Imperial College of Science, Technology and Medicine
Department of Mathematics

# Bayesian Point Processes models with Applications in the COVID-19 Pandemic

Stamatina Lamprinakou

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy of
Imperial College London and the Diploma of Imperial College, November 2022

# Statement of Originality

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas, techniques, quotations, or any other material from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

# Copyright

# Abstract

A point process is a set of points randomly located in a space, such as time or abstract spaces. Point process models have found numerous applications in epidemiology, ecology, geophysics, social networks and many other areas.

The Poisson process is the most widely known point process. Poisson intensity estimation is a vital task in various applications including medical imaging, astrophysics and network traffic analysis. A Bayesian Additive Regression Trees (BART) scheme for estimating the intensity of inhomogeneous Poisson processes is introduced. The new approach enables full posterior inference of the intensity in a non-parametric regression setting. The performance of the novel scheme is demonstrated through simulation studies on synthetic and real datasets up to five dimensions, and the new scheme is compared with alternative approaches. A drawback of the proposed algorithm is its axis-alignment nature. We discuss this problem and suggest alternative approaches to remedy the drawback.

The novel coronavirus disease (COVID-19) has been declared a Global Health Emergency of International Concern with over 557 million cases and 6.36 million deaths as of 3 August 2022 according to the World Health Organization. Understanding the spread of COVID-19 has been the subject of numerous studies, highlighting the significance of reliable epidemic models. We introduce a novel epidemic model using a latent Hawkes process with temporal covariates for modelling the infections. Unlike other Hawkes models, we model the reported cases via a probability distribution driven by the underlying Hawkes process. Modelling the infections via a Hawkes process allows us to estimate by whom an infected individual was infected. We propose a Kernel Density Particle Filter (KDPF) for inference of both latent cases and reproduction number and for predicting new cases in the near future. The computational effort is proportional to the number of infections making it possible to use particle filter-type algorithms, such as the KDPF. We demonstrate the performance of the proposed algorithm on synthetic data sets and COVID-19 reported cases in various local authorities in the UK, and benchmark our model to alternative approaches.

We extend the unstructured homogeneously mixing epidemic model considering a finite population stratified by age bands. We model the actual unobserved infections using a latent marked Hawkes process and the reported aggregated infections as random quantities driven by the underlying Hawkes process. We apply a Kernel Density Particle Filter (KDPF) to infer the marked counting process, the instantaneous reproduction number for each age group and forecast the epidemic's future trajectory in the near future. We demonstrate the performance of the proposed inference algorithm on synthetic data sets and COVID-19 reported cases in various local authorities in the UK. Taking into account the individual heterogeneity in age provides a real-time measurement of interventions and behavioural changes.

*To my family and all people who have dreams, pursue them through ethical and legitimate means and overcome the obstacles in their path, making them stronger and more determined.*

# Acknowledgements

First and foremost, I would like to thank my supervisors, Prof McCoy and Prof Gandy, for giving me the opportunity to pursue this area of research and for their support. I would also like to thank Prof Young for his advice.

I would like to express gratitude to the mathematics department and the Data Science Institute for their financial support (Roth-Data Science Institute Scholarship).

I can't express how thankful I am to my family. My family's support, love, encouragement, and belief in me gave me the strength and courage to pursue this thesis. Without my family's help, I couldn't have completed my PhD.

# Contents

**1 Introduction** **15**
- 1.1 Motivation and contributions . . . . . 15
- 1.2 Publications and working projects . . . . . 17

**2 Background** **19**
- 2.1 Point process . . . . . 19
  - 2.1.1 Poisson process . . . . . 20
  - 2.1.2 Hawkes process . . . . . 20
  - 2.1.3 Log-Gaussian Cox process . . . . . 22
  - 2.1.4 p-thinning . . . . . 22
- 2.2 Branching process . . . . . 23
- 2.3 Frequentist methods for estimating the intensity of a Poisson process 24
- 2.4 Markov Chain Monte Carlo . . . . . 24
  - 2.4.1 Metropolis-Hastings algorithm . . . . . 25
  - 2.4.2 Gibbs sampling . . . . . 25
  - 2.4.3 Difficulties of inference from MCMC . . . . . 25
- 2.5 Particle algorithms for estimating parameters of state-space models . 26
- 2.6 Particle Marginal Metropolis-Hastings sampler . . . . . 28
- 2.7 Bayesian Additive Regression Trees . . . . . 30
  - 2.7.1 Log-linear BART for count data . . . . . 32

**3 BART-based inference for Poisson processes** **33**
- 3.1 The BART Model for Poisson processes . . . . . 34
- 3.2 The Inference Algorithm . . . . . 36
  - 3.2.1 Fixing the hyperparameters of the model . . . . . 40
  - 3.2.2 Metropolis Hastings Proposals . . . . . 41
    - 3.2.2.1 GROW Proposal . . . . . 41
    - 3.2.2.2 PRUNE Proposal . . . . . 43
    - 3.2.2.3 CHANGE Proposal . . . . . 44
- 3.3 Simulation Study on Synthetic Data . . . . . 45
  - 3.3.1 One dimensional Poisson process with stepwise intensity . . . 48
  - 3.3.2 Two-dimensional Poisson process with stepwise intensity function . . . . . 51
  - 3.3.3 Inhomogeneous three-dimensional Poisson process with Gaussian intensity . . . . . 53

3.3.4 Inhomogeneous five dimensional Poisson process with sparsity assumption . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
3.4 Intensity estimation for Real Data . . . . . . . . . . . . . . . . . . . . 59
3.4.1 Earthquakes Data . . . . . . . . . . . . . . . . . . . . . . . . 59
3.4.2 Lansing Data . . . . . . . . . . . . . . . . . . . . . . . . . . 62
3.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

**4 Axis-alignment effect 65**
4.1 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
4.2 Ad-hoc Inference Algorithm using helpful rotations . . . . . . . . . . 67
4.3 Simulation Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . 67
4.3.1 Inhomogeneous Poisson processes in a disk . . . . . . . . . . 68
4.3.2 Inhomogeneous Poisson process in a square . . . . . . . . . . . 72
4.4 Intensity estimation for Real Data . . . . . . . . . . . . . . . . . . . . 72
4.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 75

**5 Using a latent Hawkes process for epidemiological modelling 78**
5.1 Related work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
5.2 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
5.3 Inference algorithm . . . . . . . . . . . . . . . . . . . . . . . . . . 82
5.3.1 Kernel Density Particle Filter . . . . . . . . . . . . . . . . . . 84
5.4 Simulation Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . 87
5.5 Real Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 93
5.6 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 102

**6 Age-stratified epidemic model using a marked latent Hawkes process 108**
6.1 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 110
6.2 Inference algorithm . . . . . . . . . . . . . . . . . . . . . . . . . . 111
6.3 Simulation Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . 114
6.3.1 Two age groups . . . . . . . . . . . . . . . . . . . . . . . . . 116
6.4 Real Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117
6.5 Epidemic Dynamics . . . . . . . . . . . . . . . . . . . . . . . . . . 131
6.6 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 133

**7 Conclusions and Further Work 135**

**A Appendix of Chapter 3 137**
A.1 The model for the case of one tree . . . . . . . . . . . . . . . . . . . 137
A.1.1 Poisson process conditional likelihood . . . . . . . . . . . . . . 137
A.1.2 Inference Algorithm . . . . . . . . . . . . . . . . . . . . . . . 138
A.2 Simulation results on synthetic data with various number of sampling iterations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 139
A.2.1 One dimensional Poisson process with stepwise intensity . . . 140
A.2.2 One dimensional Poisson process with continuously varying intensity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 141

A.2.3 Two dimensional Poisson process with stepwise intensity function . . . 142
A.2.4 Inhomogeneous two dimensional Poisson process with Gaussian intensity . . . 142
A.3 Intensity estimation for real data . . . 146
A.3.1 Coal Data . . . 146
A.3.2 Redwoodfull Data . . . 146
A.4 Simulation study on synthetic data . . . 147
A.4.1 One dimensional Poisson process with continuously varying intensity . . . 147
A.4.2 Inhomogeneous two-dimensional Poisson process with Gaussian intensity . . . 148
A.4.3 Inhomogeneous five dimensional Poisson process with Gaussian intensity . . . 151
A.5 Additional simulation results on synthetic data . . . 151
A.5.1 One dimensional Poisson process with stepwise intensity . . . 152
A.5.1.1 5 Trees . . . 152
A.5.1.2 7 Trees . . . 154
A.5.2 One dimensional Poisson process with with continuously varying intensity . . . 156
A.5.2.1 5 Trees . . . 156
A.5.2.2 10 Trees . . . 158
A.5.3 Inhomogeneous two dimensional Poisson process with Gaussian intensity . . . 160
A.5.3.1 8 Trees . . . 160
A.5.3.2 10 Trees . . . 162
A.5.4 Two dimensional Poisson process with stepwise intensity function . . . 165
A.5.4.1 4 Trees . . . 165
A.6 Additional simulation results on real data . . . 167
A.6.1 Coal Data . . . 167
A.6.1.1 8 Trees . . . 167
A.6.1.2 10 Trees . . . 169
A.6.2 Earthquakes Data . . . 171
A.6.2.1 10 Trees . . . 171
A.6.3 Mapples . . . 173
A.6.3.1 5Trees . . . 174
A.6.3.2 10 Trees . . . 175
A.6.4 Redwood . . . 176
A.6.4.1 5 Trees . . . 176
A.6.4.2 10 Trees . . . 178

**B Appendix of Chapter 5 180**

## C Appendix of Chapter 6

**C Appendix of Chapter 6 . . . 182**
C.1 Simulation Analysis . . . 182
  C.1.1 Four age groups . . . 182
C.2 Independent processes $\{\gamma(t,a)\}_a$ on age groups . . . 187
  C.2.1 Six age groups . . . 187
  C.2.2 Nine age groups . . . 192
C.3 Real Data . . . 198
  C.3.1 Ashford . . . 198
  C.3.2 Kingston upon the Thames . . . 207
  C.3.3 Leicester . . . 207
C.4 Alternative method of estimating the susceptible population in Ashford and Kingston upon Thames . . . 212
Bibliography . . . 218

# Chapter 1

# Introduction

## 1.1 Motivation and contributions

The thesis is concerned with the analysis of point processes, with a particular emphasis on the data that has arisen from the COVID-19 pandemic. A point process is a set of points randomly located in a space such as time or abstract spaces. Point process models have found numerous applications in epidemiology, ecology, geophysics, social networks and many other areas. An event can be, for example, an infection time in epidemiology; an observed species location in ecology; a buy or sell transaction of a stock in finance. The statistical properties and inference for point processes have been widely investigated in the literature [Daley and Vere-Jones, 2003; Cox and Isham, 1980; Baddeley et al., 2015].

The Poisson process is the most widely known point process. The question of estimating the intensity of Poisson processes has a long history, including both frequentist and Bayesian methods. Frequentist methods include fixed-bandwidth and adaptive bandwidth kernel estimators with edge correction [Diggle et al., 2003], and wavelet-based methods (e.g. Fryzlewicz and Nason [2004], Patil et al. [2004]). Bayesian methods include using a sigmoidal Gaussian Cox process model for intensity inference [Adams et al., 2009], variational Bayesian intensity inference [Lloyd et al., 2015], and non-parametric Bayesian estimations of the intensity via piecewise functions with either random or fixed partitions of constant intensity [Arjas and Gasbarra, 1994; Heikkinen and Arjas, 1998; Gugushvili et al., 2018].

In Chapter 2, we introduce a novel Bayesian Additive Regression Tree (BART) scheme for estimating the intensity of inhomogeneous Poisson processes that is published in the Journal Computational Statistics and Data Analysis [Lamprinakou et al., 2023a]. The effectiveness of BART has been demonstrated in a variety of contexts including non parametric regression and classification. We demonstrate the performance of the scheme through simulation studies on synthetic and real datasets up to five dimensions, and compare the approach to alternative approaches. The simulation studies demonstrate that our algorithm is competitive with the Haar-Fisz algorithm in one dimension, kernel smoothing in two dimensions, and outperforms the kernel approach for multidimensional intensities. The simulation analysis also demonstrates that our proposed algorithm is competitive with the inference via spatial log-Gaussian Cox processes

A drawback of the proposed algorithm is its axis-alignment nature. A simulation study shows that points close to jumps are estimated with less reliability, which is expected due to that drawback. Recently, many Bayesian and non-Bayesian studies including [Ge et al., 2019; Fan et al., 2016; Tomita et al., 2020; Rainforth and Wood, 2015; Rodriguez et al., 2006; Blaser and Fryzlewicz, 2021, 2016] have proposed methods for dealing with the axis-alignment effects. Chapter 4 discusses this problem and proposes alternative approaches to remedy the drawback.

In Chapter 5, we move to developing epidemic models with applications using data from COVID-19, an infectious disease which has declared as a Global Health Emergency of International Concern with over 557 million cases and 6.36 million deaths as of 3 August 2022 according to the World Health Organization. In the absence of vaccines, countries initially followed mitigation strategies to prevent the rapid spread of COVID-19, such as social distancing, quarantine, mask wearing, and city lock-downs.

Understanding the spread of COVID-19 has been the subject of numerous studies, highlighting the significance of reliable epidemic models. A large number of studies have been carried out to understand the spread of COVID-19, forecast new cases and when the peak of the pandemic will occur, and investigate "what-if-scenarios". A report of Imperial [Ferguson et al., 2020] presented the results of epidemiological modelling looking at a variety of nonpharmaceutical interventions. Several compartmental models [Zou et al., 2020; Chen et al., 2020; Wangping et al., 2020; Roques et al., 2020] using ordinary differential questions (ODE) have been proposed for modelling the spread of COVID-19. Various models using Hawkes processes [Garetto et al., 2021; Kresin et al.; Escobar, 2020; Chiang et al., 2021; Browning et al., 2021; Koyama et al., 2021; Bertozzi et al., 2020], widely used to model contagion patterns, have been introduced as an alternative to ODE models. Others have used Poisson autoregression model for the daily new observed cases [Agosto and Giudici, 2020] and a Bayesian model linking the infection cycle to observed deaths [Flaxman et al., 2020].

In Chapter 5, we introduce a novel epidemic model using a latent Hawkes process with temporal covariates for modelling the infections. Unlike other Hawkes models, we model the reported cases via a probability distribution driven by the underlying Hawkes process. Modelling the infections via a Hawkes process allows us to estimate by whom an infected individual was infected. We propose a Kernel Density Particle Filter (KDPF) [Sheinson et al., 2014; Liu and West, 2001] for inference of both latent cases and reproduction number and for predicting the new cases in the near future. The computational effort is proportional to the number of infections making it possible to use particle filter type algorithms, such as the KDPF. We demonstrate the performance of the proposed algorithm on synthetic data sets and COVID-19 reported cases in various local authorities in the UK, and benchmark the model to alternative approaches. The simulation analysis shows that the proposed algorithm provides comparable estimates of observed case fluctuations compared with those of Koyama et al. [2021]. The method of Koyama et al. [2021] and EpiEstim provide similar estimates of the reproduction number as the proposed algorithm. The model is published in the journal PLOS ONE [Lamprinakou et al., 2023b].

