1 \\ &= \frac{m-\tilde{m}}{(\tilde{m}+1)h(r_j - i'_j|\lambda,\gamma)} \frac{\pi(\mathbf{i} + i_j, \mathbf{r}|\beta,\lambda,\gamma)}{\pi(\mathbf{i},\mathbf{r}|\beta,\lambda,\gamma)} \wedge 1. \end{aligned}$$

To derive the proposal ratio, suppose there are m individuals with unknown infection statuses, and $\tilde{m}$ have had infection times added by the algorithm. The numerator is the probability of individual j having an infection time and it being deleted. This

is equivalent to the probability of there being $\tilde{m}+1$ possible infection times to be deleted and randomly choosing the time belonging to j, which is given by $\frac{1}{\tilde{m}+1}$. The denominator is the probability we choose to add an infection time for individual j, given by $\frac{1}{m-\tilde{m}}$ and we propose i'_j, given by $h(r_j - i'_j|\lambda,\gamma)$.

5.2.6.3 Deleting Infection Times

We randomly choose an individual j who at the current iteration has an infection time added and propose removing their infection time. Should there be no individuals with an unknown infection status, who, at the current iteration of the algorithm, have had an infection time added, the step is abandoned. We accept this proposal with probability

$$\begin{aligned} p_{acc} &= \frac{1/(m-(\tilde{m}-1))h(r_j - i_j|\lambda,\gamma)}{1/\tilde{m}} \frac{\pi(\mathbf{i}-i_j,\mathbf{r}|\beta,\lambda,\gamma)}{\pi(\mathbf{i},\mathbf{r}|\beta,\lambda,\gamma)} \wedge 1 \\ &= \frac{h(r_j - i_j|\lambda,\gamma)\tilde{m}}{m-(\tilde{m}-1)} \frac{\pi(\mathbf{i}-i_j,\mathbf{r}|\beta,\lambda,\gamma)}{\pi(\mathbf{i},\mathbf{r}|\beta,\lambda,\gamma)} \wedge 1. \end{aligned}$$

The proposal ratio is given in a similar fashion to the ratio from the inserting proposal. The numerator is the probability we choose to insert an infection time for individual j and we propose i_j. Suppose j is not infected, then there are $m-(\tilde{m}-1)$ individuals to choose from, so the proposal probability is given by $\frac{1}{m-(\tilde{m}-1)}h(r_j - i_j|\lambda,\gamma)$. The denominator is the probability we choose individual k from the $\tilde{m}$ possible individuals, and delete their infection time. This probability is given by $\frac{1}{\tilde{m}}$.

5.2.7 Results

We ran the MCMC algorithm for 20,000 iterations, including a burn-in period of 500 iterations. For each update of the MCMC algorithm, we repeat the step which updates, adds or deletes infection times 200 times. Following the results with fixed infection times, we fix the length scale parameter $l = 3$km as this reduces the computational time required. The results for the infection rate are shown in figure 5.3, where the function decays to zero. From this, we estimate that after approximately 6km the probability of infection is negligible. From the credible interval, we can see

samples from the posterior distribution will take a variety of shapes, with functions that have a high infection rate over short distance decaying quickly and functions that have a lower rate over short distances taking a logistic-type form. The reason for the difference in shapes over short distances is because there is little data to learn from. In a small radius of an infected farm we observe that most other farms are infected, so we learn that the infection rate is high but we do not learn the shape of the function well.

To assess our results against existing methods, we compare our results to the best fitting parametric model with fixed infection times, namely model 3 in table 5.2. The results are shown in figure 5.3 and one clear difference between the parametric and nonparametric methods is the associated uncertainty. Although the nonparametric method allows for a greater degree of flexibility, it also induces a greater degree of uncertainty. Despite this, both estimates are of similar shape and scale and our results broadly agree with existing work.

As we assume the infection times to be unknown, we infer the times farms were infected alongside the infection rate. We estimate the mean infectious period to be 6.4 days, and figure 5.4 shows the distribution of infectious periods by culling status. We see that on farms that were subject to pre-emptive culling, the average infectious period is shorter than on those were identified to be infected. This is expected as the culling times are similar to censored times. The distribution of the infectious periods for the pre-emptively culled farms assigned infection times is more positively skewed than that of the farms that were confirmed to be infected. The pre-emptively culled farms with longer infectious periods were infected towards the end of the epidemic. This is partly due to the infection rate not being time dependent, but also an increase in the number of farms being culled at the end of the outbreak. For each of the pre-emptively culled farms, we estimate the probability they were infected and the probabilities are shown on the map in figure 5.4. All of the farms with non-zero probability are located in the two main infection clusters. Our results show that the transmission to the southern cluster cannot be explained by a path of shorter distance infections that were censored by preventive culling. This is consistent with

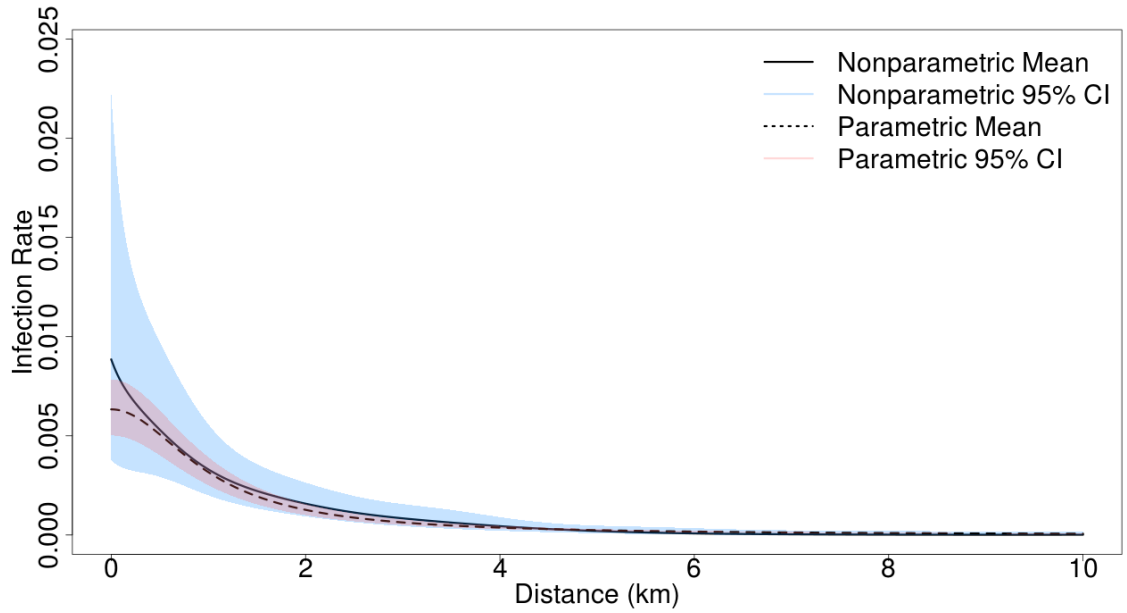
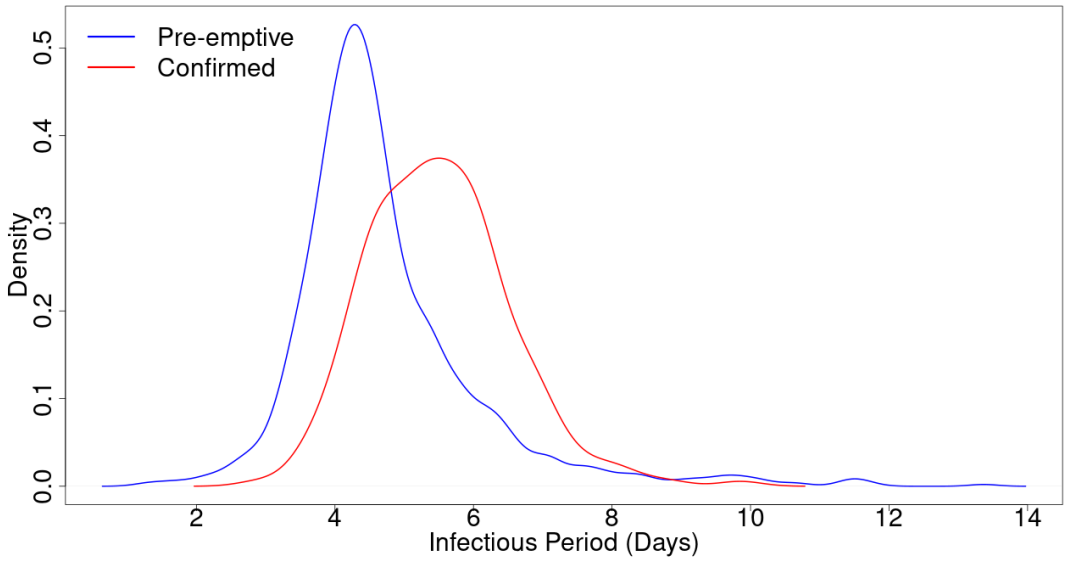


Figure 5.3: The posterior median for the nonparametric (solid) and parametric (dashed) infection rate functions for the Avian Influenza data set with unknown infection times and allowing for pre-emptive culling.

the hypothesis proposed in Bataille et al. (2011) that this long distance transmission event of avian influenza was the result of human-mediated transport of the virus. Long range transmissions require a non-zero probability for pairs of individuals that are far apart from each other. In a parametric setting, this is difficult to achieve when using an infection rate function that decays to 0. Typically, modellers include an extra non-zero shifting term to allow for long range terms and propose models of the form:

$$\beta_{i,j} = f(d_{i,j}) + \epsilon,$$

where $f(x) \downarrow 0$ as $x \to \infty$ and $\epsilon > 0$. Using our Bayesian nonparametric method, we do not need to this as the infection rate function learns the non-zero infection rate from the observed long range transmissions.