In Chapter 6, we extend the unstructured homogeneously mixing epidemic model introduced in Chapter 5 considering a finite population stratified by age bands. We model the actual unobserved infections using a latent marked Hawkes process and the reported aggregated infections as random quantities driven by the underlying Hawkes process. Following the inference approach of Chapter 5, we apply a KDPF to infer the marked counting process, the instantaneous reproduction number for each age group and forecast the epidemic's future trajectory over short time horizons. We demonstrate the performance of the proposed inference algorithm on synthetic data sets and COVID-19 reported cases in various local authorities in the UK. We illustrate that taking into account the individual heterogeneity in age decreases the uncertainty of estimates and provides a real-time measurement of interventions and behavioural changes; that conclusion is under investigation by increasing the number of particles, proposing alternative inference methods and priors. The model is available on arXiv [Lamprinakou and Gandy, 2022].

The above research was conducted during the first, second and fourth years of this PhD. In the second term of the 2nd year, an extensive literature review of the community detection research area was conducted and possible applications of the BART scheme in that area were investigated. A novel application of BART could be detection communities and inference edges on directed or undirected graphs. According to Handcock et al. [2007], many social networks exhibit clustering beyond what can be explained by transitivity (the property of being connected with the friend of my friend is far more likely than with some randomly chosen member of the population) and homophily by metadata (the tendency to associate with others whom they share common metadata). To tackle this difficulty, they assume that each node has an unobserved position in a d-dimensional Euclidean latent social space, and nodes belonging to the same community are not too far apart in the unobserved latent space of characteristics. Their model considers transitivity via latent space, homophily on metadata and clustering. They employ a logistic regression model with covariates the metadata and the distance of latent positions for inferring an edge between two vertices. Alternatively, the BART model can be applied for inferring the existence of an edge by modelling it as a Poisson random variable in line with Newman and Reinert [2016]; Riolo et al. [2017] with mean a regression function of metadata, group memberships and locations.

## 1.2 Publications and working projects

This thesis includes research that either is accepted for publication or has been in review for publication at the time of writing. Specifically, the work in Chapter 3 is published in the Journal of Computational Statistics and Data Analysis [Lamprinakou et al., 2023a]. The model presented in Chapter 5 is published in the Journal PLOS ONE [Lamprinakou et al., 2023b]. The model presented in Chapter 6 is under revision, and available as preprint under the title "Age-stratified epidemic model using a marked latent Hawkes process" [Lamprinakou and Gandy, 2022].

The thesis also introduces research that is in progress. This includes the application of the BART model to community detection problems introduced earlier in this Chapter, and the methods to address the axis-alignment effect of the BART model discussed in Chapter 4.

# Chapter 2

# Background

This chapter introduces the necessary theory on the fundamental aspects of the thesis, including point processes (Section 2.1), branching processes (Section 2.2), Markov Chain Monte Carlo Methods (Section 2.4), Particle Filters (Section 2.5) and Bayesian Additive Regression Trees (Section 2.7).

## 2.1 Point process

A point process (temporal or spatial) is a stochastic model of a collection of points that is typically interpreted as a sample from (or a realization of) the point process. Fundamental introductions to the theory of point processes include Cox and Isham [1980], Daley and Vere-Jones [2003, 2007], Van Lieshout [2000], Møller and Waagepetersen [2003], Baddeley et al. [2007] and Illian et al. [2008]. The theoretical framework presented here is mainly based on Illian et al. [2008] and Baddeley et al. [2007].

A point process in one dimension ("time"), usually called a temporal process, is a stochastic model of random event times. For example, a 1-dimensional point process can be used for modelling the emergency calls received by a hospital. One way to formulate a temporal point process is in terms of a counting process. The following definition is from Ross [1995].

**Definition 1.** *A stochastic process* $\{N(t), t \geq 0\}$ *is said to be a counting process if* $N(t)$ *represents the total number of events that have occurred up to time* $t$*. Hence, a counting process* $N(t)$ *must satisfy*

1. $N(t) \geq 0$.
2. $N(t)$ *is integer valued.*
3. *If* $s < t$*, then* $N(s) \leq N(t)$.
4. *For* $s < t$*,* $N(t) - N(s)$ *equals the number of events that have occurred in the interval* $(s, t]$.

A spatial point process is a stochastic model of a random pattern of points in $d$-dimensional space, where $d \geq 2$. For example, a 3-dimensional point process can be used for modelling the locations and the times of emergency calls received by the hospital.

The points of a point process might be labelled with extra information called marks. For example, emergency calls might carry their duration.

**Definition 2.** *A marked point process on a space $S$ with marks in a space $M$ composed of a point process $X$ with associate marks in $M$ is a locally finite counting process $Y$ on the Cartesian product $S \times M$, that is, the number of points of $Y$ falling into $K \times M$ is finite, for all $K \subset S$.*

A common approach to building the joint model for the marks and point process intensity is

$$\phi(x, m) = \lambda(x)h(m|x)$$

where $\lambda(\cdot)$ is the intensity of the point process $X$ and $h(m|x)$ denotes the conditional mark density at point $x \in S$ with $m \in M$.

### 2.1.1 Poisson process

The archetypal point processes are Poisson processes.

**Definition 3.** *A non-homogeneous Poisson process on a d-dimensional domain $S \subset R^d$, $d \geq 1$, with intensity $\lambda : S \to R^+$ and $\int_S \lambda(x)dx < \infty$ is a point process on $S$ with the following properties.*

1. *For every compact set $B \subset S$, the number of points falling into $B$, denoted by $N(B)$, has a Poisson distribution with mean $\int_B \lambda(x)dx$.*
2. *If $B_1, \ldots, B_m$ are disjoints sets, then $N(B_1), \ldots, N(B_m)$ are independent.*

Hence, the number of points in a region of $S$ is Poisson distributed, and the numbers of points are independent of each other. A homogeneous Poisson process is a special case with intensity $\lambda(x) = \lambda_0, \forall x \in S$.

In the next chapters, we apply an extension of the widely-used thinning technique to generate realizations of Poisson processes. Thinning [Lewis and Shedler, 1979] requires determining a constant $\lambda^*$ such that $\lambda^* \geq \lambda(x)$ for all $x \in S$. Each point $x \in S$ of a homogeneous Poisson process with intensity $\lambda^*$ is retained, independent of all other points, with probability $\lambda(x)/\lambda^*$. To generate a realization of a Poisson process with intensity $\lambda^*$ in $S$, we first generate a random variable $M$ with a Poisson distribution with mean $\lambda^*|S|$, where $|S|$ denotes the volume of the region $S$. Given $M = m$, we then generate $m$ uniform points in $S$.

### 2.1.2 Hawkes process

The Hawkes process [Hawkes, 1971a,b, 2018; Laub et al., 2021] is a well-known self-exciting process based on a counting process in which the intensity depends on the history of the process. The Hawkes process has been applied for modelling a wide range of fields: earthquakes and aftershocks, dynamics of crime, epidemic dynamics, finance and many others.

**Definition 4.** *Consider* $(N(t) : t \geq 0)$ *a counting process with associate history* $\mathcal{H}_t = \{t_i : t_i < t\}$ *being the times of events prior to time* $t$*, that satisfies*

$$P\left(N(t+h) - N(t) = m|\mathcal{H}_t\right) = \begin{cases} 1 - \lambda(t)h + o(h) & \text{, m=0} \\ \lambda(t)h + o(h) & \text{, m=1} \\ o(h) & \text{, m>1} \end{cases}$$

*Suppose the process' conditional intensity function of the form*

$$\lambda(t) = v + \sum_{t_i \in \mathcal{H}_t} \mu(t - t_i)$$

*for some* $v > 0$ *and* $\mu : R^+ \to R^+$ *which are called the background intensity and excitation function, respectively. Such a process* $N$ *is a (linear) Hawkes process.*

Common choices for the excitation function are the exponential function and the power-law function. Hawkes [1971a] used the exponential function for the theoretical derivations. In this case, the Hawkes conditional intensity takes the form

$$\lambda(t) = v + \sum_{t_i \in \mathcal{H}_t} a e^{-\beta(t-t_i)}$$

where $a, \beta > 0$. Each new event increases the intensity by $a$, and over time this event's influence decays exponentially at rate $\beta$.

The power-law function was popularized by the geological model, called Omori's law [Ogata, 1999], and was used to predict the rate of aftershocks caused by an earthquake. In this case, the Hawkes conditional intensity takes the form

$$\lambda(t) = v + \sum_{t_i < t} \frac{k}{(c + (t - t_i))^p}$$

where $c, k, p > 0$.

Various extensions of the (linear) Hawkes process have been proposed, such as the non-linear Hawkes process [Brémaud and Massoulié, 1996] and the neural Hawkes process [Mei and Eisner, 2017]. A simple extension of the Hawkes process is the mutually exciting Hawkes process, which is a collection of one-dimensional Hawkes processes which excite themselves and each other.

**Definition 5.** *Consider a collection of* $m$ *counting processes* $\{N_1(), \cdots, N_m()\}$ *denoted by* $\mathbf{N}$*. Say* $\{T_{i,j} : i \in \{1, \ldots, m\}, j \in \mathbb{N}\}$ *are the random arrival times for each counting process (and* $t_{i,j}$ *for observed arrivals). If for each* $i = 1, \ldots, m$*,* $N_i()$ *has a conditional intensity of the form*

$$\lambda_i(t) = v_i + \sum_{j=1}^{m} \sum_{t_{j,k} < t} \mu_{i,j}(t - t_{j,k})$$

*for some* $v_i > 0$ *and* $\mu_{i,j} : R^+ \to R^+$ *then* $\mathbf{N}$ *is called a mutually exciting Hawkes process.*

Various methods, including thinning and superposition of Poisson processes, have been proposed to generate Hawkes process realizations [Laub et al., 2021]. In Chapters 5 and 6, we have applied the Hawkes process and the mutually exciting Hawkes process for modelling the spread of an epidemic. We have generated realizations of Hawkes processes using the method of superpositions of Poisson processes on multiple subintervals; the union of them give the finite horizon under consideration.

**Marked Hawkes process** A common approach to model the intensity of a marked Hawkes process is

$$\lambda(t) = v + \sum_{t_i<t} \phi_{m_i}(t - t_i)$$

where $m_i$ is the mark of event at time $t_i$ and $\phi_{m_i}(\cdot)$ an excitation kernel dependent on the mark of event at $t_i$.

### 2.1.3 Log-Gaussian Cox process

Log-Gaussian Cox process (LGCP) is an inhomogeneous Poisson process with stochastic intensity, $\Lambda$ where $\log \Lambda$ is a Gaussian process. LGCP is commonly used to model spatial point pattern data like tree locations and lightning data. Consider a point pattern with locations $(s_1, \ldots, s_l)$ within the domain $S \subset R^2$. A common approach for estimating the intensity is to partition the domain into $n$ disjoint subregions $\{B_1, \ldots, B_n\}$ and their associated centroids $\{c_1, \ldots, c_n\}$. We count the observed points per subregion via

$$Y_i = \sum_{j=1}^{l} \mathbb{1}\{s_i \in B_i\}$$

for $i = 1, .., n$, where $\mathbb{1}\{\cdot\}$ is the indicator function. Let $X(c_i)$ be a $(p+1)$-dimensional vector of an intercept and covariates constant within the $i_{th}$ subregion. The LGCP model for the point pattern $\{s_1, \ldots, s_n\}$ is

$$\begin{aligned} Y_i|\lambda &\sim \text{Poisson}\left(|B_i|\lambda(c_i)\right) \\ \lambda(c_i) &= \exp\left(X(c_i)\beta + Z(c_i)\right) \\ (Z(c_1), \ldots, Z(c_n)) &\sim \text{GP}(0, \Sigma) \end{aligned}$$

where $\lambda(\cdot)$ is an intensity function that is constant within subregion, $\beta$ is a $(p+1)$-dimensional vector of coefficients and $\Sigma$ a covariance matrix. Various methods have been proposed for fitting LGCP (e.g.,Brix and Diggle [2001], Diggle et al. [2005], Illian et al. [2012], Gelsinger et al. [2022]).

### 2.1.4 p-thinning

p-thinning [Illian et al., 2008] is the simplest form of thinning a counting process $N_b$; each point of $N_b$ is deleted with probability $1 - p$ independent of the other points of $N_b$. If $\lambda_b(\cdot)$ is the intensity function of the process $N_b$, the intensity function of the thinned process is $\lambda(x) = p\lambda_b(x)$.

## 2.2 Branching process

The study of branching processes grew out of Watson and Galton's collaboration to give a solution to the problem of the survival of family names. Branching processes [Harris, 1963; Dobrow, 2016] are a class of Markov chains that model the growth of populations and are widely used in biology and epidemiology to study the spread of infectious diseases and epidemics [Bartoszynski, 1967; Jacob, 2010]. The basic branching process is named after Galton and Watson, and described here.

We consider a population of individuals, each of which independently produces a random number of children according to the offspring distribution. Let $Z_n$ be the number of individuals (size) of the $n_{th}$ generation. Assume $Z_0 = 1$, as the population starts with one individual. The sequence $Z_0, Z_1, \ldots$ is a branching process.

**Mean Generation Size** The size of the $n_{th}$ generation is given by

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_i,$$

where $X_i$ denotes the number of offspring of the $i_{th}$ individual belonging to the $(n-1)_{th}$ generation. To find the mean size of the $n_{th}$ generation $E(Z_n)$, we apply the law of total expectation.

$$\begin{aligned} E(Z_n) &= \sum_{k=0}^{\infty} E\left(Z_n | Z_{n-1} = k\right) P(Z_{n-1} = k) = \sum_{k=0}^{\infty} E\left(\sum_{i=1}^{Z_{n-1}} X_i | Z_{n-1} = k\right) P(Z_{n-1} = k) \\ &= \sum_{k=0}^{\infty} E\left(\sum_{i=1}^{k} X_i\right) P(Z_{n-1} = k) = \mu E(Z_{n-1}), \end{aligned}$$

where $\mu = E(X_i)$. Iterating the recurrence relation gives

$$E(Z_n) = \mu E(Z_{n-1}) = \mu^2 E(Z_{n-2}) = \ldots = \mu^n E(Z_0) = \mu^n, \ n \geq 0.$$

The long-term expected generation size is

$$\lim_{n\to\infty} E(Z_n) = \lim \mu^n = \begin{cases} 0 & , \mu < 1 \\ 1 & , \mu = 1 \\ \infty & , \mu > 1. \end{cases}$$

A branching process is said to be subcritical if $\mu < 1$, critical if $\mu = 1$ and supercritical if $\mu > 1$.

### Extinction Probability

**Theorem 2.2.1.** *Given a branching process, let $G$ be the probability generating function of the offspring distribution. Then, the probability of eventual extinction is the smallest positive root of the equation $s = G(s)$. If $\mu \leq 1$, that is, in the subcritical and critical cases, the extinction probability is equal to 1.*

The theorem gives that the population becomes extinct with probability 1 in the subcritical and critical cases ($\mu \leq 1$), while there is a positive probability of the population's extinction in the supercritical case.

## 2.3 Frequentist methods for estimating the intensity of a Poisson process

Popular frequentist methods for estimating the intensity of an inhomogeneous Poisson process are fixed-bandwidth and adaptive-bandwidth kernel estimators with edge corrections [Diggle et al., 2003], and wavelet-based methods [e.g. Fryzlewicz and Nason, 2004; Patil et al., 2004] given a realization of the process. In Chapters 3 and 4, we apply the Haar-Fisz algorithm and kernel methods for estimating the intensity of one-dimensional and multidimensional Poisson processes, respectively.

**Haar-Fisz algorithm** The Haar-Fisz algorithm [Fryzlewicz and Nason, 2004] is a wavelet-based method applied to counts of points falling into consecutive subintervals and derives a deterministic discretized version of the intensity for one-dimensional Poisson processes. The **R** package **haarfisz** [Fryzlewicz, 2010] is a software to perform the Haar-Fisz algorithm.

**Kernel estimators** The kernel-based intensity estimators are non-parametric methods that estimate the intensity at each point $x$ of the domain via a sum of kernels centred at $x$. The kernels are spatial probability density functions weighing the influence of the process' realization on $x$. The bandwidth of the kernel function determines the collection of realization' points considered in estimating the intensity at $x$. There is a trade-off between sensitivity to noise for small bandwidth and over-smoothing for large bandwidth. A problem with kernel estimators arises when points of interest are close to the boundary of the domain. Edge-corrected kernel estimators (e.g. [Zheng et al., 2004] ) have been proposed. A non-edge corrected kernel estimator takes the form

$$\hat{\lambda}(x) = \sum_{s_i} K\left(\frac{x - s_i}{h}\right),$$

where $s = (s_1, .., s_n)$ is a realization of the process, and $K(\cdot)$ the kernel function with bandwidth $h$.

## 2.4 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a powerful framework that allows sampling from a large class of distributions. The key to Markov Chain simulation is to create a Markov process whose stationary distribution is the target distribution $p(\theta|y)$ and to run the simulations long enough that the drawn values of the parameter $\theta$ are samples from $p(\theta|y)$, where $y$ is the data. Two widely used MCMC algorithms are the Metropolis-Hastings algorithm and Gibbs sampling [Gelman et al., 2013; Robert and Casella, 1999].

### 2.4.1 Metropolis-Hastings algorithm

The algorithm proceeds as follows

1. Draw a starting value $\theta^{(0)}$ from a starting distribution $p_0(\theta)$.
2. For t=1,2,...
    - Sample a proposal $\theta^*$ from the distribution $q(\theta|\theta^{(t-1)})$ where $\theta^{(t-1)}$ is the value of the parameter at iteration $t-1$.
    - Calculate the Hastings-ratio, $r = \frac{p(\theta^*|y)q(\theta^{(t-1)}|\theta^*)}{p(\theta^{(t-1)}|y)q(\theta^*|\theta^{(t-1)})}$.
    - Set $\theta^{(t)} = \begin{cases} \theta^* & \text{, with probability } \min(r,1) \\ \theta^{(t-1)} & \text{, otherwise.} \end{cases}$

The sequence of iterations $\theta^{(1)}, \theta^{(2)}, \ldots$ converges to the target distribution. The convergence to the target requires conditions, and we recommend the books of Gelman et al. [2013] and Robert and Casella [1999] for more details.

### 2.4.2 Gibbs sampling

Suppose the parameter vector $\theta$ has been divided into $d$ components or subvectors, $\theta = (\theta_1, \ldots, \theta_d)$. Gibbs sampler is a special case of the Metropolis-Hastings algorithm. At each iteration $t$, each $\theta_j^{(t)}$ is sampled from the conditional distribution given all other components of $\theta$, $p\left(\theta_j|\theta_{-j}^{(t-1)}, y\right)$, where $\theta_{-j}^{(t-1)} = \left(\theta_1^{(t)}, \ldots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)}, \ldots, \theta_d^{(t-1)}\right)$. Each step of the Gibbs sampler updates one of the subvectors by drawing its conditional distribution conditioned to the latest values of the other components of $\theta$.

### 2.4.3 Difficulties of inference from MCMC

Iterative simulation adds three challenges to inference. First, the within-sequence correlation of draws can cause less precise inference than the same number of independent draws. The second difficulty is that early iterations will reflect the influence of the starting values, even when the chain has converged to the target distribution. To diminish the influence of starting approximations, we discard the early iterations. A conservative choice is to discard the first half of iterations and treat the second half as a sample from the target distribution. The third difficulty is assessing convergence, as it is impractical to run MCMC for an infinite number of iterations. Two widely used MCMC convergence diagnostic tools are the Gelman-Rubin diagnostic and the trace plot [Gelman et al., 2013], which we apply in Chapters 3 and 4 to monitor the convergence of chains.

**Trace plot** A trace plot illustrates the sampled values per chain throughout iterations. The trace plot is a useful tool to assess the convergence and mixing of the chains.

**Gelman-Rubin diagnostic** Let $m$ be the number of chains with length $n$ and $\psi$ be the quantity of interest. We label its simulations as $\psi_{ij}$, $i = 1, \ldots, n$ and $j = 1, \ldots, m$, after having discarded the first halves of chains. The between- and within- sequence variances are

$$B = \frac{n}{m-1} \sum_{j=1}^{m} \left(\bar{\psi}_{.j} - \bar{\psi}_{..}\right)^2, \text{ where } \bar{\psi}_{.j} = \frac{1}{n} \sum_{i=1}^{n} \psi_{ij}, \; \bar{\psi}_{..} = \frac{1}{m} \sum_{j=1}^{m} \bar{\psi}_{.j}$$

$$W = \frac{1}{m} \sum_{j=1}^{m} s_j^2, \; s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\psi_{ij} - \bar{\psi}_{.j}\right)^2.$$

The marginal posterior variance of $\psi$ is estimated by a weighted average of $W$ and $B$,

$$\hat{\text{var}}(\psi|y) = \frac{n-1}{n} W + \frac{1}{n} B.$$

The Gelman-Rubin diagnostic is defined by

$$\hat{R} = \sqrt{\frac{\hat{\text{var}}(\psi|y)}{W}}.$$

Once all chains converge, the between-chain variance should be close to 0. If $n \to \infty$ and $B \to 0$, $\hat{R}$ approaches to 1, indicating that we should run the chains until the diagnostic is close to 1 (usually less than or equal to 1.1) [Gelman et al., 2013; Robert and Casella, 2010]. We should diagnose convergence for each scalar quantity of interest.

## 2.5 Particle algorithms for estimating parameters of state-space models

We assume a discrete time Markov process $\{X_n\}_{n \geq 1}$ such that

$$X_1 \sim \mu(x_1) \text{ and } X_n | X_{n-1} \sim f(x_n | x_{n-1}),$$

where $\mu(x)$ is a probability density function and $f(x_n|x_{n-1})$ the transition density of moving from $x_n$ to $x_{n-1}$. We also assume that the observations $\{Y_n\}$ are independent given $\{X_n\}$.

A state-space model consists of a latent process $\{X_n\}$ and an observation process $\{Y_n\}$. A particle algorithm [Kantas et al., 2015; Doucet et al., 2009] approximates sequentially a sequence of posteriors $p_\theta(x_{1:n}|Y_{1:n})$ employing Importance Sampling (IS), as follows

$$\hat{p}_\theta(x_{1:n}|Y_{1:n}) = \sum_{k=1}^{N} W_n^k \delta_{X_{1:n}^k}(x_{1:n}),$$

where $z_{i:j}$ denotes the components $(z_i, z_{i+1}, \ldots, z_j)$ of a sequence $\{z_n\}$. The distribution $p_\theta(x_{1:n}|Y_{1:n})$ is approximated by using a set of $N$ weighted particles drawn from the chosen importance distribution. Each particle $X_{1:n}^k$ has an associated normalized weight $W_n^k$ that takes into account the discrepancy between the target and importance distribution.

To have $N$ equally weighted particles from $p_\theta(x_{1:n}|Y_{1:n})$, we sample $N$ times from its approximation $\hat{p}_\theta(x_{1:n}|Y_{1:n})$ ; that is, we select $X_{1:n}^k$ with probability the weight $W_n^k$. This operation is called multinomial resampling. The resampling step approximates $\hat{p}_\theta(x_{1:n}|Y_{1:n})$ by the resampled empirical measure

$$\bar{p}_\theta(x_{1:n}|Y_{1:n}) = \sum_{k=1}^{N} \frac{N_n^k}{N} \delta_{X_{1:n}^k}(x_{1:n}),$$

where $N_n^k$ is the number of times we see the particle $X_{1:n}^k$ in the set of equally weighted particles and $E\left(N_n^k|\{W_n^k\}_{k=1}^N\right) = NW_n^k$. We note that $\bar{p}_\theta(x_{1:n}|Y_{1:n})$ is an unbiased approximation of $\hat{p}_\theta(x_{1:n}|Y_{1:n})$. Improved unbiased resampling steps, including systematic resampling, with var $\left(N_n^k|\{W_n^k\}_{k=1}^N\right)$ being smaller than that obtained via the multinomial resampling have been proposed in the literature [Doucet et al., 2009; Douc and Cappe, 2005].

Particle algorithms (e.g., Auxiliary Particle Filter (APF) [Pitt and Shephard, 1999; Kantas et al., 2015; Doucet et al., 2009], Bootstrap Filter (BF) [Doucet et al., 2009; Kantas et al., 2015], Kernel Density Particle Filter (KDPF) [Sheinson et al., 2014; Liu and West, 2001]) are a combination of sequential IS and resampling. At step 1, we sample $N$ times from the importance distribution $q(x_1|Y_1)$, i.e., $X_1^i \sim q(x_1|Y_1)$, $i = 1, \ldots, N$, and form a weighted collection of particles $\{W_1^i, X_1^i\}$. Then, we apply the resampling step to form $N$ equally weighted particles, denoted by $\{\bar{X}_1^i\}_{i=1}^N$, as a set of samples approximately distributed according to $p_\theta(x_1|Y_1)$. At step 2, we approximate the posterior $p_\theta\left(x_{1:2}|Y_{1:2}\right)$. It holds that $p_\theta\left(x_{1:2}|Y_{1:2}\right) \propto p_\theta\left(x_2|x_1, Y_2\right) p_\theta\left(x_1, Y_1\right)$. Thus, a sensible approach is to extend each of the particles $\{\bar{X}_1^i\}_{i=1}^N$ by sampling $X_2^i \sim q(x_2|x_1, Y_2)$, giving a collection $\{W_2^i, \left(\bar{X}_1^i, X_2^i\right)\}$, approximately distributed according to $q_\theta\left(x_2|x_1, Y_2\right) p_\theta\left(x_1|Y_1\right)$. We apply the resampling step to form $N$ equally weighted particles, denoted by $\{\bar{X}_{1:2}^i\}$, approximately distributed according to $p_\theta\left(x_{1:2}|Y_{1:2}\right)$. We proceed until the final step.

We focus here on BF and APF and introduce KDPF in Chapter 5.

**Bootstrap Filter** BF (Algorithm 1) uses the transition density as importance distribution,
$q\left(x_n|x_{n-1}, Y_n\right) = f\left(x_n|x_{n-1}\right)$. The unnormalized importance weights are given by

$$\tilde{w}_n = \frac{p_\theta\left(x_{1:n}, Y_{1:n}\right)}{f(x_n|x_{n-1}) p_\theta(x_{1:n-1}, Y_{1:n-1})} = p_\theta\left(Y_n|x_n\right).$$

**Auxiliary Particle Filter** BF employs an importance distribution at each step $n$ which does not consider the associated observation $Y_n$. APF (Algorithm 2) gives a solution to this problem by considering the observation $Y_{n+1}$ in the resampling step $n$ via the predictive likelihood. The target distribution for APF is $p_\theta\left(x_{1:n}|Y_{1:n+1}\right)$.

**Algorithm 1** Bootstrap filter

```
Sample $N$ particles $\{X_1^j\}_{j=1}^N$ :
    for $j$ in $1:N$ do
        $X_1^j \sim \mu(x_1)$
    end for
Find the weights, $\tilde{w}_1 = \{\tilde{w}_1^j\}_{j=1}^N$:
    for $j$ in $1:N$ do
        $\tilde{w}_1^j = p_\theta(Y_1|x_1^j)$
    end for
Normalize the weights, $W_1 = \{W_1^j\}_{j=1}^N$:
    for $j$ in $1:N$ do
        $W_1^j = \frac{\tilde{w}_1^j}{\sum_{j=1}^N \tilde{w}_1^j}$
    end for
Estimate $P(Y_1) : \hat{P}(Y_1) = \frac{1}{N} \sum_{j=1}^N \tilde{w}_1^j$
for $n = 2, .., k$ do
    Resample and form $N$ equally weighted particles, $\bar{X}_{1:n-1} = \{\bar{X}_{1:n-1}^i\}_{i=1}^N$:
        for $j$ in $1:N$ do
            (i) sample index $i_j$ from a multinomial distribution with probabilities $W_{n-1}$
            (ii) $\bar{X}_{1:n-1}^j = X_{1:n-1}^{i_j}$
            (iii) $W_{n-1}^j = 1$
        end for
    Using $\bar{X}_{1:n-1}$ propagate:
        for $j$ in $1:N$ do
            $X_n^j \sim f(x_n|x_{n-1}^j)$
            Set $X_{1:n}^j = \left(\bar{X}_{1:n-1}^j, X_n^j\right)$
        end for
    Find the weights, $\tilde{w}_n = \{\tilde{w}_n^j\}_{j=1}^N$:
        for $j$ in $1:N$ do
            $\tilde{w}_n^j = W_{n-1}^j p_\theta(Y_n|x_n^j)$
        end for
    Normalize the weights, $W_n = \{W_n^j\}_{j=1}^N$:
        for $j$ in $1:N$ do
            $W_{j,n} = \frac{\tilde{w}_n^j}{\sum_{j=1}^N \tilde{w}_n^j}$
        end for
    Estimate $P(Y_n|Y_{1:n-1}) : \hat{P}(Y_n|Y_{1:n-1}) = \frac{1}{N} \sum_{j=1}^N \tilde{w}_n^j$.
end for
Resample and form $N$ equally weighted particles, $\bar{X}_{1:k} = \{\bar{X}_{1:k}^i\}_{i=1}^N$:
    for $j$ in $1:N$ do
        (i) sample index $i_j$ from a multinomial distribution with probabilities $W_k$
        (ii) $\bar{X}_{1:k}^j = X_{1:k}^{i_j}$
        (iii) $W_k^j = 1$
    end for
Estimate the marginal likelihood $P(Y_{1:k})$: $\hat{P}(Y_{1:k}) = \hat{P}(Y_1) \prod_{n=2}^k \hat{P}(Y_n|Y_{1:n-1})$.
```

The unnormalized auxiliary importance weights are given by

$$\tilde{g}_n = \frac{p_\theta\left(Y_{n+1}|x_n\right) p_\theta\left(x_{1:n}, Y_{1:n}\right)}{f(x_n|x_{n-1}) p_\theta\left(Y_n|x_{n-1}\right) p_\theta(x_{1:n-1}, Y_{1:n-1})} = \frac{p_\theta\left(Y_n|x_n\right)}{p_\theta\left(Y_n|x_{n-1}\right)} p_\theta\left(Y_{n+1}|x_n\right),$$

and the unnormalized weights for the approximation of $p_\theta\left(x_{1:n}|Y_{1:n}\right)$ given by

$$\tilde{w}_n = \frac{p_\theta\left(x_{1:n}, Y_{1:n}\right)}{f(x_n|x_{n-1}) p_\theta\left(Y_n|x_{n-1}\right) p_\theta(x_{1:n-1}, Y_{1:n-1})} = \frac{p_\theta\left(Y_n|x_n\right)}{p_\theta\left(Y_n|x_{n-1}\right)}.$$

## 2.6 Particle Marginal Metropolis-Hastings sampler

The Particle Marginal Metropolis-Hastings (PMMH) sampler is a particle MCMC method that uses the output of a particle algorithm to form the proposal density for a Metropolis-Hastings update [Andrieu et al., 2010]. The PMMH is as shown in Algorithm 3. The PMMH update leaves $p\left(\theta, x_{1:T}|Y_{1:T}\right)$ invariant.