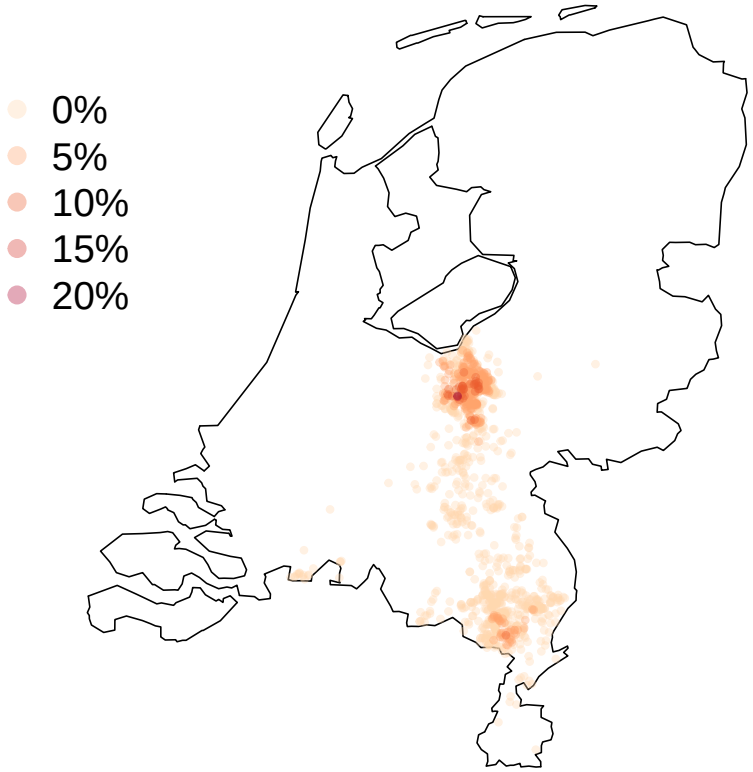


Figure 5.4: Top: the distribution for how long farms remained infectious for. Bottom: the probability each pre-emptively culled farm was infected. Farms which were susceptible throughout the outbreak are not included.

5.2.8 Culling Strategies

We can use these results to investigate different intervention strategies. Intervention strategies are used to control and limit the spread of the disease and there are two commonly used strategies.

The first is to cull all farms within a specified radius of an infected farm, this is sometimes referred to as stamping out or depopulation. When implemented successfully, this method can quickly and effectively stop the spread of the disease, but can be expensive as governments typically pay out compensation for culled animals. It can also be difficult to implement as it needs quick and accurate diseases tests, and can generate negative publicity and headlines for both the farming industry and the government. This therefore involves balancing the two effects, culling too many farms can result in increased costs, damage to the poultry industry and be unnecessary, whereas culling too few farms can be a public and animal health risk. The second option is to vaccinate farms against the disease. This option is low cost as compensation is not required, and better received by the public, but takes longer to implement. Once the disease has been identified, farmers can add vaccination feed to the animals usual feed and after approximately a week the animals will be vaccinated. Considering the timescale of the Avian Influenza outbreak, this strategy cannot respond quickly enough. For this reason, we only investigate the effects of culling.

5.2.8.1 The Posterior Predictive Distribution

To simulate the effect of culling, we sample from the posterior predictive distribution for the infection and removal times. Given the observed removal times, and the posterior distributions of g, γ and κ, we wish to generate new infection times $\mathbf{i}^*$ for all individuals and corresponding removal times $\mathbf{r}^*$. We do this using the posterior predictive distribution, which is given by:

$$\pi(\mathbf{i}^*, \mathbf{r}^* | \mathbf{i}, \mathbf{r}) = \iiint \pi(\mathbf{i}^*, \mathbf{r}^* | f, \gamma, \mathbf{i}, \kappa, i_\kappa, \mathbf{r}, \mathbf{d}) \pi(g, \gamma, \kappa | \mathbf{i}, \mathbf{r}, \mathbf{d}) \mathrm{d}f \, \mathrm{d}\gamma \, d\kappa.$$

Total Number of Infected Farms (I)	Maximum Number of Farms Culled per day	Proportion of Culling Radius Implemented
$I \leq 33$	0	0
$33 \leq I < 54$	3	$\frac{1}{2}$
$54 < I$	6	1

Table 5.4: The number of infected farms and the corresponding culling radius and maximum number of farms culled per day.

To generate samples from this distribution, we generate sample from the posterior distributions for g and γ and simulate an outbreak of avian influenza. We fix the first infected farm, κ, to be the first culled farm in the observed outbreak. To simulate culling, we assume that once an infected farm reaches the end of its infectious period and enters the removed class all farms up to $r$km away are simultaneously culled and enter the removed class. Culling cannot start early in the outbreak as it may take time for the authorities to be notified and put measures into place. Furthermore, whereas previous work (e.g. Backer et al. (2015)) has used the date after the initial infected to initiate the culling measures, we allow for stochasticity in the disease take-off and assume culling takes place once a certain number of farms have been infected. As the resources may not immediately be available to the authorities, authorities may not be able to cull all farms within $r$km and we simulate this by fixing a maximum number of farms the can be culled per day. We then increase this number over the course of the outbreak as the authorities have more available resources. The numbers are given in table 5.4 and are based on the number of farms we estimate to have been infected in the observed outbreak. Similarly, we assume the authorities will not have sufficient resources to cull all farms within the chosen radius at the start of the outbreak and we model this by assuming farms within a radius half as large are culled initially.

5.2.8.2 The Posterior Predictive Distribution as an Assessment of Model Fit

As we are using the posterior predictive distribution to simulate outbreaks, we can use these simulations to asses how well our model is fitting the data. We use the removal curves as a measure of fit, and the removal curves from simulated data sets to the curve from the observed data. The removal curve is computed by

$$R(t) = \sum_{j=1}^{n} 1_{r_j<t},$$

which computes the cumulative number of removed individuals up to time t. If our model provides a good fit to the data, we should seems removal curves generated from the posterior predictive distribution being similar to the observed removal curve. Although there are pre-emptively removed individuals in the data set, we only consider confirmed infected individuals in the removal curve. This is because the culling strategy used in the outbreak is not clear and so we cannot simulate this exactly. We instead simulate 20 outbreaks using a 5km culling radius and the constraints in table 5.4. The number of simulations is small as in order to compare similar outbreaks we condition on the final size being within 10% of the observed final size (233). This increases the time required significantly.

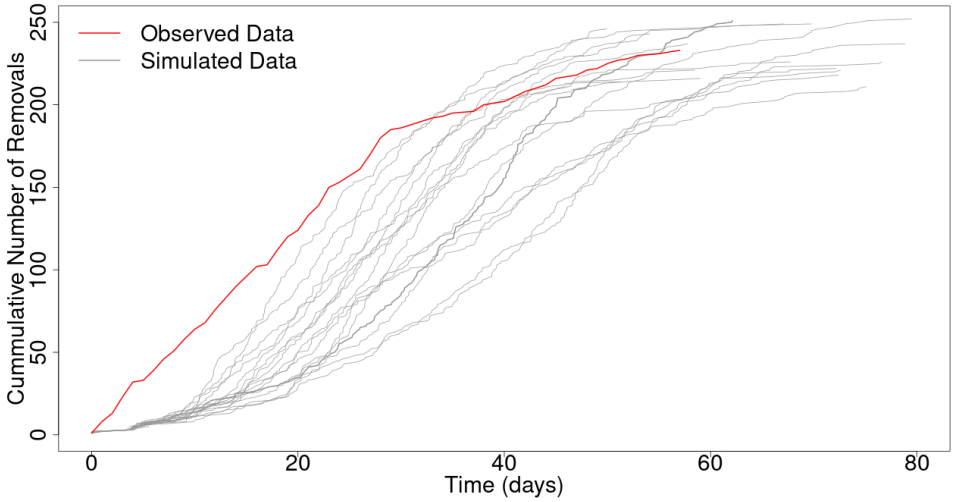
The removal curves are shown in figure 5.5. In figure 5.5(a) we display the true removal curves and it is clear that the simulated outbreaks do not match the observed outbreak. This is due to the difference in times for the outbreaks to take off and become serious epidemics. This difference occurs for two reasons. The first is that we are modelling a stochastic process and so the time for each outbreak to take off will be different. The second reason is that in the observed outbreak the authorities were not aware the disease was spreading at the start of the outbreak, so there are a considerable number of removals recorded on the day when they became aware of the outbreak. As we are not interested in comparing take off times, merely the course of the epidemic, we perform time-shifting on the simulated removal curves. To do this, we minimise the sum of squares between each simulated removal curve and the observed removal curve. This method is being developed by G. Aristotelous (private

communication). The shifted simulated removal curves are shown in figure 5.5(b) and we can see that they match the observed curve, and the observed outbreak is fairly typical of samples from the posterior predictive distribution. Considering the shifted removal curves, the main difference in the curves is the start of the outbreak. In the simulated removal curves we see a period where the number of cases increases gradually, however we do not see this in the observed outbreak. This is due to how the data was collected, as at the start of the outbreak, the authorities were not aware of the outbreak, so were not collecting data. This effect has been minimised by the time-shifting, but is still somewhat evident. After this initial period, both the observed and simulated removal curves, show a fairly linear increase in the number of removals, before reaching a critical point, where the number of removals decreases considerably. Comparing these two curves, we are more confident that our Bayesian nonparametric method gives a good fit.

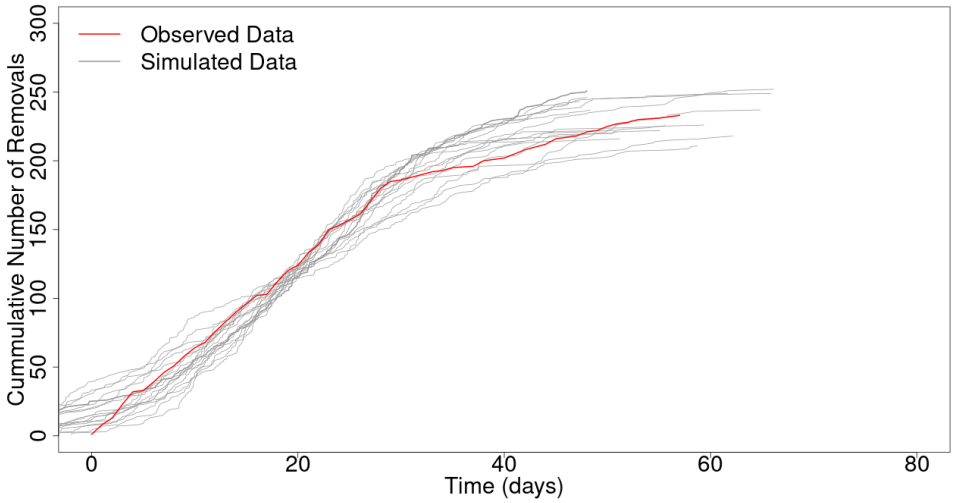
5.2.8.3 Analysis of Culling Strategies

To investigate the economic consequences of these strategies, we assume each farmer is compensated for their culled livestock. The value of the compensation depends of the type of bird culled, the number of birds culled, their age in weeks and, for turkeys, their gender. We follow Backer et al. (2015), and use the approximate rates shown in table 5.5. We acknowledge this method is crude and does not take into account any of the wider economic impacts. However, it will allow us to simulate the number of farms that are infected, the number of farms that are culled, and the compensation paid to farmers. These three values can be used to compare the risk to public health, the impact of the poultry industry, and the cost to the authorities.

Table 5.6 shows the results of the culling strategies for radius between 0 km and 5 km. A culling radius of 0 km is when the authorities take no action. It is clear that taking any course of action is better than taking none, however we also see that more ambitious strategies show little gain in reducing the median number of farms infected in an outbreak. The effect of culling at larger radii results in a larger number of culled farms and a higher amount of compensation, but does not result in a considerable



(a) The removal curves for the simulated outbreaks (grey) and the observed outbreak (red).



(b) The shifted removal curves for the simulated outbreaks (grey) and the observed outbreak (red).

Figure 5.5: Removals curves constructed from samples from the posterior predictive distribution.

Poultry Type	Compensation (€per bird)
Broiler	0.98
Duck	2.09
Turkey	10.63
Layer	2.05

Table 5.5: Estimates of compensation per bird paid to farmers during the Avian Influenza outbreak. The estimates were obtained from table 2 in Backer et al. (2015).

Radius (km)	Total Number of Infected Farms	Total Number of Culled Farms	Compensation (€millions)
0	443 (151, 644)	443 (151, 644)	24.8 (8.62, 35.9)
1	297 (110, 535)	489 (215, 709)	27.2 (12.2, 38.9)
2	283 (108, 608)	488 (217, 740)	27.5 (12.2, 41.7)
3	283 (112, 582)	517 (242, 775)	29.0 (13.2, 43.1)
4	274 (105, 564)	512 (228, 793)	28.5 (12.3, 43.9)
5	280 (109, 549)	527 (226, 797)	39.2 (12.4, 41.9)

Table 5.6: Posterior predictive medians for the number of infected and culled farms and the amount of compensation paid.

reduction in the number of infected farms. This is because the maximum number of farms culled per day is quickly reached, even for small culling radii. In the data set, the average density of farms was approximately 2 per km^2, whereas a culling radius of 2 km covers over $12km^2$.

These results are broadly in line with those of Backer et al. (2015), who also suggest that larger culling radii do not result in a considerable reduction in the number of infected farms. However, as we use a much smaller estimate for the maximum number of farms culled per day, we do not find a large difference between culling radii of 1 km and 2 km.

5.2.9 Discussion

Our main conclusion is that it is possible to model the spatially-heterogenous infection rate for epidemic diseases nonparametrically and that GPs provide a flexible framework for doing so. This nonparametric methodology allows us to reduce the need for strict parametric assumptions, which are made for mathematical convenience rather than any characteristics in the data. We were also able to incorporate current MCMC and data augmentation methods alongside our nonparametric approach to allow for missing data without making assumptions about which farms were infected.

The method however requires more time and computational power than the standard parametric methods, especially as we are using an MCMC method. We have taken steps to alleviate these issues by utilising the MPA approximation method, which reduces the dimension of the covariance matrix. This fits a GP onto a smaller set of uniformly spaced pair-wise distances and then projects this onto the complete set of pair-wise distances.

Concerning the Avian Influenza data set, our methodology has allowed us to approach the infection process in a more flexible way than previous methods. Our estimates are in line with previous work and combining this method with previously developed MCMC techniques and data augmentation allows us to analyse this data set in more detail than has previously been possible, including determining whether pre-emptively culled farms had been infected. The uncertainty around our estimates is much larger than that of previous parametric methods. However, we believe we are better quantifying the uncertainty. As parametric models we use strict assumptions, the uncertainty is the results is often a lot smaller compared to Bayesian Nonparametric methods. Underestimating this uncertainty can result in misplaced confidence in results, particularly in predictions, which many then be taken into account by practitioners and policy makers. We were able to use the posterior predictive distribution to analyse the effect of different control strategies which can be used to inform policy in this area. Using the posterior predictive distribution also allowed us to exploit the larger uncertainty in our estimates. Previous methods used asymptotic properties of Maximum Likelihood Estimates and used quantiles of

this distribution to simulate forward and make predictions. The posterior predictive distribution means we can explore the entire posterior distribution of the model parameters, including the better quantified uncertainty around these estimates.

For this outbreak, we only considered the spatial heterogeneity. It is possible that the number of animals or the type of animals on the farms impacted the infection rate. Given sufficient data, further work could build a model which uses this data as covariates. One way of doing this would be considering each covariate as a separate dimension of the GP. We also used the squared exponential function throughout to analyse this data set, and we could allow for a non-stationary infection rate function by considering different covariance functions.

5.3 Foot and Mouth Disease

Foot and Mouth Disease (FMD) is a viral infection affecting a wide variety of animals, including cattle, sheep and pigs. Infected animals can suffer from a loss of appetite followed by blisters in the mouth and on the legs. They may also have a high fever and the disease can be fatal in some cases (Alexandersen et al., 2003). In 2001 there was a large outbreak of FMD among sheep and cattle farms in the UK and over the course of seven months over six million animals were infected. The outbreak affected farms in Cumbria and Devon, and in this section we consider the outbreak in Cumbria.

In Cumbria, a county in the north-west of the UK, there were 5,436 farms consisting of: 1,061 sheep farms, 1,064 cattle farms, and 3,253 farms with both sheep and cattle. Of these farms, 1,021 were infected including 8% of the sheep farms, 13% of the cattle farms, and 24% of farms where both sheep and cattle were present.

FMD can be spread through several mechanisms. We outline the four mechanisms given in Alexandersen et al. (2003). The first is direct contact between infected and susceptible animals, this can occur when animals are moved between farms. The second is feeding contaminated feed to susceptible livestock. Thirdly, it is possible for humans to contribute to the spread of the disease by carrying the disease from an

infected farm to a susceptible farm. This can include carrying the virus on clothing or machinery between farms. The final mechanism is for the virus to be carried for short distances on the wind, but this is only under certain weather conditions and is unlikely to occur over long distances.

5.3.1 Literature Review

The data set has been studied extensively from a parametric standpoint. In Keeling (2001), the authors proposed a parametric model where the infection rate between farms depended on the Euclidean distance between them as well as the number of sheep and cattle on each of them two farms. They estimated the model parameters using a least squares method. In Diggle (2006), the author proposed a similar parametric model and used a partial likelihood approach to reduce the computational complexity for the inference.

A network-based approach was proposed in Ferguson (2001). In this model, a susceptible farm receives infectious pressure from local neighbouring infected farms and a lower level of pressure from farms that are further away. This was based on a two-level mixing model proposed in Ball et al. (1997). The model in Ferguson (2001) also depends on which type of animals were on the farm.

In Kypraios (2007, §.3) and Jewell et al. (2009), a Bayesian approach was adopted and the authors proposed a parametric model consisting of. a spatial kernel and components for the number of sheep and cattle on each farm. They assigned prior distributions to the model parameters to infer their values and used a data augmentation technique to infer the unobserved infection times. In Stockdale et al. (2019), the author used the same parametric form proposed in Kypraios (2007, §.3), but implemented a likelihood approximation method. This data set has not previously been analysed using Bayesian nonparametric methods. This outbreak has been previously analysed parametrically using a Bayesian framework in Probert et al. (2018), and the authors use an individual-level model where the infection rate between any two farms depends on the distance between them, as well as the number of sheep, cattle and pigs on both farms. The authors also take into account the ban on movements

between farms the authorities introduced to control the spread of the disease. The authors develop methods for making real time decisions about control strategies. Methods for performing real time analysis for this outbreak are also developed in Welding and Neal (2019) using a Sequential Monte Carlo approach.

In contrast to previous work, we will estimate the infection rate functions non-parametrically, in particular describing how the relationship in susceptibility between farms with one type of animal and farms with two types of animals differ. This has previously been done by including a parameter in the model which compares the susceptibility of sheep to the susceptibility of cattle, and assuming the spatial kernel is the same for all types of animals. We use our multi-type models to allow these kernels to be dependent and learn how they are related.

5.3.2 Data

The data set we used was previously used in Kypraios (2007), Jewell et al. (2009) and Stockdale et al. (2019). Kypraios had obtained from the Department for Food and Rural Affairs (DEFRA), however it is no longer publicly available. The data consists of the location of all 5,436 sheep and cattle farms in Cumbria and the surrounding areas, their status at the end of the infection, and which types of animals were on the farms. For farms that were infected, we have the date on which the animals were culled. The animals on a small number of farms were culled without their infection status being known. We follow Kypraios (2007) and assume they were infected. A map with the locations of the farms in shown in figure 5.6.

5.3.3 Single Type Model

We propose modelling the infection rate from farm j to farm k by a function of the Euclidean distance between them:

$$\beta_{j,k} = \beta(d_{j,k}).$$

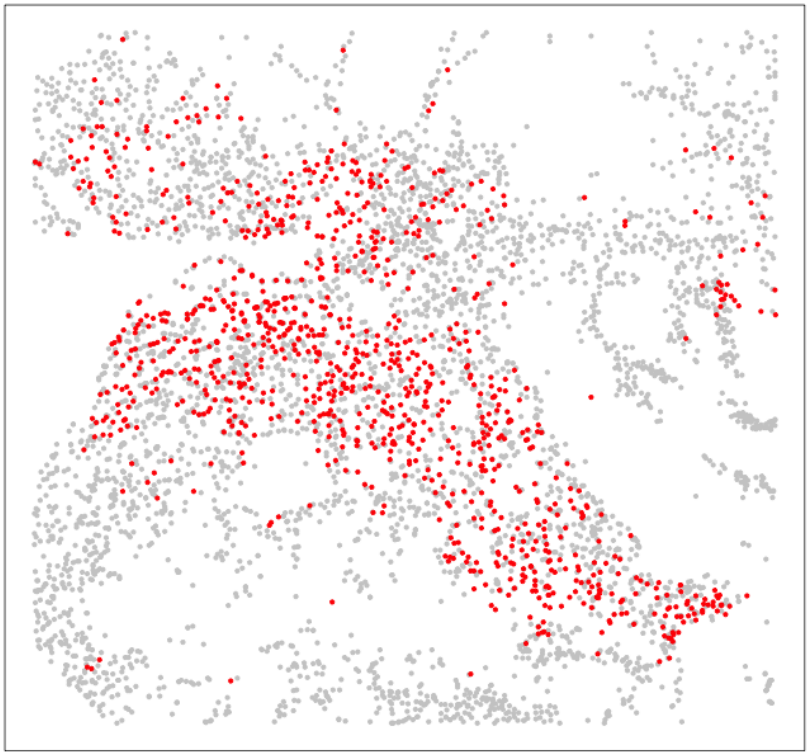


Figure 5.6: A map of the farms in Cumbria, during the outbreak. Grey farms were not infected with the virus, red farms were infected with the virus.