**Algorithm 2** Auxiliary particle filter

```
Sample $N$ particles $\{X_1^j\}_{j=1}^N$ :
    for $j$ in $1:N$ do
        $X_1^j \sim \mu(x_1)$
    end for
Find the weights, $\tilde{w}_1 = \{\tilde{w}_1^j\}_{j=1}^N$:
    for $j$ in $1:N$ do
        $\tilde{w}_1^j = p_\theta(Y_1|x_1^j)$
    end for
Normalize the weights, $W_1 = \{W_1^j\}_{j=1}^N$:
    for $j$ in $1:N$ do
        $W_1^j = \frac{\tilde{w}_1^j}{\sum_{j=1}^{N} \tilde{w}_1^j}$
    end for
Estimate $P(Y_1) : \hat{P}(Y_1) = \frac{1}{N} \sum_{j=1}^{N} \tilde{w}_1^j$
for $n = 1, .., k$ do
    For each particle $j$, we calculate an estimate of $X_{n+1}^j$ called $\tilde{X}_{n+1}^j$ by drawing a sample from $f(x_{n+1}|x_n)$:
        for $j$ in $1:N$ do
            $\tilde{X}_{n+1}^j \sim f(x_{n+1}|x_n^j)$
        end for
    Find the auxiliary weights, $\tilde{g}_{n+1} = \{\tilde{g}_{n+1}^j\}_{j=1}^N$:
        for $j$ in $1:N$ do
            $\tilde{g}_{n+1}^j = g_n^j W_n^j p_\theta\left(Y_{n+1}|\tilde{x}_{n+1}^j\right)$
        end for
    Normalize the auxiliary weights, $g_{n+1} = \{g_{n+1}^j\}_{j=1}^N$:
        for $j$ in $1:N$ do
            $g_{n+1}^j = \frac{\tilde{g}_{n+1}^j}{\sum_{j=1}^{N} \tilde{g}_{n+1}^j}$
        end for
    Resample and form $N$ equally weighted particles, $\bar{X}_{1:n} = \{\bar{X}_{1:n}^i\}_{i=1}^N$:
        for $j$ in $1:N$ do
            (i) sample index $i_j$ from a multinomial distribution with probabilities $g_{n+1}$
            (ii) $\bar{X}_{1:n}^j = X_{1:n}^{i_j}$
            (iii) $g_{n+1}^j = 1$
        end for
    Using $\bar{X}_{1:n}$ propagate:
        for $j$ in $1:N$ do
            $X_{n+1}^j \sim f(x_{n+1}|x_n^j)$
            Set $X_{1:n+1}^j = \left(\bar{X}_{1:n}^j, X_{n+1}^j\right)$
        end for
    Find the weights, $\tilde{w}_{n+1} = \{\tilde{w}_{n+1}^j\}_{j=1}^N$:
        for $j$ in $1:N$ do
            $\tilde{w}_{n+1}^j = \frac{p_\theta\left(Y_{n+1}|x_{n+1}^j\right)}{p_\theta\left(Y_{n+1}|\tilde{x}_{n+1}^{i_j}\right)}$
        end for
    Normalize the weights, $W_{n+1} = \{W_{n+1}^j\}_{j=1}^N$:
        for $j$ in $1:N$ do
            $W_{n+1}^j = \frac{\tilde{w}_{n+1}^j}{\sum_{j=1}^{N} \tilde{w}_{n+1}^j}$
        end for
    Estimate $P(Y_{n+1}|Y_{1:n}) : \hat{P}(Y_{n+1}|Y_{1:n}) = \left(\sum_{i=1}^{N} W_n^i p_\theta(Y_{n+1}|\tilde{x}_{n+1}^i)\right)\left(\frac{1}{N}\sum_{i=1}^{N} \tilde{w}_{n+1}^i\right)$.
end for
Estimate the marginal likelihood $P(Y_{1:k})$: $\hat{P}(Y_{1:k}) = \hat{P}(Y_1) \prod_{n=2}^{k} \hat{P}(Y_n|Y_{1:n-1})$.
```

**Algorithm 3** Particle marginal Metropolis-Hastings sampler

```
Step 1: Initialization,$i = 0$,
    (a) set $\theta^{(0)}$ arbitrarily
    (b) - run a particle algorithm targetting $p_{\theta^{(0)}}(x_{1:T}|Y_{1:T})$
        -Sample $X_{1:T}(0) \sim \hat{p}_{\theta^{(0)}}(\cdot|Y_{1:T})$
        -Let $\hat{p}_{\theta^{(0)}}(Y_{1:T})$ denote the marginal likelihood estimate.
Step 2: for iteration $i \geq 1$
    (a) Sample $\theta^* \sim q\left(\cdot|\theta^{(i-1)}\right)$
    (b) - run a SMC algorithm targetting $p_{\theta^*}(x_{1:T}|Y_{1:T})$
        -Sample $X^*_{1:T} \sim \hat{p}_{\theta^*}(\cdot|Y_{1:T})$
        -Let $\hat{p}_{\theta^*}(Y_{1:T})$ denote the marginal likelihood estimate.
    (c) with probability $\min\left(1, \frac{\hat{p}_{\theta^*}(Y_{1:T})}{\hat{p}_{\theta^{(i-1)}}(Y_{1:T})} \frac{p(\theta^*)}{p(\theta^{(i-1)})} \frac{q(\theta^{(i-1)}|\theta^*)}{q(\theta^*|\theta^{(i-1)})}\right)$.
     set $\theta^{(i)} = \theta^*$, $X^{(i)}_{1:n} = X^*_{1:n}$, $\hat{p}_{\theta^{(i)}}(Y_{1:T}) = \hat{p}_{\theta^*}(Y_{1:T})$. Otherwise,
    $\theta^{(i)} = \theta^{(i-1)}$, $X^{(i)}_{1:n} = X^{(i-1)}_{1:n}$, $\hat{p}_{\theta^{(i)}}(Y_{1:T}) = \hat{p}_{\theta^{(i-1)}}(Y_{1:T})$.
```

## 2.7 Bayesian Additive Regression Trees

The Bayesian Additive Regression Trees (BART) model, introduced by Chipman et al. [2010], uses a sum of trees to model the conditional mean of a response Y given a $p$-dimensional covariate $x$, $E(Y|x)$.

Let $T_j$ denote a binary tree that partitions the predictor space into $b_j$ subregions denoted $T_j = \{\Omega_{jt}\}_{t=1}^{b_j}$, each associated with a leaf parameter $\mu_{jt}$. Let $M_j = \left(\mu_{j1}, \mu_{j2}, \ldots, \mu_{jb_j}\right)$ denote the collection of leaf parameters of tree $T_j$. We denote the set of all trees except $T_j$ as $T_{(j)} = \{T_k\}_{k=1,k\neq j}^{m}$ and their leaf parameters as $M_{(j)} = \{M_k\}_{k=1,k\neq j}^{m}$. Chipman et al. [2010] scale the response to lie between -0.5 and 0.5, and model $E(Y|x)$ as follows

$$
\begin{aligned}
Y &= \sum_{j=1}^{m} \sum_{t=1}^{b_j} \mu_{jt} \mathbb{1}\left(x \in \Omega_{jt}\right) \\
T_j &\sim \text{heterogeneous Galton-Watson process} \\
\mu_{jt}|T_j &\overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\mu^2),\ \sigma_\mu = 0.5/2\sqrt{m} \\
\sigma^2 &\sim \text{Inverse Chi-square distribution}
\end{aligned}
$$

Chipman et al. [2010] recommend $k = 2$ and $m = 200$ as automatic default choices.

**Regularization prior** The parameters of the model are the tree components $(\Lambda, T) = \{(T_j, \Lambda_j)\}_{j=1}^{m}$, and variance $\sigma^2$. Chipman et al. [2010] assume that the tree components are independent of each other and of $\sigma^2$, and that the leaf parameters of every tree are independent of each other, so that the prior can be factorized as

$$P(\Lambda, T, \sigma) = \left( \prod_{j=1}^{m} P(\Lambda_j | T_j) P(T_j) \right) P(\sigma) = \left( \prod_{j=1}^{m} \prod_{t=1}^{b_j} P(\lambda_{jt} | T_j) P(T_j) \right) P(\sigma).$$

The imposed prior on trees is a Galton-Watson process in which each node has either zero or two offspring, and a node $\eta$ splits with probability

$$p_{split}(\eta) = \frac{\alpha}{(1 + d(\eta))^{\beta}} \ \alpha \in (0,1), \beta \in [0, \infty),$$

where $d(\eta)$ is the depth of node $\eta$. The default prior specification is $\alpha = 0.95$ and $\beta = 2$. The imposed prior on leaf parameters and $\sigma^2$ are the conjugate priors; that is, the normal distribution and the inverse chi-square distribution, respectively.

**Inference Method** Chipman et al. [2010] employ a Gibbs sampler to sample from $P\left((T_1, M_1), (T_2, M_2), \ldots, (T_m, M_m), \sigma | Y\right)$. The sampler requires m successive draws from $(T_j, M_j)|T_{(j)}, M_{(j)}, \sigma, Y$ followed by a draw from $\sigma | T_1, M_1, \ldots, T_m, M_m, Y$ where $R_j$ is the partial residual based on a fit that excludes the $j_{th}$ tree, given by $R_j = Y - \sum_{k \neq j} \sum_{t=1}^{b_k} \mu_{kt} \mathbb{1}\left(x \in \Omega_{kt}\right)$. A draw from $(T_j, M_j)|R_j, \sigma$ can be achieved in $(b_j + 1)$ successive steps as

- sampling $T_j | R_j, \sigma$ using a Metropolis-Hastings algorithm and
- sampling $\mu_{jt} | T_j, R_j, \sigma$ from a normal distribution for $t = 1, \ldots, b_j$.

The transition kernel $q$ of the proposed Metropolis-Hastings is chosen from four proposals: GROW, PRUNE, CHANGE and SWAP. The GROW proposal randomly picks a terminal node, splits it into two new nodes, and assigns a decision rule to it. The PRUNE proposal randomly picks a parent of two terminal nodes and turns it into a terminal node by collapsing the nodes below it. The CHANGE proposal randomly picks an internal node and reassigns to it a splitting rule. The SWAP proposal randomly picks a parent-child pair that are both internal nodes, and swaps their splitting rules.

**Simulation Analysis** Chipman et al. [2010] compared BART with boosting, neural nets and random forests on 42 different real data sets. They created 20 independent train/test splits for each of the data sets. The analysis illustrated similar performance between BART and the other alternative algorithms when they used the default values of BART hyperparameters. BART performance was better than the alternatives when they applied cross-validation to define BART hyperparameters.

Various extensions of BART have been proposed. Here, we focus on log-linear BART for count and categorical data introduced by Murray [2021].

### 2.7.1 Log-linear BART for count data

The original BART model assumes Gaussian data and has been applied in a wide range of applications. Murray [2021] adapted BART to count data $Y$ via a log-linear transformation. Let $\Lambda_j = \left(\lambda_{j1}, \lambda_{j2}, \dots, \lambda_{jb_j}\right)$ denote the collection of leaf parameters of tree $T_j$. Murray [2021] models $E(Y|x)$ as follows

$$\begin{aligned} E(Y|x) &= \prod_{j=1}^{m} \prod_{t=1}^{b_j} \lambda_{jt}^{\mathbb{1}(x \in \Omega_{jt})} \\ T_j &\sim \text{heterogeneous Galton-Watson process} \\ \lambda_{jt}|T_j &\overset{\text{iid}}{\sim} \text{mixture of generalized inverse Gaussian distributions} \end{aligned}$$

Murray [2021] applies an efficient block Gibbs sampler to sample from $P\left((T_1, \Lambda_1), (T_2, \Lambda_2), \dots, (T_m, \Lambda_m)|Y\right)$. The sampler requires $m$ successive draws from $(T_j, \Lambda_j)|T_{(j)}, \Lambda_{(j)}, Y$.

# Chapter 3

# BART-based inference for Poisson processes

The Bayesian Additive Regression Trees (BART) model is a Bayesian framework, which uses a sum of trees to predict the posterior distribution of a response $y$ given a $p$-dimensional covariate $X$ and priors on the function relating the covariates to the response.

Chipman et al. [2010] proposed an inference procedure using Metropolis Hastings within a Gibbs Sampler, whereas Lakshminarayanan et al. [2015] used a Particle Gibbs Sampler to increase mixing when the true posterior consists of deep trees or when the dimensionality of the data is high. Several theoretical studies of BART models [Ročková and van der Pas, 2020; Ročková and Saha, 2019; Linero and Yang, 2018] have recently established optimal posterior convergence rates.

The BART model has been applied in various contexts including non-parametric mean regression [Chipman et al., 2010], classification [Chipman et al., 2010; Zhang and Härdle, 2010; Kindo et al., 2016], variable selection[Chipman et al., 2010; Bleich et al., 2014; Linero, 2018], estimation of monotone functions [Chipman et al., 2021], causal inference [Hill, 2011], survival analysis [Sparapani et al., 2016], and heteroscedasticity [Bleich and Kapelner, 2014; Pratola et al., 2016].

Linero and Yang [2018] illustrated how the BART model suffers from a lack of smoothness and the curse of dimensionality, and overcome both potential shortcomings by considering a sparsity assumption similar to [Linero, 2018] and treating decisions at branches probabilistically.

The original BART model [Chipman et al., 2010] assume that the response has a Gaussian distribution and the majority of applications have used this framework. Murray [2021] adapted the BART model to count data and categorical data via a log-linear transformation, and provided an efficient MCMC sampler. Our focus here is on extending this methodology to estimate the intensity function of inhomogeneous Poisson processes.

The question of estimating the intensity of Poisson processes has a long history, including both frequentist and Bayesian methods. Frequentist methods include fixed-bandwidth and adaptive bandwidth kernel estimators with edge correction [Diggle et al., 2003], and wavelet-based methods [e.g. Fryzlewicz and Nason, 2004; Patil et al., 2004]. Bayesian methods include using a sigmoidal Gaussian Cox process model for intensity inference [Adams et al., 2009], a Markov random field (MRF) with Laplace prior [Sardy and Tseng, 2004], variational Bayesian intensity inference [Lloyd et al., 2015], and non-parametric Bayesian estimations of the intensity via piecewise functions with either random or fixed partitions of constant intensity [Arjas and Gasbarra, 1994; Heikkinen and Arjas, 1998; Gugushvili et al., 2018].

In this chapter, we introduce an extension of the BART model [Chipman et al., 2010] for Poisson processes whose intensity at each point is estimated via an ensemble of trees. Specifically, the logarithm of the intensity at each point is modelled via a sum of trees (and hence the intensity is a product of trees). This approach enables full posterior inference of the intensity in a non-parametric regression setting. Our main contribution is a novel BART scheme for estimating the intensity of an inhomogeneous Poisson process. The simulation studies demonstrate that our algorithm is competitive with the Haar-Fisz algorithm in one dimension, kernel smoothing in two dimensions, and outperforms the kernel approach for multidimensional intensities. The simulation analysis also demonstrates that our proposed algorithm is competitive with the inference via spatial log-Gaussian Cox processes. We also demonstrate its ability to track varying intensity in synthetic and real data.

The outline of the chapter is as follows. Section 3.1 introduces our approach for estimating the intensity of a Poisson process through the BART model, and Section 3.2 presents the proposed inference algorithm. Sections 3.3 and 3.4 present the application of the algorithm to synthetic data and real data sets, respectively. Section 3.5 provides our conclusions.

## 3.1 The BART Model for Poisson processes

Consider an inhomogeneous Poisson process defined on a $d$-dimensional domain $S \subset \mathbb{R}^d$, $d \geq 1$, with intensity $\lambda : S \to \mathbb{R}^+$. For such a process, the number of points within a subregion $B \subset S$ has a Poisson distribution with mean $\lambda_B = \int_B \lambda(s)\,\mathrm{d}s$, and the number of points in disjoint subregions are independent [Daley and Vere-Jones, 2003]. The homogeneous Poisson process is a special case with constant intensity $\lambda(s) = \lambda_0$, $\forall s \in S$.

To estimate the intensity of the inhomogeneous Poisson process, we use $m$ partitions of the domain $S$, each associated with a tree $T_h$, $h = 1, \ldots, m$. The partitions are denoted $T_h = \{\Omega_{ht}\}_{t=1}^{b_h}$, where $b_h$ is the number of terminal nodes in the corresponding tree $T_h$, and each leaf node $t$ corresponds to one of the subregions $\Omega_{ht}$ of the partition $T_h$. Being a partition, every tree covers the full domain, i.e. $S = \cup_{t=1}^{b_h} \Omega_{ht}$ for every $h$. Each subregion $\Omega_{ht}$ has an associated parameter $\lambda_{ht}$, and hence each tree $T_h$ has an associated vector of leaf intensities $\Lambda_h = (\lambda_{h1}, \lambda_{h2}, .., \lambda_{hb_h})$.

We model the intensity of $s \in S$ as:

$$\begin{aligned}
\log(\lambda(s)) &= \sum_{h=1}^{m} \sum_{t=1}^{b_h} \log\left(\lambda_{ht}\right)\, I(s \in \Omega_{ht}) && (3.1)\\
T_h &\sim \text{heterogeneous Galton-Watson process for a partition of } S && (3.2)\\
\lambda_{ht} | T_h &\overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta) && (3.3)
\end{aligned}$$

where $I(\cdot)$ denotes the indicator function. Equivalently, (3.1) can be expressed as

$$\lambda(s) = \prod_{h=1}^{m} \prod_{t=1}^{b_h} \lambda_{ht}^{I(s \in \Omega_{ht})}. \tag{3.4}$$

Given a fixed number of trees, $m$, the parameters of the model are thus the regression trees $T = \{T_h\}_{h=1}^{m}$ and their corresponding intensities $\Lambda = \{\Lambda_h\}_{h=1}^{m}$. Following Chipman et al. [2010], we assume that the tree components $(T_h, \Lambda_h)$ are independent of each other, and that the terminal node parameters of every tree are independent, so that the prior can be factorized as:

$$P(\Lambda, T) = \prod_{h=1}^{m} P(\Lambda_h, T_h) = \prod_{h=1}^{m} P(\Lambda_h | T_h) P(T_h) = \prod_{h=1}^{m} \left[ \prod_{t=1}^{b_h} P(\lambda_{ht} | T_h) \right] P(T_h). \tag{3.5}$$

**Prior on the trees** The trees $T_h$ of the BART model are stochastic regression trees generated through a heterogeneous Galton-Watson (GW) process [Harris, 1963; Ročková and Saha, 2019]. The GW process is the simplest branching process concerning the evolution of a population in discrete time. Individuals (tree nodes) of a generation (tree depth) give birth to a random number of individuals (tree nodes), called offspring, mutually independent and all with the same offspring distribution that may vary from generation (depth) to generation (depth). In our case, we use the prior introduced by Chipman et al. [1998], that is a GW process in which each node has either zero or two offspring and the probability of a node splitting depends on its depth in the tree. Specifically, a node $\eta \in T_h$ splits into two offspring with probability

$$p_{\text{split}}(\eta) = \frac{\gamma}{\left(1 + d(\eta)\right)^{\delta}}, \tag{3.6}$$

where $d(\eta)$ is the depth of the node $\eta$ in the tree, and $\gamma \in (0, 1)$ and $\delta \geq 0$ are parameters of the model. Classic results from the theory of branching processes show that $\gamma \leq 0.5$ guarantees that the expected depth of the tree is finite. In our construction, each tree $T_h$ is associated with a partition of $S$. Namely, if node $\eta$ splits, we select uniformly at random one of the $d$ dimensions of the space of the Poisson process, followed by uniform selection from the available split values associated with that dimension respecting the splitting rules higher in the tree.

**Prior on the leaf intensities** Our choice of a Gamma prior for the leaf parameters $\lambda_{ht}$ builds upon previous work by Murray [2021], who used a mixture of Generalized Inverse Gaussian (GIG) distributions as the prior on leaf parameters in a BART model for count regression. Here we impose a Gamma prior (a special case of GIG) on the leaf parameters, which simplifies the model and leads to a closed form of the conditional integrated likelihood below (see Section 3.2) as the Gamma distribution is the conjugate prior for the Poisson likelihood. We discuss the selection of it hyperparameters $\alpha$ and $\beta$ in Section 3.2.1.

## 3.2 The Inference Algorithm

Given a finite realization of an inhomogeneous Poisson process with $n$ sample points $\boldsymbol{s} = s_1, \dots, s_n \in S \subset \mathbb{R}^d$, we seek to infer the parameters of the model $(\Lambda, T)$ by sampling from the posterior $P(\Lambda, T|\boldsymbol{s})$.

Before presenting the sampling algorithm, we summarize a preliminary result. To simplify our notation, let us define

$$g(s_i; T_h, \Lambda_h) = \prod_{t=1}^{b_h} \lambda_{ht}^{I(s_i \in \Omega_{ht})},$$

so that Eq. (3.4) becomes $\lambda(s_i) = \prod_{h=1}^{m} g(s_i; T_h, \Lambda_h)$.

Let us choose any arbitrary tree $T_h$ in our ensemble $T$, and let us denote the set with the rest of the trees as $T_{(h)} = \{T_j\}_{j=1, j\neq h}^{m}$ and their leaf parameters as $\Lambda_{(h)} = \{\Lambda_j\}_{j=1, j\neq h}^{m}$. The intersection of all the partitions associated with the trees in $T_{(h)}$ gives us a global partition $\{\overline{\Omega}_k^{(h)}\}_{k=1}^{K(T_{(h)})}$ with $K(T_{(h)})$ subregions [Rockova and Pas, 2017].

Then we have the following result.

**Remark 1.** *(i) The conditional likelihood of the realization is given by*

$$P(\boldsymbol{s}|\Lambda, T) = c_h \prod_{t=1}^{b_h} \lambda_{ht}^{n_{ht}} \, e^{-\lambda_{ht} c_{ht}}, \tag{3.7}$$

$$\text{with} \quad c_h = \prod_{i=1}^{n} \prod_{j=1, j\neq h}^{m} g(s_i; T_j, \Lambda_j),$$

$$c_{ht} = \sum_{k=1}^{K(T_{(h)})} \overline{\lambda}_k^{(h)} |\overline{\Omega}_k^{(h)} \cap \Omega_{ht}|,$$

*where* $\overline{\lambda}_k^{(h)} = \prod_{t=1, t\neq h}^{m} \prod_{l=1}^{b_t} \lambda_{tl}^{I(\Omega_{tl} \cap \overline{\Omega}_k^{(h)} \neq 0)}$, $n_{ht}$ *is the cardinality of the set* $\{i : s_i \in \Omega_{ht}\}$, *and* $|\overline{\Omega}_k^{(h)} \cap \Omega_{ht}|$ *is the volume of the region* $\overline{\Omega}_k^{(h)} \cap \Omega_{ht}$.

*(ii) For a tree $h$, the conditional integrated likelihood obtained by integrating out $\Lambda_h$ is*

$$P(\boldsymbol{s}|T_h, T_{(h)}, \Lambda_{(h)}) = c_h \left( \frac{\beta^{\alpha}}{\Gamma(\alpha)} \right)^{b_h} \prod_{t=1}^{b_h} \frac{\Gamma(n_{ht} + \alpha)}{(c_{ht} + \beta)^{n_{ht} + \alpha}}. \tag{3.8}$$

*Proof.* Let us consider a finite realization of an inhomogeneous Poisson process with $n$ points $\boldsymbol{s}$. Given the tree components $(T, \Lambda)$, and approximating the intensity of a point $s_i \in S$ by a product of $m$ trees $\lambda(s_i) = \prod_{j=1}^{m} g(s_i; T_j, \Lambda_j)$, the likelihood is:

$$\begin{aligned} P(\boldsymbol{s}|\Lambda, T) &= \prod_{i=1}^{n} \lambda(s_i) \exp\left(-\int_S \lambda(s) ds\right) \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m} g(s_i; T_j, \Lambda_j) \exp\left(-\int_S \prod_{j=1}^{m} g(s; T_j, \Lambda_j) ds\right). \end{aligned} \tag{3.9}$$

The first term of the above equation can be written as follows

$$\begin{aligned} \prod_{i=1}^{n}\prod_{j=1}^{m} g(s_i; T_j, \Lambda_j) &= \prod_{i=1}^{n} \prod_{j=1, j\neq h}^{m} g(s_i; T_j, \Lambda_j) g(s_i; T_h, \Lambda_h) \\ &= \prod_{i=1}^{n} \prod_{j=1, j\neq h}^{m} g(s_i; T_j, \Lambda_j) \left(\prod_{i=1}^{n} g(s_i; T_h, \Lambda_h)\right) = c_h \prod_{t=1}^{b_h} \lambda_{ht}^{n_{ht}} \end{aligned}$$

where $c_h = \prod_{i=1}^{n} \prod_{j=1, j\neq h}^{m} g(s_i; T_j, \Lambda_j)$ and $n_{ht}$ is the cardinality of the set $\{i : s_i \in \Omega_{ht}\}$.

The exponential term of (3.9) can be expressed as:

$$\begin{aligned} \exp\left(-\int_S \prod_{j=1}^{m} g(s; T_j, \Lambda_j) ds\right) &= \exp\left(-\int_S \prod_{j=1, j\neq h}^{m} g(s; T_j, \Lambda_j) g(s; T_h, \Lambda_h) ds\right) \\ &= \exp\left(-\int_S \prod_{j=1, j\neq h}^{m} g(s; T_j, \Lambda_j) \left(\sum_{t=1}^{b_h} \lambda_{ht} I(s \in \Omega_{ht})\right) ds\right) \\ &= \exp\left(-\int_S \sum_{t=1}^{b_h} \lambda_{ht} \prod_{j=1, j\neq h}^{m} g(s; T_j, \Lambda_j) I(s \in \Omega_{ht}) ds\right) \end{aligned}$$

Tonelli's theorem allows the change of order between summation and integral.

$$\begin{aligned} \exp\left(-\int_S \prod_{j=1}^{m} g(s; T_j, \Lambda_j) ds\right) &= \exp\left(-\sum_{t=1}^{b_h} \lambda_{ht} \int_S \prod_{j=1, j\neq h}^{m} g(s; T_j, \Lambda_j) I(s \in \Omega_{ht}) ds\right) \\ &= \exp\left(-\sum_{t=1}^{b_h} \lambda_{ht} c_{ht}\right) \end{aligned}$$

where

$$c_{ht} = \int_S \left(\prod_{j=1, j\neq h}^{m} g(s; T_j, \Lambda_j)\right) I(s \in \Omega_{ht}) ds.$$

Let $T_{(h)} = \{T_j\}_{j=1, j\neq h}^{m}$ be an ensemble of trees not including the tree $T_h$ that defines the global partition $\{\overline{\Omega}_k^{(h)}\}_{k=1}^{K(T_{(h)})}$ by merging all cuts in $\{T_j\}_{j=1, j\neq h}^{m}$. Giving,

$$\prod_{j=1, j\neq h}^{m} g(s; T_j, \Lambda_j) = \sum_{k=1}^{K(T_h)} \overline{\lambda}_k^{(h)} I(s \in \overline{\Omega}_k^{(h)})$$

where

$$\overline{\lambda}_k^{(h)} = \prod_{t=1,t\neq h}^{m} \prod_{l=1}^{b_t} \lambda_{tl}^{I(\Omega_{tl}\cap\overline{\Omega}_k^{(h)}\neq\emptyset)},$$

leading to the following expression for $c_{ht}$,

$$\begin{aligned} c_{ht} &= \int_S \left( \prod_{j=1,j\neq h}^{m} g(s,T_j,\Lambda_j) \right) I(s\in\Omega_{ht})ds = \int_S \left( \sum_{k=1}^{K(T_{(h)})} \overline{\lambda}_k^{(h)} I(s\in\overline{\Omega}_k^{(h)}) \right) I(s\in\Omega_{ht})ds \\ &= \sum_{k=1}^{K(T_{(h)})} \overline{\lambda}_k^{(h)} \int_S I(s\in\overline{\Omega}_k^{(h)}\cap\Omega_{ht})ds = \sum_{k=1}^{K(T_{(h)})} \overline{\lambda}_k^{(h)} |\overline{\Omega}_k^{(h)}\cap\Omega_{ht}|, \end{aligned}$$

where $|\overline{\Omega}_k^{(h)}\cap\Omega_{ht}|$ is the volume of the region $\Omega_k^{(h)}\cap\Omega_{ht}$. Hence the conditional likelihood can be written as follows

$$P(\boldsymbol{s}|\Lambda,T) = c_h \prod_{t=1}^{b_h} \lambda_{ht}^{n_{ht}} e^{-\lambda_{ht}c_{ht}}.$$

The conditional integrated likelihood is given by

$$\begin{aligned} P(\boldsymbol{s}|T_h,T_{(h)},\Lambda_{(h)}) &= \int_0^\infty P(\boldsymbol{s},\Lambda_h|T_h,T_{(h)},\Lambda_{(h)})d\Lambda_h \\ &= \int_0^\infty P(\boldsymbol{s}|\Lambda,T)\, P(\Lambda_h|T_h,T_{(h)},\Lambda_{(h)})d\Lambda_h \\ &= c_h \int_0^\infty \cdots \int_0^\infty \prod_{t=1}^{b_h} \lambda_{ht}^{n_{ht}} e^{-\lambda_{ht}c_{ht}} \prod_{t=1}^{b_h} \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\lambda_{ht}} \lambda_{ht}^{\alpha-1} d\lambda_{h1}\dots d\lambda_{hb_h} \\ &= c_h \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{b_h} \prod_{t=1}^{b_h} \int_0^\infty \lambda_{ht}^{n_{ht}+\alpha-1} e^{-(c_{ht}+\beta)\lambda_{ht}} d\lambda_{ht} \\ &= c_h \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{b_h} \prod_{t=1}^{b_h} \frac{\Gamma(n_{ht}+\alpha)}{(c_{ht}+\beta)^{n_{ht}+\alpha}} \end{aligned}$$

□

We now summarize our sampling algorithm. To sample from $P(\Lambda,T|\boldsymbol{s})$, we implement a Metropolis-Hastings within block Gibbs sampler (Algorithm 4), which requires $m$ successive draws from $(T_h, \Lambda_h)|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s}$. Note that

$$\begin{aligned}
P(T_h, \Lambda_h|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s}) &= P(T_h|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s})\, P(\Lambda_h|T_h, T_{(h)}, \Lambda_{(h)}, \boldsymbol{s}) \\
&\propto P(T_h|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s})\, P(\boldsymbol{s}|\Lambda, T) P(\Lambda_h|T_h) \\
&= P(T_h|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s})\, P(\boldsymbol{s}|\Lambda, T) \prod_{t=1}^{b_h} P(\lambda_{ht}|T_h) \\
&= P(T_h|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s})\, c_h \prod_{t=1}^{b_h} \lambda_{ht}^{n_{ht}}\, e^{-\lambda_{ht} c_{ht}} \prod_{t=1}^{b_h} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_{ht}^{\alpha-1}\, e^{-\beta\lambda_{ht}} \\
&\propto P(T_h|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s}) \prod_{t=1}^{b_h} \lambda_{ht}^{n_{ht}+\alpha-1} e^{-(c_{ht}+\beta)\lambda_{ht}} \qquad (3.10)
\end{aligned}$$

which follows directly from Bayes' rule and Eqs. (3.5) and (3.3).