We model the function β nonparametrically through the function f using a transformation function to ensure it is positive:

$$\beta = \exp(f), \quad f \sim \mathcal{GP}(0,\, \Sigma), \quad \Sigma_{j,k} = k(d_j, d_k; \alpha, l).$$

We place a vague exponential prior distribution on l such that

$$l \sim \text{Exp}(0.01).$$

We assume the infectious period distribution is given by a $\Gamma(\lambda, \gamma)$, where λ is fixed and known. We place a vague exponential prior distribution on γ such that

$$\gamma \sim \text{Exp}(0.01).$$

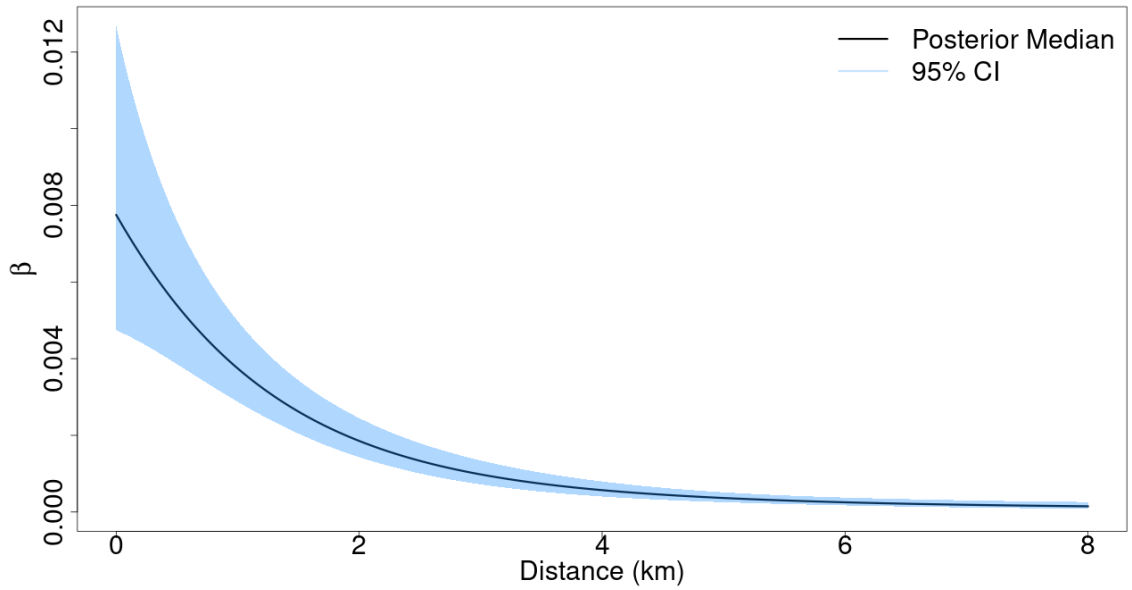


Figure 5.7: The posterior median and 95% credible interval for the infection rate function for the FMD dataset.

As we are inferring plausible values for the length scale parameter, we fix the infection times. This reduces the computational time required. We follow the results of Jewell et al. (2009) and (Stockdale et al., 2018, §4.4.2) and assume once infected, each farm remains so for $7\frac{1}{2}$ days.

We run the MCMC algorithm (algorithm 5) for 20,000 iterations, removing the first 5,000 iterations as a burn-in period. This took approximately seven days. The results for the infection rate function are shown in figure 5.7. The infection rate function is a decreasing function of distance and we estimate the infection rate between farms further than 7km apart to be negligible. We see from the credible interval that the function could have several shapes. The upper bound of the credible interval decays roughly exponentially, whereas the lower bound does not. Instead, it is similar to a logistic function. This analysis is not possible with parametric methods as we must specify the exact form the of function. For the length scale parameter, the posterior median is 8.53 km (95% CI: (8.02, 8.93)).

We now compare our results with those in Jewell et al. (2009). The authors

assume the spatial component is the same for all types of farm and we therefore compare their results to our method with one type. The spatial element of the model in Jewell et al. (2009) is given by:

$$f(d_{ij}) = \frac{\beta_0}{d_{ij}^2 + \beta_0^2}.$$

The comparison is shown in figure 5.8. We show the relative infection rate functions, which are the infection rate functions normalised by their values for immediate neighbours, i.e. $\beta_{\text{relative}} = \frac{\beta}{\beta(0)}$. We see the results are broadly similar, with our result having a heavier tail. In Kypraios (2007, §3.3.2), where the parametric results are also presented, the author justifies their choice of infection rate function by stating the need for a function with a heavy tail without including extra parameters. Although this choice of functions fulfils these requirements, as it only has a single parameter, the scale of the function depends on the shape of the tail. As we have used a Bayesian nonparametric method, we can avoid assuming the type of tails we require and how to parameterise the infection rate function. This is an advantage over parametric methods as we can estimate the scale of the function independently of the shape of the tail, and we do not need to propose of shape of the tail at all.

5.3.4 Multi-Type Model

Using our MOGP models, we fit a multi-type model with three types; sheep-only farms, cattle-only farms, and farms with both sheep and cattle. Due to the substantially different number of farms in each type, we normalise the functions by the number of farms in each type and fit a density dependent model. This is because there are approximately three times farms with two types of animals as single type farms. The susceptibility model is given by:

$$\beta_{j,k} = \begin{cases} \frac{1}{N_1} \exp\left(f^{(1)}(d_{j,k})\right) & \text{if } k \text{ is a sheep-only farm} \\ \frac{1}{N_2} \exp\left(f^{(2)}(d_{j,k})\right) & \text{if } k \text{ is a cattle-only farm} \\ \frac{1}{N_3} \exp\left(f^{(3)}(d_{j,k})\right) & \text{if } k \text{ has both sheep and cattle,} \end{cases}$$

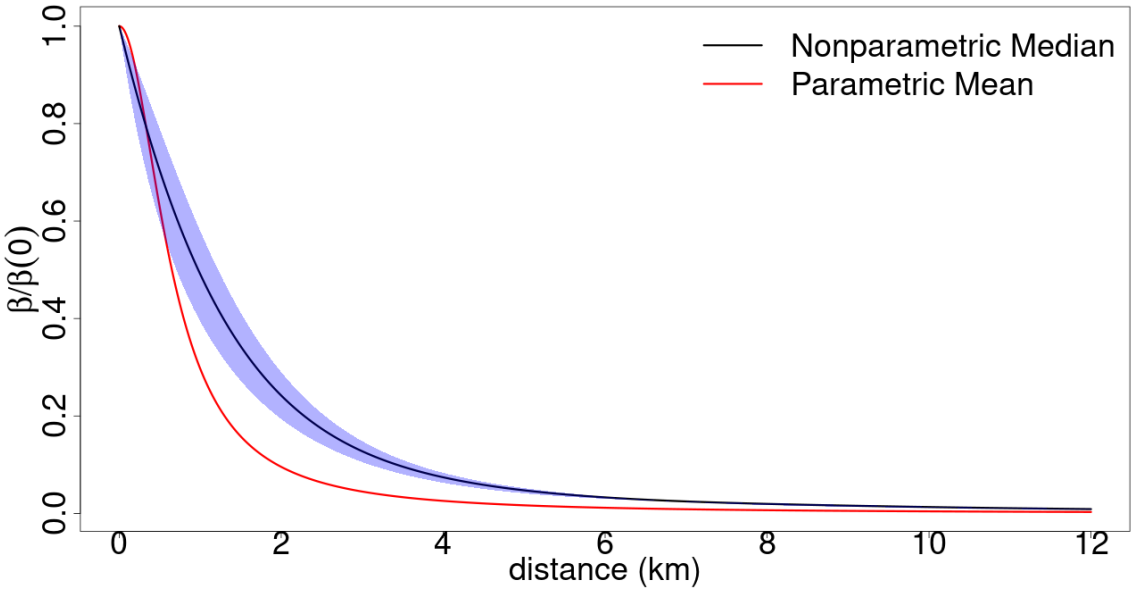


Figure 5.8: The relative infection rate function for the Bayesian nonparametric method (black with blue credible intervals) and the parametric results from Jewell et al. (2009).

where $N_1 = 1061$, $N_2 = 1064$ and $N_3 = 3253$. We implement both the MOC model and the DBM and infer the model parameters.

5.3.4.1 Multi-Output Covariance Model

For the MOC model, we place a joint prior distribution on the functions $f^{(1)}$, $f^{(2)}$, and $f^{(3)}$. In the MOC model, we can control the correlation between the infection rate functions. Initial runs of this model show that the infection rate functions for sheep-only and cattle-only farms have higher correlation than that of sheep-only farms and cattle-only farms with farms with both sheep and cattle. We therefore specify two correlation parameters and use the following prior distribution:

$$\begin{pmatrix} f^{(1)} \\ f^{(2)} \\ f^{(3)} \end{pmatrix} \sim \mathcal{GP}\left(0, \begin{pmatrix} \Sigma^{(1,1)} & \rho_1\Sigma^{(1,2)} & \rho_2\Sigma^{(1,3)} \\ \rho_1\Sigma^{(2,1)} & \Sigma^{(2,2)} & \rho_2\Sigma^{(2,3)} \\ \rho_2\Sigma^{(3,1)} & \rho_2\Sigma^{(3,2)} & \Sigma^{(3,3)} \end{pmatrix}\right),$$

where $\Sigma^{(i,j)}$ is the covariance matrix for types i and j. We assume the correlation parameters are positive and place identical and independent uniform prior distributions on both correlation parameters such that:

$$\rho_1 \sim U[0,1] \quad \text{and} \quad \rho_2 \sim U[0,1].$$

For the label of the first infected farm, we place a discrete uniform prior distribution over the labels of the infected farms, and given this label the prior distribution of the initial infected time is

$$i_\kappa|\kappa = -z, \quad z \sim \text{Exp}(0.01).$$

It is negative as it must occur before the first removal, which occurs at time $t = 0$. We place a vague, conjugate exponential prior distribution on the infectious period distribution rate parameter, such that $\gamma \sim \text{Exp}(0.01)$. As inferring both the infection times and length scale parameter is computationally expensive, based on the inference for the single type model, we fix the length scale parameter to be $l = 8.5$km. This is justified as we used sensible estimates for the infection times, so we do not expect the infection rate function to be considerably different when inferring the infection