From (3.10), it is clear that a draw from $(T_h, \Lambda_h)|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s}$ can be achieved in $(b_h+1)$ successive steps consisting of:

- sampling $T_h|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s}$ using Metropolis-Hastings (Algorithm 5)
- sampling $\lambda_{ht}|T_h, T_{(h)}, \Lambda_{(h)}, \boldsymbol{s}$ from a Gamma distribution with shape $n_{ht} + \alpha$ and rate $c_{ht} + \beta$ for $t = 1, .., b_h$.

These steps are implemented through Metropolis-Hastings in Algorithm 4. Note also that

$$P(T_h|T_{(h)}, \Lambda_{(h)}, \boldsymbol{s}) \propto P(\boldsymbol{s}|T_h, T_{(h)}, \Lambda_{(h)})\, P(T_h),$$

so that the conditional integrated likelihood (3.8) is required to compute the Hastings ratio.

**Algorithm 4** Metropolis-Hastings within Gibbs sampler

```
for $v = 1, 2, 3, ..$ do
  for $h = 1$ to $m$ do
    Sample $T_h^{(v+1)}|\boldsymbol{s}, \{T_j^{(v+1)}\}_{j=1}^{h-1}, \{T_j^{(v)}\}_{j=h+1}^{m}, \{\Lambda_j^{(v+1)}\}_{j=1}^{h-1}, \{\Lambda_j^{(v)}\}_{j=h+1}^{m}$
     using Algorithm 5
    for $t = 1$ to $b_h$ do
      Sample $\lambda_{ht}^{(v+1)}|\boldsymbol{s}, \{T_j^{(v+1)}\}_{j=1}^{h}, \{T_j^{(v)}\}_{j=h+1}^{m}, \{\Lambda_j^{(v+1)}\}_{j=1}^{h-1}, \{\Lambda_j^{(v)}\}_{j=h+1}^{m}$
      from Gamma($n_{ht} + \alpha$, $c_{ht} + \beta$)
    end for
  end for
end for
```

**Algorithm 5** Metropolis-Hastings Algorithm for sampling from the posterior $P(T_j|\boldsymbol{s}, T_{(j)}, \Lambda_{(j)})$

```
Generate a candidate value $T_j^*$ with probability $q(T_j^*|T_j^{(v)})$.
Set $T_j^{(v+1)} = T_j^*$ with probability
    $\alpha(T_j^{(v)}, T_j^*) = \min\left\{1, \frac{q(T_j^{(v)}|T_j^*)}{q(T_j^*|T_j^{(v)})} \frac{P(\boldsymbol{s}|T_j^*, T_{(j)}, \Lambda_{(j)})}{P(\boldsymbol{s}|T_j^{(v)}, T_{(j)}, \Lambda_{(j)})} \frac{P(T_j^*)}{P(T_j^{(v)})}\right\}$
Otherwise, set $T_j^{(v+1)} = T_j^{(v)}$.
```

The transition kernel $q$ in Algorithm 5 is chosen from the three proposals: GROW, PRUNE, CHANGE [Chipman et al., 2010; Kapelner and Bleich, 2016]. The GROW proposal randomly picks a terminal node, splits the chosen terminal into two new nodes and assigns a decision rule to it. The PRUNE proposal randomly picks a parent of two terminal nodes and turns it into a terminal node by collapsing the nodes below it. The CHANGE proposal randomly picks an internal node and randomly reassigns to it a splitting rule. We describe the implementation of the proposals in Subsection 3.2.2.

For completeness, in Appendix A.1, we present the full development of the algorithm for inference of the intensity of inhomogeneous Poisson processes via only one tree.

### 3.2.1 Fixing the hyperparameters of the model

**Hyperparameters of the Gamma distribution for the leaf intensities** We use a simple data-informed approach to fix the hyperparameters $\alpha$ and $\beta$ of the Gamma distribution (3.3). We discretize the domain into $N_G$ subregions of equal volume ($N_G = (\lceil 100^{1/d} \rceil)^d$ works well in practice up to 5 dimensions) and count the number of samples $s_i$ per subregion. We thus obtain the empirical densities in each of the subregions: $\xi_i$, $i = 1, \dots, N_G$. Given the form of the intensity (3.4) as a product of $m$ trees, we consider the $m$-th roots $\Xi = \{\xi_i^{1/m}\}_{i=1}^{N_G}$ as candidates for the intensity of each tree. Taking the sample mean $\widehat{\mu}_\Xi$ and sample variance $\widehat{\sigma}^2_\Xi$, we choose the model hyperparameters $\alpha$ and $\beta$ to correspond to those of a Gamma distribution with the same mean and variance, i.e., $\alpha = \widehat{\mu}^2_\Xi / \widehat{\sigma}^2_\Xi$ and $\beta = \widehat{\mu}_\Xi / \widehat{\sigma}^2_\Xi$, although fixing $\beta = 1$ can also give good estimates of the intensity. Although setting $N_G = (\lceil 100^{1/d} \rceil)^d$ leads to convergence and good estimates of the intensity in our simulation studies below, there are other possibilities. Alternatively, we can bin the data based on a criterion that takes into account the number of samples, $n$, and the number of dimensions, $d$. For example, the number of bins per dimension, $n_b$, can be computed as [Scott, 2015; Wand, 1997]: (i) $n_b = \lceil n^{1/(d+1)} \rceil$, (ii) $n_b = \lceil n^{1/(d+2)} \rceil$, or (iii) $n_b = \max_{k \in \{1,2,..,d\}} [\lceil DR_k \cdot n^{1/(d+2)} / (2 \cdot \mathrm{IQR}(\{s_{i,k}\})) \rceil]$, where IQR denotes the interquartile range of the sample, $DR_k$ is the range of the domain in dimension $k$ (here we scale the initial domain to a unit hypercube so that $DR_k = 1$, $\forall k$), and by extension $N_G = {n_b}^d$. In our simulation scenarios below, all these approaches lead to comparable convergence times and estimates of the intensity.

**Hyperparameters of the stochastic ensemble of regression trees** The GW stochastic process that generates our tree ensemble has several hyperparameters. The parameters $(\gamma, \delta)$ control the shape of trees. The parameter $\gamma > 0$ controls the probability that the root of a tree will split into two offspring, while the parameter $\delta > 0$ penalizes against deep trees. As noted in [Chipman et al., 2010], for a sum-of-trees model, we want to keep the depth of the tree small whilst ensuring non-trivial trees, hence, in our simulation study we fix $\gamma = 0.98$ and $\delta = 2$. Second, each of the $d$ dimensions has to be assigned a grid of split values, from which the subregions of the partition are randomly chosen, yet always respecting the consistency of the ancestors in the tree (that is respecting the splitting rules higher in the tree). Here, we use a simple uniform grid for each of the $d$-dimensions [Pratola et al., 2016]: we normalize each dimension of the space from (0,1) and discretize each dimension into $N_d$ segments. ($N_d = 100$ works well in practice and is used throughout our examples.) More sophisticated, data-informed grids are also possible, although using, e.g., the sample points as split values does not improve noticeably the performance in our examples. Finally, the number of trees $m$ also needs to be fixed as in Chipman et al. [2010]. In our examples below, we have checked the performance of our algorithm with varying number of trees $m$ between 2 and 50. We find that good performance can be achieved with a moderate number of trees, $m$, between 3 and 10 depending on the particular example.

### 3.2.2 Metropolis Hastings Proposals

We describe here the proposals of Algorithm 5. The Hastings ratio can be expressed as the product of three terms [Kapelner and Bleich, 2016]:

- Transition Ratio:
$$\mathrm{TR} = \frac{q(T_j^{(t)} | T_j^*)}{q(T_j^* | T_j^{(t)})}$$
- Likelihood Ratio:
$$\mathrm{LR} = \frac{P(\boldsymbol{s} | T_j^*, T_{(j)}, \Lambda_{(j)})}{P(\boldsymbol{s} | T_j^{(t)}, T_{(j)}, \Lambda_{(j)})}$$
- Tree Structure Ratio:
$$\mathrm{TSR} = \frac{P(T_j^*)}{P(T_j^{(t)})}$$

#### 3.2.2.1 GROW Proposal

This proposal randomly picks a terminal node, splits the chosen terminal into two new nodes and assigns a decision rule to it.

Let $\eta$ be the randomly picked terminal node in tree $T_j^{(t)}$. We denote the new nodes as $\eta_L$ and $\eta_R$. We now derive the expressions for the transition ratio (TR), tree structure ratio (TSR) and likelihood ratio (LR).

**Transition Ratio** It holds that:

(i) $q(T_j^*|T_j^{(t)}) = \mathrm{P(GROW)}$
$\times$ P(selecting a leaf $\eta$ to grow from)
$\times$ P(selecting an available dimension $j$ to split on)
$\times$ P(selecting the slitting value given the chosen dimension to split on)
$= \mathrm{P(GROW)}\ \frac{1}{b_j}\frac{1}{\mathrm{card}(k_\eta)}\frac{1}{\mathrm{card}(\tau_\eta)}$
where $b_j$ is the number of terminal nodes in the tree $T_j^{(t)}$, $k_h$ the set of all available dimensions to split the node $\eta$, $\tau_\eta$ the set of all available splitting values given the chosen dimension for splitting the node $\eta$ and $\mathrm{card}(S)$ the cardinality of a set $S$.

(ii) $q(T_j^{(t)}|T_j^*) = \mathrm{P(PRUNE)}$
$\times$ P(selecting a node $\eta$ having two terminal nodes to prune from)
$= \mathrm{P(PRUNE)}\ \frac{1}{w^*}$
where $w^*$ is the number of internal nodes with two terminal nodes as children in the tree $T_j^*$.

Hence the transition ratio is given by

$$\mathrm{TR} = \frac{P(\mathrm{PRUNE})\frac{1}{w^*}}{P(\mathrm{GROW})\frac{1}{b_j}\frac{1}{\mathrm{card}(k_\eta)}\frac{1}{\mathrm{card}(\tau_\eta)}}.$$

**Tree Structure Ratio:** The difference between the structures of the proposed tree $T_j^{(t)}$ and the tree $T_j^*$ is the two offsprings $\eta_L$ and $\eta_R$. Thus the tree structure ratio is:

$$\begin{aligned}\mathrm{TSR} = \frac{P(T_j^*)}{P(T_j^{(t)})} &= \frac{(1-p_{\mathrm{SPLIT}}(\eta_L))\,(1-p_{\mathrm{SPLIT}}(\eta_R))\,p_{\mathrm{SPLIT}}(\eta)\,p_{RULE}(\eta)}{(1-p_{\mathrm{SPLIT}}(\eta))}\\ &= \frac{\left(1-\frac{\gamma}{(1+d(\eta_L))^\delta}\right)\left(1-\frac{\gamma}{(1+d(\eta_R))^\delta}\right)\frac{\gamma}{(1+d(\eta))^\delta}\frac{1}{\mathrm{card}(k_\eta)}\frac{1}{\mathrm{card}(\tau_\eta)}}{1-\frac{\gamma}{(1+d(\eta))^\delta}},\end{aligned}$$

where $p_{\mathrm{SPLIT}}(\eta)$ is the splitting probability for a node $\eta$ and $p_{RULE}(\eta)$ the distribution of decision rule associated to node $\eta$.

**Likelihood Ratio** The likelihood ratio is an application of equation 3.8 twice, that is once considering the proposed tree, $T_j^*$ (numerator) and the other considering the tree of the current iteration $t$, $T_j^{(t)}$ (denominator), which can be simplified as follows

$$\begin{aligned}\mathrm{LR} &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\frac{\Gamma(n_{j\eta_L}+\alpha)}{(c_{j\eta_L}+\beta)^{n_{j\eta_L}+\alpha}}\frac{\Gamma(n_{j\eta_R}+\alpha)}{(c_{j\eta_R}+\beta)^{n_{j\eta_R}+\alpha}}}{\frac{\Gamma(n_{j\eta}+\alpha)}{(c_{j\eta}+\beta)^{n_{j\eta}+\alpha}}}\\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(n_{j\eta_L}+\alpha)\Gamma(n_{j\eta_R}+\alpha)}{\Gamma(n_{j\eta}+\alpha)}\frac{(c_{j\eta}+\beta)^{n_{j\eta}+\alpha}}{(c_{j\eta_L}+\beta)^{n_{j\eta_L}+\alpha}(c_{j\eta_R}+\beta)^{n_{j\eta_R}+\alpha}}\end{aligned}$$

#### 3.2.2.2 PRUNE Proposal

This proposal randomly picks a parent of two terminal nodes and turns it into a terminal node by collapsing the nodes below it.

Let $\eta$ be the picked parent of two terminal nodes, $y$ and $c$ the dimension and splitting value of the rule linked to the node $\eta$.

**Transition Ratio** It holds that:

(i) $$\begin{aligned} q(T_j^*|T_j^{(t)}) &= \text{P(PRUNE)} \\ &\quad \times \text{P(selecting a parent of two terminal nodes to prune from)} \\ &= \text{P(PRUNE)}\ \frac{1}{w} \end{aligned}$$

where $w$ is the number of nodes with two terminal nodes as children in the tree $T_j^{(t)}$.