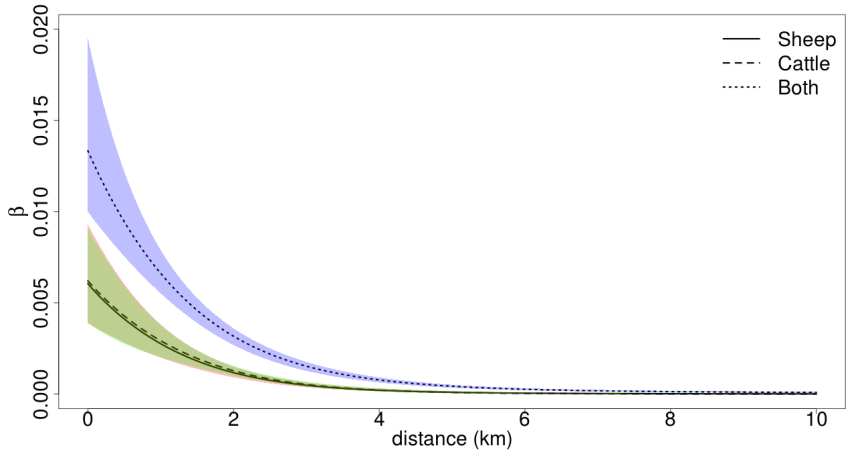
times. The posterior distribution is therefore given by

$$\begin{aligned}
&\pi(f^{(1)}, f^{(2)}, f^{(3)}, \rho_1, \rho_2, \gamma, \mathbf{i}, i_\kappa, \kappa | \mathbf{r}, \lambda, \mathbf{c}) \propto \\
&\quad \exp\left\{ -\sum_{j=1}^{n} \sum_{k=1}^{N} f^{(c_k)}(d_{j,k}) \left((r_j \wedge i_k) - (i_j \wedge i_k) \right) \right\} \prod_{\substack{j=1 \\ j \neq \kappa}}^{n} \left(\sum_{k \in \mathcal{Y}_j} f^{(c_j)}(d_{k,j}) \right) \\
&\quad \times \prod_{j=1}^{n} h(r_j - i_j | \lambda, \gamma) \mathcal{GP} \left(\begin{pmatrix} f^{(1)} \\ f^{(2)} \\ f^{(3)} \end{pmatrix} ; 0, \begin{pmatrix} \Sigma^{(1,1)} & \rho_1 \Sigma^{(1,2)} & \rho_2 \Sigma^{(1,3)} \\ \rho_1 \Sigma^{(2,1)} & \Sigma^{(2,2)} & \rho_2 \Sigma^{(2,3)} \\ \rho_2 \Sigma^{(3,1)} & \rho_2 \Sigma^{(3,2)} & \Sigma^{(3,3)} \end{pmatrix} \right) \\
&\quad \times \exp\{-0.01\gamma\} \exp\{0.01\kappa\}.
\end{aligned}$$

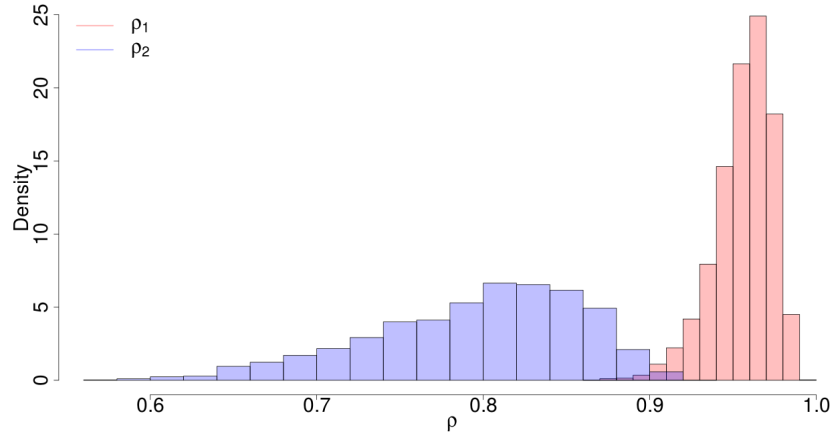
We run the MCMC algorithm outlined in algorithm 8 for 25,000 iterations, removing the first 5,000 as a burn-in period. The results in figure 5.9 show that farms with both sheep and cattle are more susceptible to contracting the disease than farms with only one type of animal. With regard to the shape of the infection rate functions, the function of sheep and cattle farms decays more quickly than the other two functions, and for farms of all types the probability of an infected farm infecting a susceptible further than 7 km away is negligible.

The MOC model allows us to asses the relationship between the infection rate functions. In figure 5.9(a) it is striking how similar the infection rate functions for sheep-only and cattle-only farms are. Figure 5.9(b) shows the correlation between these two functions is high and the 95% credible interval is (0.914, 0.982). The correlation between the functions for farms with one type of animal and the function for farms with both types of animals is not as high, but these functions are still highly correlated (95% CI: (0.652, 0.891)).

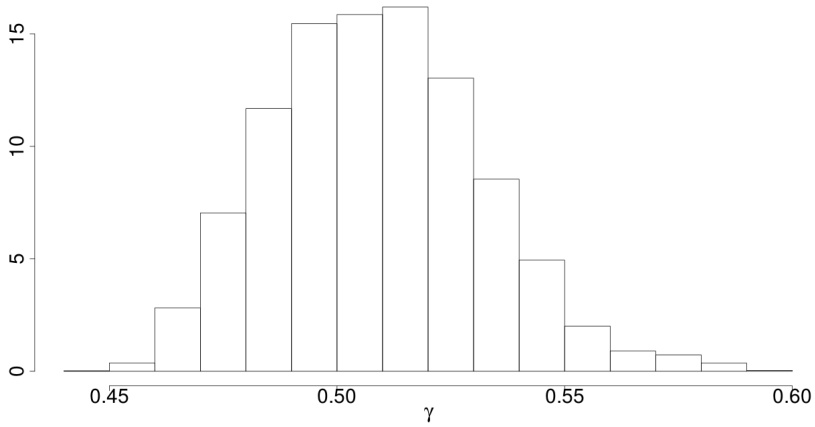
With regard to the infection times, the posterior median for the infectious period distribution rate parameter is 0.508, which gives an expected infectious period of 7.86 days. This is in line with the results in Jewell et al. (2009) and Stockdale et al. (2019).



(a) The posterior medians and 95% credible intervals for the infection rate functions.



(b) The posterior distributions for the correlation parameters ρ_1 and ρ_2.



(c) The posterior distribution for the infectious period distribution rate parameter γ.

Figure 5.9: Results of the MOC model applied to the FMD dataset.

5.3.4.2 Discrepancy Based Model

We choose the sheep-only farm infection rate function to be the baseline function, and set up the prior distributions as follows:

$$\begin{aligned}
&f^{(1)} \sim \mathcal{GP}(0, \Sigma^{(1)}), \quad \Sigma^{(1)}_{jk} = k(d^{(1)}_j, d^{(1)}_k; \alpha, l) \\
&f^{(2)} = f^{(1)} + u^{(2)}, \quad u^{(2)} \sim \mathcal{GP}(0, \Sigma^{(2)}), \quad \Sigma^{(2)}_{jk} = k(d^{(2)}_j, d^{(2)}_k; \alpha, l) \\
&f^{(3)} = f^{(1)} + u^{(3)}, \quad u^{(3)} \sim \mathcal{GP}(0, \Sigma^{(3)}), \quad \Sigma^{(3)}_{jk} = k(d^{(3)}_j, d^{(3)}_k; \alpha, l),
\end{aligned}$$

where $d^{(\tau)}_j$ is the j^{th} pair-wise distance for type τ farms. This construction allows us to compare the susceptibility of both cattle-only farms and sheep and cattle farms to sheep-only farms.

As in the MOC model, we use the following prior distributions on the remaining model parameters:

$$\begin{aligned}
&\gamma \sim \text{Exp}(0.01), \\
&\kappa \sim U[1, \ldots, n], \\
&i_\kappa | \kappa = -z, \quad z \sim \text{Exp}(0.01).
\end{aligned}$$

Again, we set the length scale parameter to be $l = 8.5$km. The posterior distribution is therefore given by:

$$\begin{aligned}
\pi(f^{(1)}, f^{(2)}, f^{(3)}, \rho, \gamma, \mathbf{i}, i_\kappa, \kappa | \mathbf{r}, \lambda, \mathbf{c}) \propto \exp &\left\{ -\sum_{j=1}^{n} \sum_{k=1}^{N} f^{(c_k)}(d_{j,k}) \left((r_j \wedge i_k) - (i_j \wedge i_k) \right) \right\} \\
&\times \prod_{\substack{j=1 \\ j \neq \kappa}}^{n} \left(\sum_{k \in \mathcal{Y}_j} f^{(c_j)}(d_{k,j}) \right) \prod_{j=1}^{n} h(r_j - i_j | \lambda, \gamma) \\
&\times \mathcal{GP}(f^{(1)}; 0, \Sigma^{(1)}) \mathcal{GP}(u^{(2)}; 0, \Sigma^{(2)}) \mathcal{GP}(u^{(3)}; 0, \Sigma^{(3)}) \\
&\times \exp\{-0.01\gamma\} \exp\{0.01 \mathbb{1}_\kappa\}.
\end{aligned}$$

We run the MCMC algorithm outlined in algorithm 9 for 25,000 iterations and remove the first 5,000 as a burn-in period. The results are shown in figure 5.10 and they are similar to the results of the MOC model. In contrast to the MOC model, we can compare the functions to a baseline. We have chosen the infection rate

function for sheep-only farms to be a baseline and we see in figure 5.10(b) compared to cattle-only farms, there is no significant difference between the functions, as their ratio is near 1. However, the infection rate function for farms with both sheep and cattle is significantly higher across all distances compared to the sheep-only infection rate function. We need to be cautious when interpreting this as in figure 5.10(a) all three infection rate functions tend to 0 and after 7km the infection rate is negligible. For example, the relative error for $\beta^{(1)} = 10^{-7}$ and $\beta^{(2)} = 10^{-8}$ is large, although the absolute error is small. This means for larger distances the discrepancy shown 5.10(b) is not as meaningful as it may seem on first impressions. However, we can say the infection rates between cattle-only and sheep-only and cattle and sheep farms and sheep-only farms are different.