(ii) $$\begin{aligned} q(T_j^{(t)}|T_j^*) &= \text{P(GROW)} \\ &\quad \times \text{P(selecting the node } \eta \text{ to grow from)} \\ &\quad \times \text{P(selecting the dimension } y) \\ &\quad \times \text{P(selecting the slitting value c given the chosen dimension } y) \\ &= \text{P(GROW)}\ \frac{1}{w^*}\frac{1}{\text{card}(k_\eta)}\frac{1}{\text{card}(\tau_\eta)} \end{aligned}$$

where $w^*$ is the number of terminal nodes in the tree $T_j^*$, $k_h$ the set of all available dimensions to split the node $\eta$ and $\tau_\eta$ the set of all available splitting values given the chosen dimension y for splitting the node $\eta$.

Hence the transition ratio is given by

$$\text{TR} = \frac{P(\text{GROW})\frac{1}{w^*}\frac{1}{\text{card}(k_\eta)}\frac{1}{\text{card}(\tau_\eta)}}{P(\text{PRUNE})\frac{1}{w}}.$$

**Tree Structure Ratio** The proposed tree differs by not having the two children nodes $\eta_L$ and $\eta_R$. Thus the tree structure ratio is:

$$\begin{aligned} \text{TSR} = \frac{P(T_j^*)}{P(T_j^{(t)})} &= \frac{(1 - p_{\text{SPLIT}}(\eta))}{(1 - p_{\text{SPLIT}}(\eta_L))\,(1 - p_{\text{SPLIT}}(\eta_R))\,p_{\text{SPLIT}}(\eta)\,p_{RULE}(\eta)} \\ &= \frac{1 - \frac{\gamma}{(1+d(\eta))^\delta}}{\left(1 - \frac{\gamma}{(1+d(\eta_L))^\delta}\right)\left(1 - \frac{\gamma}{(1+d(\eta_R))^\delta}\right)\frac{\gamma}{(1+d(\eta))^\delta}\frac{1}{\text{card}(k_\eta)}\frac{1}{\text{card}(\tau_\eta)}} \end{aligned}$$

**Likelihood Ratio** Similar to the GROW proposal, the likelihood ratio can be written as follows

$$\begin{aligned} \text{LR} &= \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{-1} \frac{\frac{\Gamma(n_{j\eta}+\alpha)}{(c_{j\eta}+\beta)^{n_{j\eta}+\alpha}}}{\frac{\Gamma(n_{j\eta_L}+\alpha)}{(c_{j\eta_L}+\beta)^{n_{j\eta_L}+\alpha}}\frac{\Gamma(n_{j\eta_R}+\alpha)}{(c_{j\eta_R}+\beta)^{n_{j\eta_R}+\alpha}}} \\ &= \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{-1} \frac{\Gamma(n_{j\eta}+\alpha)}{\Gamma(n_{j\eta_L}+\alpha)\Gamma(n_{j\eta_R}+\alpha)} \frac{(c_{j\eta_L}+\beta)^{n_{j\eta_L}+\alpha}(c_{j\eta_R}+\beta)^{n_{j\eta_R}+\alpha}}{(c_{j\eta}+\beta)^{n_{j\eta}+\alpha}} \end{aligned}$$

### 3.2.2.3 CHANGE Proposal

This proposal randomly picks an internal node and randomly reassigns to it a splitting rule.

Let $\eta$ be the picked internal node having rule $y < c$ and children denoted as $\eta_R$ and $\eta_L$. We assume that $\widetilde{y} < \widetilde{c}$ is its new assigned rule in the proposed tree, $T_j^*$. Following Kapelner and Bleich [2016], for simplicity we are restricted to picking an internal node having two terminal nodes as children.

**Transition Ratio** It holds that:

(i) $q(T_j^*|T_j^{(t)}) =$ P(CHANGE)
× P(selecting an internal node $\eta$ to change)
× P(selecting the new available dimension $\widetilde{y}$ to split on)
× P(selecting the new splitting value $\widetilde{c}$ given the chosen dimension $\widetilde{y}$)

(ii) $q(T_j^{(t)}|T_j^*) =$ P(CHANGE)
× P(selecting the node $\eta$ to change)
× P(selecting the dimension $y$ to split on)
× P(selecting the splitting value c given the chosen dimension $y$)

Thus the Transition Ratio is

$$\mathrm{TR} = \frac{P(\text{selecting } c \text{ to split on given the chosen dimension } y)}{P(\text{selecting } \widetilde{c} \text{ to split on given the chosen dimension } \widetilde{y})}$$

**Tree Structure Ratio** The two trees differ in the splitting rule at node $\eta$. Thus we have that

$$\begin{aligned}\mathrm{TSR} = \frac{P(T_j^*)}{P(T_j^{(t)})} &= \frac{p_{\mathrm{SPLIT}}(\eta)\, p_{\mathrm{RULE}}(\eta|T_j^*)}{p_{\mathrm{SPLIT}}(\eta)\, p_{\mathrm{RULE}}(\eta|T_j^{(t)})} \\ &= \frac{P(\text{selecting } \widetilde{y})\, P(\text{selecting } \widetilde{c} \text{ given } \widetilde{y})}{P(\text{selecting } y)\, P(\text{selecting } c \text{ given } y)} \\ &= \frac{P(\text{selecting } \widetilde{c} \text{ given } \widetilde{y})}{P(\text{selecting } c \text{ given } y)}.\end{aligned}$$

It then follows that $\mathrm{TR} \cdot \mathrm{TSR} = 1$, and hence only the likelihood ratio needs to be found to obtain the Hastings ratio.

**Likelihood Ratio** Let $n_L^* = n_{j\eta_L}^{(T_j^*)}$, $n_R^* = n_{j\eta_R}^{(T_j^*)}$, $c_L^* = c_{j\eta_L}^{(T_j^*)}$, $c_R^* = c_{j\eta_R}^{(T_j^*)}$, $n_L^{(t)} = n_{j\eta_L}^{(T_j^{(t)})}$, $n_R^{(t)} = n_{j\eta_R}^{(T_j^{(t)})}$, $c_L^{(t)} = c_{j\eta_L}^{(T_j^{(t)})}$ and $c_R^{(t)} = c_{j\eta_R}^{(T_j^{(t)})}$, where $(T_j^*)$ and $(T_j^{(t)})$ indicate that the corresponding quantities are related to the tree $T_j^*$ and $T_J^{(t)}$ respectively. Following the previous proposals, the likelihood ratio is

$$\mathrm{LR} = \frac{\frac{\Gamma(n_L^*+\alpha)}{(c_L^*+\beta)^{n_L^*+\alpha}}\,\frac{\Gamma(n_R^*+\alpha)}{(c_R^*+\beta)^{n_R^*+\alpha}}}{\frac{\Gamma(n_L^{(t)}+\alpha)}{(c_L^{(t)}+\beta)^{n_L^{(t)}+\alpha}}\,\frac{\Gamma(n_R^{(t)}+\alpha)}{(c_R^{(t)}+\beta)^{n_R^{(t)}+\alpha}}} = \frac{(c_L^{(t)}+\beta)^{n_L^{(t)}+\alpha}\,(c_R^{(t)}+\beta)^{n_R^{(t)}+\alpha}}{(c_L^*+\beta)^{n_L^*+\alpha}\,(c_R^*+\beta)^{n_R^*+\alpha}}\,\frac{\Gamma(n_L^*+\alpha)\,\Gamma(n_R^*+\alpha)}{\Gamma(n_L^{(t)}+\alpha)\,\Gamma(n_R^{(t)}+\alpha)}.$$

## 3.3 Simulation Study on Synthetic Data

We carried out a simulation study on synthetic data to illustrate the performance of Algorithm 4 to estimate first the intensity of one dimensional and two dimensional inhomogeneous Poisson processes and finally the intensity of multidimensional Poisson processes.

We simulate realizations of Poisson processes on the domain $[0,1)^d$ for $d \in \{1,2,3,4,5\}$ via thinning [Lewis and Shedler, 1979]. The hyperparameters of the model (for the trees and the leaf intensities) are fixed as described in Section 3.2.1. We initially randomly generate $m$ trees of zero depth. The probabilities of the proposals in Algorithm 5 are set to: $P(\text{GROW}) = P(\text{PRUNE}) = 0.4$ and $P(\text{CHANGE}) = 0.2$. A set $\{z_i\}$ is defined by uniformly sampling points in the domain $[0,1)^d$.

We run 3 parallel chains of the same length. We discard their first halves treating the second halves as a sample from the target distribution. We assess chain convergence using the Gelman-Rubin convergence diagnostic [Gelman and Rubin, 1992] applied to the estimated intensity for each point of the set $\{z_i\}$ , as well as trace plots and autocorrelation plots for some points of the testing set.

At each state $t$ of a simulated chain we estimate the intensity for each point $z_i$ by a product of trees denoted as

$$\widehat{\lambda}^{(t)}(z_i) = \prod_{j=1}^{m} g(z_i; T_j^{(t)}, \Lambda_j^{(t)}).$$

The induced sequence $\{\widehat{\lambda}^{(t)}(\cdot)\}_{t=1}^{\infty}$ for the sequence of draws $\{(T_1^{(t)}, \Lambda_1^{(t)}), .., (T_m^{(t)}, \Lambda_m^{(t)})\}_{t=1}^{\infty}$ converges to $P(\widehat{\lambda}|\boldsymbol{s})$. We estimate the posterior mean $E[\widehat{\lambda}(\cdot)|s_1, ..s_n]$, the posterior median of $\widehat{\lambda}(\cdot)$, and the highest density interval (hdi) using the function *hdi* provided by the **R** package **bayestestR** [Makowski et al., 2019]. To assess the performance of our algorithm, we compute the Average Absolute Error (AAE) of the computed estimate:

$$\text{AAE}(\widehat{\lambda}) = \frac{1}{N_z} \sum_{i=1}^{N_z} |\widehat{\lambda}(z_i) - \lambda(z_i)| \tag{3.11}$$

and the Root Integrated Square Error (RISE):

$$\text{RISE}(\widehat{\lambda}) = \left( \frac{1}{N_z} \sum_{i=1}^{N_z} (\widehat{\lambda}(z_i) - \lambda(z_i))^2 \right)^{1/2} \tag{3.12}$$

where $N_z$ is the number of test points.

In the spirit of Akaike information criterion (AIC) [Loader, 1999], we also introduce two diagnostics targetting the likelihood function to evaluate if increasing the number of trees leads to better intensity estimation:

$$D_g = 2\left(\log P(s_1, .., s_n) - k_g\right),$$

and

$$D_l = 2\left(\log P(s_1, .., s_n) - k_l\right),$$

where $k_g$ is the number of global cells, and $k_l$ is the overall number of leaves in the ensemble. We estimate both diagnostics using the sequence of the draws $\left(T^{(w)}, \Lambda^{(w)}\right) = \left\{\left(T_1^{(w)}, \Lambda_1^{(w)}\right), \ldots, \left(T_m^{(w)}, \Lambda_m^{(w)}\right)\right\}_{w=1}^{N_w}$ after the burn-in period as

$$D_g \approx 2\frac{1}{N_w}\sum_w \left(\log P\left(s_1, .., s_n | \left(T^{(w)}, \Lambda^{(w)}\right)\right) - k_g^{(w)}\right),$$

and

$$D_l \approx 2\frac{1}{N_w}\sum_w \left(\log P\left(s_1, .., s_n | \left(T^{(w)}, \Lambda^{(w)}\right)\right) - k_l^{(w)}\right),$$

where $k_g^{(w)}$ and $k_l^{(w)}$ are the number of global cells and the overall number of leaves in the ensemble associated to the $w_{th}$ draw, respectively.

AIC has been shown to be asymptotically equal to leave-one-out cross validation (LOO-CV) [Stone, 1977; Gelman et al., 2014]. According to Leininger and Gelfand [2017], the computational burden required for leave-one-out cross validation considering a point pattern data is impractical. We introduce a leave-partition-out (LPO) method, assuming that the initial process $N(\boldsymbol{s})$ is obtained by combining independent processes $\{N_i(\boldsymbol{s})\}_{i=1}^{N_p}$, as follows

$$D_{LPO} = \sum_{i=1}^{N_P} \log P\left(N_i(\boldsymbol{s}) | N(\boldsymbol{s}) - \{N_i(\boldsymbol{s})\}\right) \tag{3.13}$$

where $P\left(N_i(\boldsymbol{s}) | N(\boldsymbol{s}) - \{N_i(\boldsymbol{s})\}\right)$ is the leave-partition-out predictive intensity given the process $N(\boldsymbol{s})$ without the $i_{th}$ partition, $N_i(\boldsymbol{s})$. We can evaluate 3.13 as follows,

$$D_{LPO} = \sum_{i=1}^{N_P} \log\left(\frac{1}{N_w}\sum_w P\left(N_i(\boldsymbol{s}) | \left(T^{(w,i)}, \Lambda^{(w,i)}\right)\right)\right)$$

where $(T^{(w,i)}, \Lambda^{(w,i)})$ is the sequence of draws $\left\{\left(T_1^{(w,i)}, \Lambda_1^{(w,i)}\right), \ldots, \left(T_m^{(w,i)}, \Lambda_m^{(w,i)}\right)\right\}$ after the burn-in period leaving out the partition $N_i(\boldsymbol{s})$. We assume that each event of $N(\boldsymbol{s})$ is coming from $N_i(\boldsymbol{s})$ with probability $p_i$. The bias of the method is introduced by randomly splitting the process into individual processes. We can get the LOO-CV by LPO, defining appropriately the parameter $N_p$. As higher the number $N_p$ is, as less biased the method is. In the simulation scenarios, we consider that $p_i = 0.1$, $i = 1, \ldots, N_p$ and $N_p = 10$ for computational reasons. The diagnostics show that tiny ensembles of trees provide good estimates in our simulation scenarios.

To confirm the proposed diagnostics, we use $p$-thinning [Illian et al., 2008, Chapter 6] with $p = 0.8$ to create training and test datasets in two of the simulation scenarios. We employ Root Mean Chi-Square statistic (RMCS) and Rank Probability Score (RPS) [Gneiting and Raftery, 2007] with the test data set comparing observed counts in disjoint equal volume subregions $\{S_i\}_{i=1}^{N_s}$ as follows:

$$\mathrm{RMCS}(\widehat{N}) = \left(\frac{1}{N_s}\sum_{i=1}^{N_s} \frac{(\widehat{N}(S_i) - N(S_i))^2}{\widehat{N}(S_i)}\right)^{1/2} \tag{3.14}$$

and

$$\mathrm{RPS}(N(S_j)) = \sum_{u=0}^{N(S_j)-1} F(u)^2 + \sum_{u=N(S_j)}^{\infty} (F(u) - 1)^2 \,, \tag{3.15}$$

where $F$ is the Poisson cumulative distribution with parameter $m = \int_{S_j} \hat{\lambda}(s)ds$, $N(S_i)$ the actual number of testing points in $S_i$ and $\widehat{N}(S_i)$ the estimated number of testing points in $S_i$ given by

$$\widehat{N}(S_i) = \int\limits_{S_i} \frac{1-p}{p}\widehat{\lambda}(s)ds \simeq \frac{1-p}{p}\frac{1}{N_z^i}\sum_{z_j \in S_i} \widehat{\lambda}(z_j)|S_i| \tag{3.16}$$

with $N_z^i$ being the number of points $\{z_j\}$ falling in $S_i$ and estimating the intensity at each points $s$ , $\widehat{\lambda}(s)$, via the posterior mean $E[\widehat{\lambda}(\cdot)|s_1, ..s_n]$.