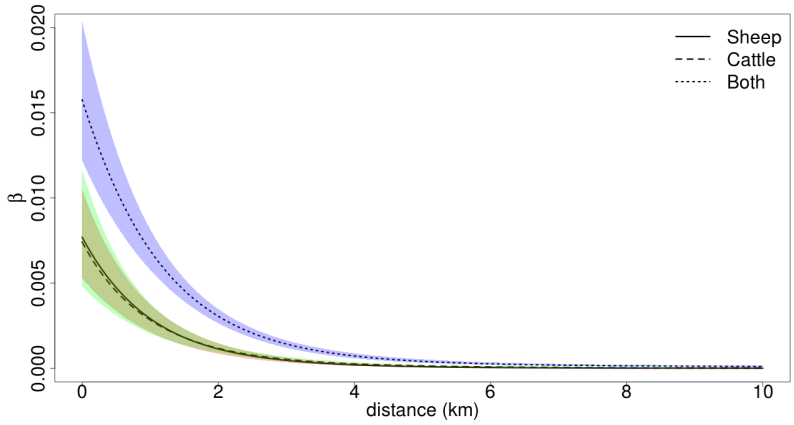
We estimate $\gamma = 0.517$ (95% CI:(0.469, 0.570)), which gives an expected infectious period of 7.74 days. This is in line with both the results from the MOC model and the results presented in Jewell et al. (2009) and Stockdale et al. (2018).

5.3.5 Discussion

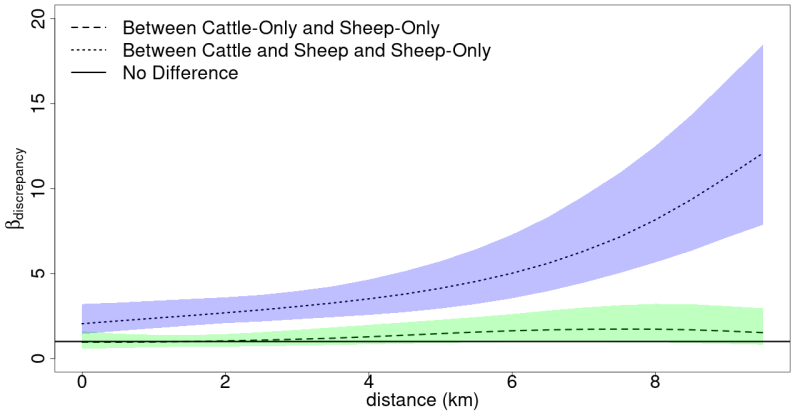
Both of the models show similar results; farms with both sheep and cattle are more susceptible to infection than those with only one type of animal on and farms with either only sheep or only cattle are equally susceptible to the disease. All farms share a similar spatial element to the infection rate function, which is a decreasing function of distance. There is little difference in the time taken for the DBM and MOC, with both taking around 5 days to run for 25,000 iterations.

The estimates for the model parameters are shown in table 5.7. We show the value of the infection rate functions for immediate neighbours that are 0km apart. Again, we see the estimates from both models are similar and the 95% credible intervals largely overlap.

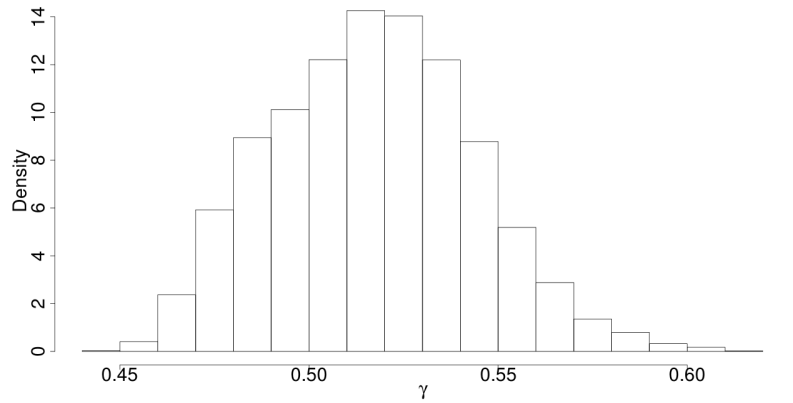
This dataset shows how the models are suitable for different audiences. The MOC model allows us to understand the correlation structure for the GPs. This is interesting from a mathematical standpoint. However, famers, the authorities and other practitioners are likely to be interested in a direct comparison of the infection



(a) The posterior medians and 95% credible intervals for the infection rate functions.



(b) The computed discrepancy between the sheep-only and cattle-only infection rate functions, and the sheep and cattle and sheep-only infection rate functions. The horizontal line at represents no difference.



(c) The posterior distribution for the infectious period distribution rate parameter γ.

Figure 5.10: Results of the DBM applied to the FMD dataset.

Parameter	MOC Estimate	DBM Estimate
γ	0.508 (0.469, 0.557)	0.517 (0.469, 0.570)
$\beta^{(0)}(0)$	6.05×10^{-3} (3.89×10^{-3}, 9.3×10^{-3})	7.77×10^{-3} (5.28×10^{-3}, 1.1×10^{-2})
$\beta^{(1)}(0)$	6.2×10^{-3} (3.92×10^{-2}, 9.07×10^{-2})	7.42×10^{-3} (4.8×10^{-3}, 1.17×10^{-2})
$\beta^{(2)}(0)$	1.33×10^{-2} (1.00×10^{-2}, 1.95×10^{-2})	1.58×10^{-2} (1.20×10^{-2}, 2.03×10^{-2})
ρ_1	0.95 (0.941, 0.982)	- -
ρ_2	0.806 (0.652, 0.891)	- -

Table 5.7: The posterior median values and 95% credible intervals for the model parameters for the Foot and Mouth disease dataset.

rate functions. The DBM allows us to do this in a presentable and concise fashion.

We can also see the advantages of using a multi-type model over model with an infection rate function that only depends on the distance. Not only can we analyse the difference between the various types of farms, but we risk fitting a misspecified model when assuming the infection rate function is the same between types. When fitting the single covariate model, we estimate the infection rate between immediate neighbours to be 0.00798 (95% CI: (0.00478, 0.0127)) farms per day. This falls in between the estimates for the different types in the multi-type model as some kind of average for the infection rate for each type. Using the multi-type model, we are better able to analyse the infection rate for each type of farm and quantify their susceptibility to the disease.

There is a subtle difference between the analysis presented here and that of Jewell et al. (2009) and Stockdale et al. (2019). Previous analyses have investigated the

relative susceptibility of cows to sheep, whereas the our analysis with the DBM describes the relative susceptibility of cattle-only farms to sheep-only farms, as well as farms with both cattle and sheep to sheep-only farms. Our analysis assumes the susceptibility of a farm does not depend on the number of animals present on the farm. However, our model allows the spatial component to be distinct for each type.

5.4 Conclusion

We have applied our methods to two datasets. The first was a large outbreak of Avian Influenza in the Netherlands. There were several ways in which our analysis was novel. The first was the use of Bayesian nonparametric methods. Our method provided a flexible framework for modelling the infection rate function and, compared to previous parametric methods, allowed us to fully capture the uncertainty around the function. Using this uncertainty, we then analysed various culling strategies. This had been done previously using the asymptotic distribution of maximum likelihood estimators. We used the posterior predictive distribution of our Bayesian nonparametric method, which gives a richer analysis. Our method showed that culling is in principle an effective strategy, but the high density of farms in some areas can cause problems. Using a data augmentation method allowed us to analyse the infection times and infection status of pre-emptively culled farms, which has not previously been investigated. We were able to estimate the probability each pre-emptively culled farm was infected, and we concluded the pre-emptively culled farms had on average a shorter infectious period length than farms which were confirmed to be infected.

Finally, we applied our methods to an outbreak of Foot and Mouth disease in Cumbria, UK. We were successfully able to use our multi-type method to infer the infection rate functions for cattle-only farms, sheep-only farms, and farms with both sheep and cattle. This allowed us to compare the functions and understand how they were correlated. We showed the infection rate functions for sheep-only farms and cattle-only farms were similar to each other and that farms with both sheep and cattle were much more susceptible to the disease. The dataset also demonstrated the

difference in our two models. The output for the DBM is more suited to practitioners as we can clearly show the difference in susceptibility of the different types of farms. The MOC model allows us to analyse the correlation between the infection rate functions.

In this chapter, we have shown that we can successfully apply our methods to datasets and we have produced new results. Our Bayesian nonparametric methods can be implemented alongside existing data augmentation methods to better understand the spread of a disease. Using a Bayesian framework also has advantages as we are able to make use of the posterior predictive distribution to simulate outbreaks and analyse culling strategies.

CHAPTER 6

Conclusion

In this thesis, we have developed Bayesian nonparametric methodology for inferring infection rate functions for individual-level stochastic epidemic models. Our new methods allow us to infer infection rate functions for individual-level stochastic epidemic models without making strict assumptions about the parametric forms or choosing forms which lack any biological or epidemiological justification. Our methodology can be combined with existing data-augmentation Bayesian methods for inference for stochastic epidemic models such as inferring the infection times of individuals and the parameters of the infectious period distribution. The methodological advancements in this thesis can be split into three parts. We first developed a method for inferring infection rates, where the infection rate from one individual to another can be modelled as a continuous function of a characteristic of the relationship of the individuals. We then developed a method for nonparametrically modelling infection rate functions for multi-type epidemics. The final methodological advancement came through extending our nonparametric method to modelling infection rates where the infection rate from one individual to another can be considered as a continuous function of any number of continuous variables. We then demonstrated our methodology on two real data

sets, one of Avian Influenza in the Netherlands and the other on Foot and Mouth Disease in the UK. Using our Bayesian nonparametric methods, we were able to provide new analysis and insight into these outbreaks.

6.1 Main Findings

In chapter 2, we considered stochastic epidemic models, where the infection rate between any two individuals could be modelled as a function of the relationship between the individuals, for example the Euclidean distance between the individuals. Methods for these models had been exclusively parametric (see e.g Boender et al., 2007; Jewell et al., 2009). Using a parametric framework means we have to make assumptions about the infection rate function, which are arbitrary as we do not directly observe the infection process. To avoid making such strict assumptions, we developed a Bayesian nonparametric method and placed a Gaussian process prior distribution on the function modelling the infection rate. We then developed an MCMC algorithm to infer the infection rate function alongside the infectious period distribution parameter and the times at which individuals were infected. One difficulty with using the Gaussian process prior distribution is setting the prior distribution hyperparameters, so we place a prior distribution on the length scale hyperparameter and infer this alongside the other model parameters. Another challenge with both the Gaussian process prior distribution and MCMC algorithms is the computational complexity. To overcome this, we developed an approximation method based on the Deterministic and Fully Independent Conditional Approximation (Rasmussen and Williams, 2006, §8) called the Mean Projection Approximation. This fits a Gaussian Process prior distribution to a function over a pseudo data set and then projects the function onto the full data set. Furthermore, we allowed more assumptions to be included the model, such as assuming the infection rate function is monotonic by incorporating a method described in Riihimäki and Vehtari (2010) or assuming the function has some asymptotic behaviour. We demonstrated our methods with several simulation studies, which showed our method yields good results, but can be time

consuming.

Chapter 3 concerned multi-type epidemics, where the infection rate from one individual to another depends on some continuous variable describing the relationship between them as well as the type of individual being infected. This can be used to model outbreaks of diseases such as Avian Influenza, where the type of birds on the farm may influence how susceptible they are to the disease. We developed two sets of methods for modelling these types of diseases in a Bayesian nonparametric framework. The first method was a fixed-effects model, where the function governing the infection rate is the same across all types, but the scale is different. To avoid assuming the infection rate function is identical between types, we introduced Multi-Output Gaussian Process. This is a way of modelling several functions simultaneously and allowing a correlation structure between them. The first method we introduced was the Multi-Output covariance model, where we allow the functions to be correlated. The second method we implemented was the Discrepancy Based model, where we chose one type to be a baseline and compute the discrepancy for each type based on the baseline type. These models allow us to learn more information from the data by describing how the infection rate functions for the different types are related to each other. We demonstrated our methods through several simulation studies.

In chapter 4, we developed a Bayesian nonparametric framework for modelling heterogeneously mixing epidemics where the infection rate between any two individuals can be considered as a function of more than one continuous variable describing the relationship between them. For example, an outbreak of a disease where the infection rate from one individual to another is a function of the distance between them and the size of the individual being infected. We extended the model developed in chapter 2 to allow for n dimensional functions. We also developed methods where we model the effect of the variables separately. This framework allowed us to make more assumptions about the form for the infection rate functions. This is often needed due to the limited data we observe during the outbreak of an epidemic. This method may also be useful to practitioners as they can develop disease control strategies based on each variable.

In chapter 5, we turned our attention to real data sets. We analysed two data sets. The first was an outbreak of Avian Influenza that occurred in the Netherlands in 2003. This data set had previously been analysed using parametric methods (e.g. Boender et al., 2007). We analysed this using Bayesian Nonparametric methods, which, to our knowledge, has not previously been done. Our results were broadly similar to the parametric results with some important differences; we were better able to quantify the uncertainty around the estimate for infection rate function. We were also the first to investigate the role of pre-emptively culled farms, in particular which farms were infected with the virus. Culling strategies for this outbreak have been investigated using maximum likelihood methods (Backer et al., 2015), and we were able to produce richer results for culling methods using the posterior predictive distribution. This used the full distribution from our nonparametric model, including the wider uncertainty regions. We also then investigated an outbreak of Foot and Mouth disease from the UK in 2001. This outbreak had not been previously investigated from a nonparametric standpoint, and we were able to use our Multi-Output Gaussian process methods to analyse the susceptibility of farms with different animals.

6.2 Limitations and Further Work

Our method is not without limitations, and we now discuss three limitations we encountered: lack of data, long runtime compared to parametric methods, and accessibility of Bayesian Nonparametric methods. We require a large amount of data for the method to return suitable results. This is due to us making fewer assumptions compared to equivalent parametric models and the model being more flexible. This can pose difficulties when implementing the multi-dimensional methods discussed in chapter 4. The lack of informative data also affects the multi-type models we can use in chapter 3. We considered susceptibility multi-type outbreaks, where the infection rate depends on the type of susceptible individual being infected. This type of outbreak is often more evident in the data than infectivity models as we observed different proportions of each type becoming infected. In an infectivity model, the data

is less informative about the outbreak as we do not observe the infection process and cannot determine who infected whom easily. Therefore, further work could be done to improve methods for determining which individuals were more infectious than others. To improve the results we could include a parametric mean function in the GP prior distribution and reduce the prior variance making the distribution more informative. In Rasmussen and Williams (2006, §.7), the authors discuss incorporating explicit mean functions and how these can be learned as data becomes available.

The second limitation was competitiveness compared to parametric methods regarding runtime. One of the bottlenecks in the nonparametric MCMC algorithm was inferring length scale parameters, as this involves decomposing and inverting the covariance matrix repeatedly. Future work can look at more efficient computational methods for this. In this thesis, I used the GCC compiler and the GSL BLAS libraries, but other alternatives, which may be more efficient, include the Intel compiler and the LAPACK libraries. It would be interesting to explore better implementations of the GP prior distribution, for example using a covariance operator instead of covariance function, or drawing samples using singular value decomposition instead of Cholesky decomposition.

Although we have tried to made GPs more accessible to practitioners in this thesis by investigating the length scale parameter and constructing methods by which we can compare functions, GPs and Bayesian Nonparametric methods as a whole are still inaccessible. One method to make these more accessible would be to develop an R package to implement nonparametric methods for spatial epidemics. This package would implement a GP based method given observed removal times and spatial coordinates. This would reduce the expertise needed to run these methods. Another way in which these methods are inaccessible is the large uncertainty compared to parametric methods. This can be improved in two ways. The first is to reassure practitioners that uncertainty is a natural consequence of these methods and that parametric methods require many unjustifiable assumptions. The second is to again work on methods of including mean functions in the prior distribution.

Throughout this thesis, we have assumed the outbreaks are complete. These

methods can be extended to analyse the spread of infectious diseases early on in an outbreak. We would however suffer from the same problem as mentioned above – a severe lack of data. One way to mitigate this is to include assumptions in the model, such a monotonicity, asymptotic behaviour or, in extreme cases, using covariance functions which only model certain types of functions, such as the linear covariance function. The credible intervals around our estimates would also be much larger than current parametric estimates, due to the use of nonparametric methods and the lack of data. Further work would look at how to make nonparametric methods competitive against parametric methods for real-time inference. In low data situations, such as the start of an outbreak, the GP prior distribution will regress to its mean, and so we can include an informative mean function in the prior distribution.

We can also extend this work to consider outbreaks in discrete time. This has already been done for time-dependent infection rate functions using GPs O'Neill and Kypraios (2018). As the Bayesian nonparametric method concerns the infection rate, modelling outbreaks in discrete time will only involve changing the likelihood function, in particular the infectious period distribution.

We can extend the structure of the model to allow for recent technological advances. When modelling the outbreak of a disease among livestock using a spatial covariate, we have only considered the Euclidean distance between farms. As discussed in Kypraios (2007, §3.3.2), the Euclidean distance may be an insufficient measure of distance as it does not take the physical geography of a location into account. Using Geographic Information Systems (GIS) or mapping Application Programming Interfaces (APIs), we can now easily compute the walking and driving distance between locations. This is particularly pertinent to outbreaks of disease among livestock where the disease may be spread by humans travelling between farms. Our method can be extended to include this information by using a mixture model. We would consider the infection rate function a weighted sum of several functions, each measuring the infection rate using different distances. Although we have only used a Euclidean distance metric, we can also consider other types of metrics. A further application is considering outbreaks on a network where the relationship between the individuals is described

by the edge between them. In a similar vein, we could consider pandemics, where we model the spread of a disease between cities. The weighting of each function could indicate the importance of each distance metric. As well as technological advances in mapping software, we could make use of better computer software and hardware for the MCMC algorithm. For example, we have used a GCC compiler with the GSL BLAS library, but improvements could be made by using an Intel compiler or the LAPACK library. It may be possible to shorten the time the GP element converges in the MCMC algorithm by using a more efficient algorithm. We implemented the underrelaxed proposal mechanism, as this updates the function as a block and requires reduces computational complexity. Other, more computationally complex methods, such a Hamiltonian Monte Carlo or Stochastic Gradient MCMC, may be more efficient at sampling from the target density and reduce convergence time.

There are many modelling choices and weak assumptions which can be made when implementing this method. Many of these are biological or epidemiological assumptions which are difficult to include in parametric models without being very precise. For example, in diseases which are airborne or spread by migratory birds, we may want to include the direction of prevailing wind or the general direction of the migratory path. Instead of computing the Euclidean distance between each pair of individuals, we can use a modified Euclidean metric, which is given by:

$$\rho(i,j) = (\mathbf{x}_i - \mathbf{x}_j)^T \Lambda (\mathbf{x}_i - \mathbf{x}_j).$$

Here, $\mathbf{x}_i$ is the coordinates of individual i and Λ is a matrix which defines the direction component. Setting $\Lambda = I$ gives the standard Euclidean distance metric, the diagonal terms give weighting to the x and y directions and non-zero off-diagonal terms allow for interactions between the directions. We can also change the underlying covariance structure by weighting individuals and treating some of them as 'super infectors', who are much more likely to pass on the disease or mix with a larger number of individuals. This can be done by shortening the distances between each individual and the 'super infector'. For example, in an outbreaks of disease among humans,

these individuals may be airline employees or frequent travellers.

6.3 Concluding Remarks

This thesis contributes new Bayesian nonparametric methods for stochastic epidemic models. These methods can be successfully implemented alongside existing Bayesian methods in an MCMC framework, such as data augmentation techniques for inferring when individuals were infected. These Bayesian nonparametric methods use Gaussian process prior distributions to provide a more flexible framework than current parametric methods. They allow the observed data to speak for itself and for fewer assumptions to be made when modelling infection rates. Our methods do away with the need for arbitrarily choosing parametric forms, which may lack any epidemiological basis. Nevertheless, we can still allow for more general assumptions to made, such as assuming *a priori* the function is decreasing. The methods also allow us to better quantify the uncertainty by fitting far less specific models. We have demonstrated the success of our methods using both simulated data sets and real-life data sets, where we were able to make contributions beyond the Bayesian nonparametric framework.

Bibliography

Abrahamsen, P. (1997). *A review of Gaussian random fields and correlation functions*. Norwegian Computing Center, Oslo, Norway.

Adams, R. P., Murray, I., and MacKay, D. J. C. (2009). The Gaussian process density sampler. *Advances in Neural Information Processing Systems*, 21.

Alexandersen, S., Zhang, Z., Donaldson, A., and Garland, A. (2003). The pathogenesis and diagnosis of foot-and-mouth disease. *Journal of Comparative Pathology*, 129(1):1–36.

Alvarez, M., Luengo, D., and Lawrence, N. D. (2009). Latent force models. In *Artificial Intelligence and Statistics*, pages 9–16.

Álvarez, M. A., Rosasco, L., and Lawrence, N. D. (2012). Kernels for vector-valued functions: A review. *Foundations and Trends® in Machine Learning*, 4(3):195–266.

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

Bacaer, N. (2011). *A Short History of Mathematical Population Dynamics*. Springer.

Backer, J., van Roermund, H., Fischer, E., van Asseldonk, M., and Bergevoet, R.

(2015). Controlling highly pathogenic avian influenza outbreaks: An epidemiological and economic model analysis. *Preventive Veterinary Medicine*, 121(1-2):142–150.

Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications*. Griffin, London, 2nd edition.

Ball, F., Mollison, D., and Scalia-Tomba, G. (1997). Epidemics with two levels of mixing. *The Annals of Applied Probability*, 7(1):46–89.

Bataille, A., van der Meer, F., Stegeman, A., and Koch, G. (2011). Evolutionary analysis of inter-farm transmission dynamics in a highly pathogenic avian influenza epidemic. *PLoS Pathogens*, 7(6):e1002094.

Becker, N. G. (1989). *Analysis of Infectious Disease Data*. Chapman and Hall, London.

Bernardo, J. M. and Smith, A. F. M., editors (1994). *Bayesian Theory*. John Wiley & Sons, Inc.

Bernoulli, D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite verole et des avantages de l'inoculation pour la prevenir. *Mémoires de mathématiques et de physique de l'Académie royale des sciences*, pages 1–45.

Boender, G. J., Hagenaars, T. J., Bouma, A., Nodelijk, G., Elbers, A. R. W., de Jong, M. C. M., and van Boven, M. (2007). Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLoS Computational Biology*, 3(4):e71.

Boyle, P. and Frean, M. (2005). Dependent gaussian processes. In *Advances in neural information processing systems*, pages 217–224.

Britton, T. (1998). Estimation in multitype epidemics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(4):663–679.

Bui, T. D., Yan, J., and Turner, R. E. (2017). A unifying framework for Gaussian process pseudo-point approximations using power expectation propagation. *Journal of Machine Learning Research*, 18(104):1–72.

Cressie, N. A. C. (1993). *Statistics for Spatial Data: Revised Edition*. Wiley, New York.

Csato, L. and Opper, M. (2002). Sparse online Gaussian processes. *Neural Computation*, 14(3):641–668.

Diggle, P. J. (2006). Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Stat. Methods Med. Res*, 15(4):325–336.

Director Generate Health and Consumers (2003). Avian influenza (AI) in the Netherlands, Belgium and Germany – chronology of main events and list of decisions adopted by the commission. Technical report, European Union.

Elbers, A. R. W., Fabri, T. H. F., de Vries, T. S., de Wit, J. J., Pijpers, A., and Koch, G. (2004). The highly pathogenic Avian Influenza A (H7N7) virus epidemic in the netherlands in 2003—lessons learned from the first five outbreaks. *Avian Diseases*, 48(3):691–705.

Ferguson, N. M. (2001). The foot-and-mouth epidemic in great britain: Pattern of spread and impact of interventions. *Science*, 292(5519):1155–1160.

Flury, B. N. (1984). Common principal components inkGroups. *Journal of the American Statistical Association*, 79(388):892–898.

Fouchier, R. A. M., Schneeberger, P. M., Rozendaal, F. W., Broekman, J. M., Kemink, S. A. G., Munster, V., Kuiken, T., Rimmelzwaan, G. F., Schutten, M., van Doornum, G. J. J., Koch, G., Bosman, A., Koopmans, M., and Osterhaus, A. D. M. E. (2004). Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proceedings of the National Academy of Sciences*, 101(5):1356–1361.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., and Vehtari, A. (2013). *Bayesian Data Analysis*. Taylor & Francis Ltd.

Gibson, G. J. and Renshaw, E. (1998). Estimating parameters in stochastic compartmental models using markov chain methods. *IMA Journal of Mathematics Applied In Medicine and Biology*, 15(1):19–40.

Heinonen, M., Mannerström, H., Rousu, J., Kaski, S., and Lähdesmäki, H. (2016). Non-stationary gaussian process regression with hamiltonian monte carlo. In Gretton, A. and Robert, C. C., editors, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 732–740, Cadiz, Spain.

Held, L., Hens, N., O'Neill, P. D., and Wallinga, J., editors (2019). *Handbook of Infectious Disease Data Analysis*. Taylor & Francis Ltd.

Hensman, J., Fusi, N., and Lawrence, N. D. (2013). Gaussian processes for big data. In *Conference on Uncertainty in Artificial Intelligence*, pages 282–290.

Jacquez, J. A., Simon, C. P., Koopman, J., Sattenspiel, L., and Perry, T. (1988). Modeling and analyzing HIV transmission: the effect of contact patterns. *Mathematical Biosciences*, 92(2):119–199.

Jewell, C. P., Kypraios, T., Neal, P., and Roberts, G. O. (2009). Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4(3):465–496.

Keeling, M. J. (2001). Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. *Science*, 294(5543):813–817.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721.

Knock, E. S. and Kypraios, T. (2014). Bayesian non-parametric inference for infectious disease data. `arXiv:1411.2624`.

Koopmans, M., Wilbrink, B., Conyn, M., Natrop, G., van der Nat, H., Vennema, H., Meijer, A., van Steenbergen, J., Fouchier, R., Osterhaus, A., and Bosman, A.

(2004). Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the netherlands. *The Lancet*, 363(9409):587–593.

Kypraios, T. (2007). *Efficient Bayesian Inference for Partially Observed Stochastic Epidemics and A New Class of Semi-Parametric Time Series Models*. PhD thesis, Lancaster University.

Lawson, A. and Zhou, H. (2005). Spatial statistical modeling of disease outbreaks with particular reference to the UK foot and mouth disease (FMD) epidemic of 2001. *Preventive Veterinary Medicine*, 71(3-4):141–156.

Lekone, P. E. and Finkenstädt, B. F. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 62(4):1170–1177.

Liu, H., Cai, J., and Ong, Y. S. (2018). Remarks on multi-output gaussian process regression. *Knowledge-Based Systems*, 144:102–121.

Murray, I. and Adams, R. P. (2010). Slice sampling covariance hyperparameters of latent Gaussian models. In Lafferty, J. D., Williams, C. K. I., Shawe-Taylor, J., Zemel, R. S., and Culotta, A., editors, *Advances in Neural Information Processing Systems 23*, pages 1732–1740. Curran Associates, Inc.

Neal, R. M. (1995). Suppressing random walks in Markov chain Monte Carlo using ordered overrelaxation. Technical report, Department of Statistics, University of Toronto.

Nguyen, T. V. and Bonilla, E. V. (2014). Collaborative multi-output gaussian processes. In Zhang, N. and Tian, J., editors, *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pages 643 – 652.

O'Neill, P. D. and Kypraios, T. (2018). Bayesian nonparametrics for stochastic epidemic models. *Statistical Science*, 33(1):44–56.

O'Neill, P. D. and Roberts, G. (1999). Bayesian inference for partially observed stochastic epidemics. *Statistics in Society A*, 162(1):121–129.

Orbanz, P. and Teh, Y. W. (2017). *Encyclopedia of Machine Learning and Data Mining*, chapter Bayesian Nonparametric Models, pages 107–116. Springer, 2nd edition.

Probert, W. J. M., Jewell, C. P., Werkman, M., Fonnesbeck, C. J., Goto, Y., Runge, M. C., Sekiguchi, S., Shea, K., Keeling, M. J., Ferrari, M. J., and Tildesley, M. J. (2018). Real-time decision-making during emergency disease outbreaks. *PLOS Computational Biology*, 14(7):e1006202.

Quinonero-Candela, J. and Rasmussen, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, 6.

Rasmussen, C. E. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Retkute, R., Jewell, C. P., Boeckel, T. P. V., Zhang, G., Xiao, X., Thanapongtharm, W., Keeling, M., Gilbert, M., and Tildesley, M. J. (2018). Dynamics of the 2004 avian influenza H5N1 outbreak in thailand: The role of duck farming, sequential model fitting and control. *Preventive Veterinary Medicine*, 159:171–181.

Riihimäki, J. and Vehtari, A. (2010). Gaussian processes with monotonicity information. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *PMLR*, pages 645–652.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications (Chapman & Hall/CRC Monographs on Statistics and Applied Probability)*. Chapman and Hall/CRC.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shi, J., Titsias, M. K., and Mnih, A. (2019). Sparse orthogonal variational inference for Gaussian Processes. `arXiv:1910.10596`.

Stegeman, A., Bouma, A., Elbers, A. R. W., de Jong, M. C. M., Nodelijk, G., de Klerk, F., Koch, G., and van Boven, M. (2004). Avian Influenza A Virus (H7N7) epidemic in the Netherlands in 2003: Course of the epidemic and effectiveness of control measures. *The Journal of Infectious Diseases*, 190(12):2088–2095.

Stegeman, A., Elbers, A. R., Smak, J., and de Jong, M. C. (1999). Quantification of the transmission of classical swine fever virus between herds during the 1997–1998 epidemic in the Netherlands. *Preventive Veterinary Medicine*, 42(3-4):219–234.

Stockdale, J. E. (2019). *Bayesian computational methods for stochastic epidemics*. PhD thesis, Univeristy of Nottingham.

Stockdale, J. E., Kypraios, T., and O'Neill, P. D. (2018). Modelling and bayesian analysis of the Abakaliki smallpox data. *Epidemics*, 19:13–23.

Stockdale, J. E., Kypraios, T., and O'Neill, P. D. (2019). Pair-based likelihood approximations for stochastic epidemic models. *Biostatistics*.

Titsias, M. K. (2009). Variational learning of inducing variables in sparse Gaussian processes. In van Dyk, D. and Welling, M., editors, *Proceedings of the Twelfth International Workshop on Artificial Intelligence and Statistics*, volume 5, pages 567–574, Clearwater Beach, USA. JMLR W&CP 5.

Titsias, M. K., Rattray, M., and Lawrence, N. D. (2011). *Markov chain Monte Carlo algorithms for Gaussian processes*, chapter Markov chain Monte Carlo algorithms for Gaussian processes, pages 295–316. Cambridge University Press.

Velkers, F. C., Bouma, A., Matthijs, M. G. R., Koch, G., Westendorp, S. T., and Stegeman, J. A. (2006). Outbreak of avian influenza H7N3 on a turkey farm in the Netherlands. *Veterinary Record*, 159(13):403–405.

Welding, J. and Neal, P. (2019). Real time analysis of epidemic data. `arXiv:1909.11560`.

Xu, X. (2015). *Bayesian Nonparametric Inference for Stochastic Epidemic Models*. PhD thesis, University of Nottingham.

Xu, X., Kypraios, T., and O'Neill, P. D. (2016). Bayesian non-parametric inference for stochastic epidemic models using Gaussian processes. *Biostatistics*, 17(4):619–633.