For one dimensional processes, we compare the results of Algorithm 4 to the Haar-Fisz algorithm [Fryzlewicz and Nason, 2004], a wavelet based method for estimating the intensity of one dimensional Poisson processes that outperforms well known competitors. We apply the Haar-Fisz algorithm to the counts of points falling into 256 consecutive intervals using the **R** package **haarfisz** [Fryzlewicz, 2010]. Our algorithm is competitive with the Haar algorithm for smooth intensity functions and is not strongly out-performed by the Haar-Fisz algorithm when the underlying intensity is a stepwise function.

For two-dimensional processes, we compare the results of our algorithm with fixed-bandwidth estimators and log-Gaussian Cox process (LGCP) with intensity $\lambda(\boldsymbol{s}) = \exp(a + u(\boldsymbol{s}))$ where $u$ is a Gaussian process with exponential covariance function. We used a discretization version of the LGCP model defined on a regular grid over space, which we implemented using Stan-code [Gelman et al., 2015]. As noted in Davies and Baddeley [2018], the choice of the kernel is not of primary importance, we choose a Gaussian kernel for its wide applicability. In our tables of results, the smoothing bandwidth, sigma, selected using likelihood cross-validation [Loader, 1999] denoted by (LCV), and we have also included other values of sigma to demonstrate the sensitivity to bandwidth choice. The kernel estimators, and the bandwidth value given by likelihood cross-validation, were computed using the **R** package **spatstat** [Baddeley and Turner, 2005]. Our algorithm outperforms the maximum likelihood approach using linear conditional intensity, as expected. Our algorithm outperforms kernel smoothing and LGCP for stepwise functions and is competitive with them for a smooth intensity.

Finally, we examine the performance of our algorithm for multidimensional intensities by generating realizations of Poisson processes on the domain $[0, 1)^d$ for $d \in \{3, 5\}$ via thinning. Future work includes the study of intensities in higher dimensions ($d > 5$). We compare our intensity estimates with kernel smoothing estimators having isotropic standard deviation matrices with diagonal elements equal to $h$ and the methodology for applying maximum likelihood to point process models with linear conditional intensity [Peng, 2003]. We select the bandwidth $h$ using likelihood cross-validation [Loader, 1999] denoted by (LCV).

### 3.3.1 One dimensional Poisson process with stepwise intensity

Our first example is a one dimensional Poisson process with piecewise constant intensity with several steps (Fig. 3.1). We run 3 parallel chains of the same length for 200000 iterations for 2-10 trees, 100000 for 12 trees, 50000 iterations for 15 trees and 30000 iterations for 20 trees. The convergence criteria indicate convergence of the simulated chains for the majority of points (see Appendix A.5).

Our algorithm detects the change points and provides good estimates of the intensity and is competitive in terms of AAE with the Haar-Fisz algorithm, but does not perform as well in terms of RISE (see Fig. 3.1 and Tables 3.3-3.6). We have found the metrics and convergence diagnostics in a set of uniformly chosen points without excluding the points close to jumps. Due to inferring the intensity via a product of stepwise functions, it is expected that the proposed algorithm will provide estimates with higher variability close to jumps. The proposed algorithm outperforms the Haar-Fisz algorithm without considering the points close to jumps. Tables 3.4-3.5 show the metrics for various number of trees without considering the points in a distance ±0.02 from the jumps.

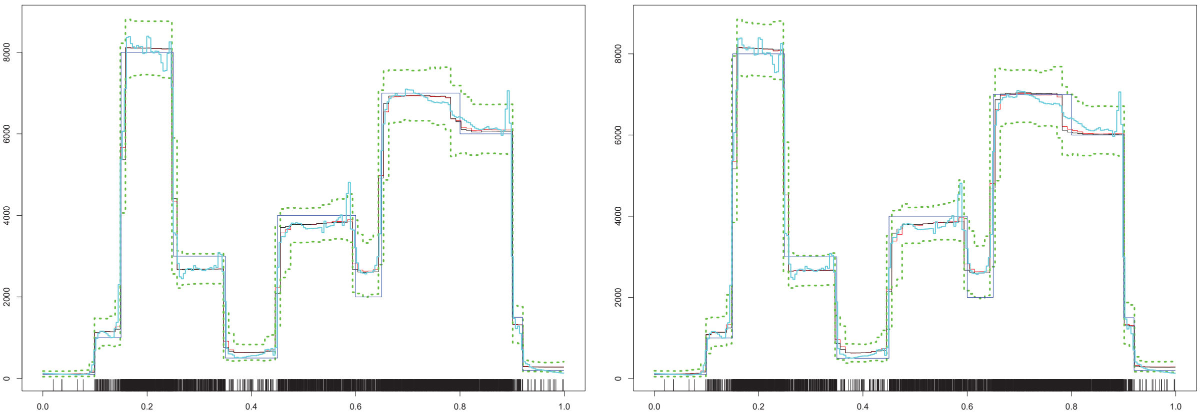
The diagnostics $D_g$, $D_l$ and $D_{LPO}$ obtain their highest values for 7, 4 and 8 trees, respectively. The analysis demonstrates only small differences between log-likelihood values as the number of trees increases, supporting results found in previous BART studies that the method is robust to the choice of $m$. The average RMCS and RPS on testing points over 7 different splits of the original data set (Tables 3.1-3.2) provide evidence that ensembles with more than seven trees do not improve the fit of the proposed algorithm.

| Proposed BART Algorithm | | | | | | |
|---|---|---|---|---|---|---|
| Number of trees | $N_s = 1$ | $N_s = 10$ | $N_s = 25$ | $N_s = 50$ | $N_s = 75$ | $N_s = 75$ |
| 3 | 17.12 | 5.25 | 3.11 | 2.08 | 1.65 | 1.43 |
| 4 | 16.98 | 5.28 | 3.09 | 2.07 | 1.64 | 1.42 |
| 5 | 16.94 | 5.26 | 3.04 | 2.06 | 1.63 | 1.41 |
| 7 | 17.01 | 5.22 | 3.00 | 2.04 | 1.62 | 1.40 |
| 8 | 17.09 | 5.20 | 2.99 | 2.03 | 1.62 | 1.40 |
| 9 | 17.07 | 5.20 | 2.98 | 2.02 | 1.61 | 1.39 |
| 10 | 20.10 | 5.78 | 3.12 | 2.12 | 1.63 | 1.43 |
| 12 | 17.74 | 5.37 | 2.99 | 2.06 | 1.63 | 1.39 |
| 15 | 17.03 | 5.15 | 2.94 | 2.01 | 1.60 | 1.38 |
| 20 | 17.08 | 5.16 | 2.93 | 2.00 | 1.60 | 1.38 |

Table 3.1: The average RPS on testing points over 7 different splits of the original data set in Fig. 3.1.

| Proposed BART Algorithm | | | | | | |
|---|---|---|---|---|---|---|
| Number of trees | $N_s = 1$ | $N_s = 10$ | $N_s = 25$ | $N_s = 50$ | $N_s = 75$ | $N_s = 75$ |
| 2 | 0.95 | 1.13 | 1.06 | 1.02 | 0.99 | 1.00 |
| 3 | 0.95 | 1.13 | 1.06 | 1.02 | 0.98 | 0.99 |
| 4 | 0.95 | 1.14 | 1.06 | 1.02 | 0.98 | 0.98 |
| 5 | 0.94 | 1.13 | 1.04 | 1.02 | 0.98 | 0.98 |
| 7 | 0.95 | 1.13 | 1.04 | 1.01 | 0.97 | 0.97 |
| 8 | 0.95 | 1.12 | 1.03 | 1.01 | 0.97 | 0.96 |
| 9 | 0.95 | 1.13 | 1.03 | 1.01 | 0.97 | 0.97 |
| 10 | 1.10 | 1.20 | 1.06 | 1.04 | 0.98 | 0.98 |
| 12 | 0.98 | 1.18 | 1.04 | 1.02 | 0.98 | 0.96 |
| 15 | 0.95 | 1.12 | 1.02 | 1.00 | 0.97 | 0.96 |
| 20 | 0.95 | 1.12 | 1.02 | 1.00 | 0.97 | 0.96 |

Table 3.2: The average RMCS on testing points over 7 different splits of the original data set in Fig. 3.1.



(a) 5 Trees

(b) 7 Trees

Figure 3.1: The original intensity (blue curve), the posterior mean (red curve), the posterior median (black curve), the 95% hdi interval of the estimated intensity illustrated by the dotted green lines and the Haar-Fisz estimator (cyan curve). The rug plot on the bottom displays the 3590 event times.

| Proposed BART Algorithm | | | | | | | |
|---|---|---|---|---|---|---|---|
| Number of trees | AAE for Posterior Mean | AAE for Posterior Median | RISE for Posterior Mean | RISE for Posterior Median | $D_g$ | $D_l$ | $D_{LPO}$ |
| 3 | 308.87 | 320.84 | 603.54 | 633.48 | 54095.1 | 54090 | -339.7 |
| 4 | 287.89 | 283.03 | 580.69 | 587.08 | 54096.5 | 54090 | -368 |
| 5 | 289.27 | 281.13 | 580.55 | 586.24 | 54098 | 54088.4 | -352.5 |
| 7 | 281.59 | 274.88 | 588.7 | 592.11 | 54098 | 54082.7 | -263.5 |
| 8 | 280.62 | 274.07 | 588.73 | 591.29 | 54097.9 | 54079.5 | -261.5 |
| 9 | 282.78 | 276.99 | 593.93 | 595.23 | 54096.9 | 54075.2 | -327.9 |
| 10 | 283.79 | 279.07 | 593.95 | 595.41 | 54095.6 | 54071.6 | -322.6 |
| 20 | 297.21 | 287.86 | 599.77 | 595.04 | 54082.9 | 54029.7 | -436 |

Table 3.3: Average Absolute Error and Root Integrated Square Error for various number of trees for the data in Fig. 3.1.

| Proposed BART Algorithm | | | | |
|---|---|---|---|---|
| Number of trees | AAE for Posterior Mean | AAE for Posterior Median | RISE for Posterior Mean | RISE for Posterior Median |
| 4 | 144.48 | 139.58 | 181.21 | 174.82 |
| 5 | 144.55 | 139.02 | 180.74 | 176.19 |
| 7 | 124.53 | 123.2 | 175.74 | 172.4 |

Table 3.4: Average Absolute Error and Root Integrated Square Error for the data in Fig. 3.1 without considering points close to steps.

| Haar-Fisz Algorithm | |
|---|---|
| AAE | RISE |
| 141.95 | 192.6 |

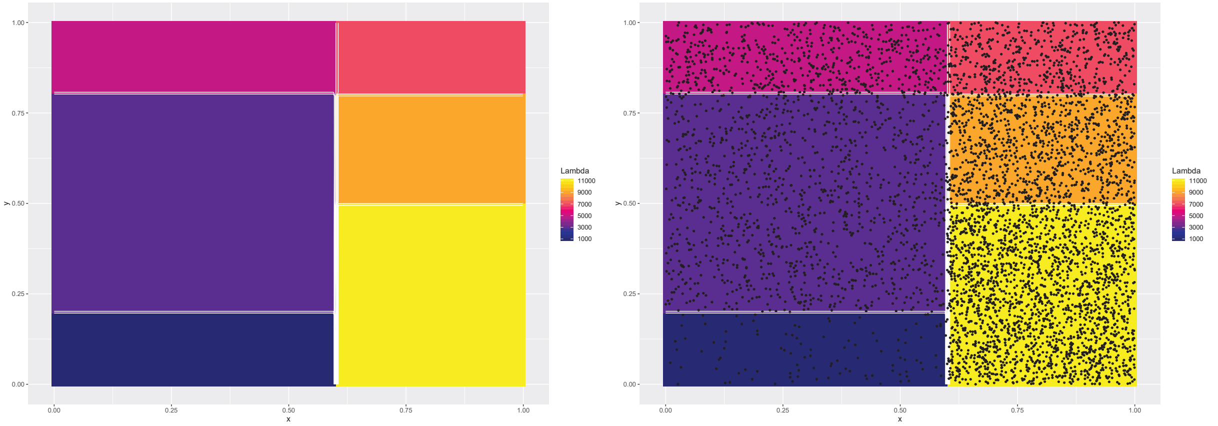
Table 3.5: Average Absolute Error and Root Integrated Square Error for Haar-Fisz estimator for the data in Fig. 3.1 without considering points close to steps.

| Haar-Fisz Algorithm | |
|---|---|
| AAE | RISE |
| 272.3 | 476.9 |

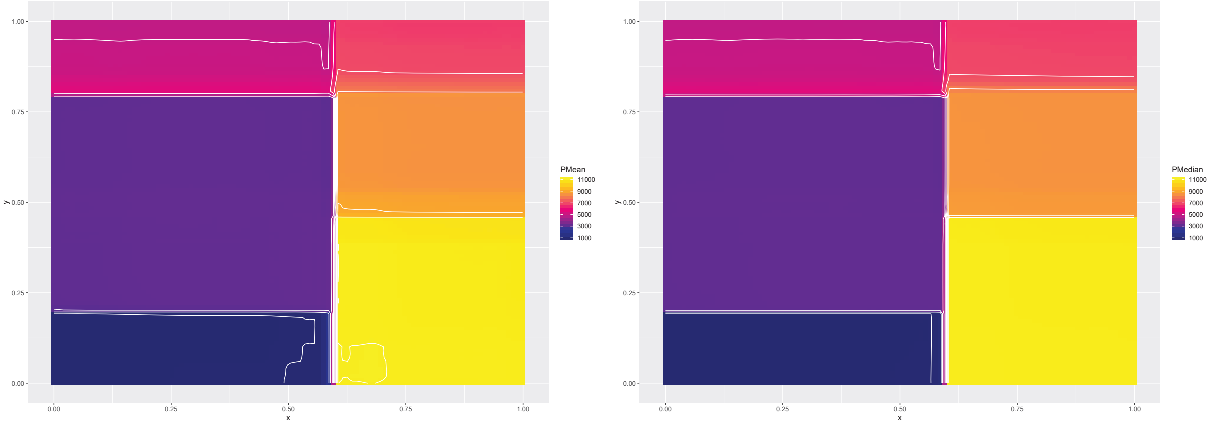
Table 3.6: Average Absolute Error and Root Integrated Square Error for Haar-Fisz estimator for the data in Fig. 3.1.

### 3.3.2 Two-dimensional Poisson process with stepwise intensity function

To demonstrate the applicability of our algorithm in a two-dimensional setting, Figures 3.2-3.3 and Tables 3.7-3.9 reveal that our algorithm outperforms kernel smoothing and inference with spatial log-Gaussian Cox processes for stepwise intensity functions. We run 3 parallel chains of the same length for 100000 iterations for 3-6 trees. The convergence criteria indicate convergence of the simulated chains for the majority of points. As may be expected, the simulation study shows that points close to jumps are estimated with less reliability. The algorithm converges less well at these points, as demonstrated by the Gelman-Rubin diagnostic (see Appendix A.5). The diagnostics $D_g$, $D_l$ and $D_{LPO}$ obtain their highest values for three trees, respectively. The diagnostics indicate that small ensembles of trees can provide a good estimate of the intensity.

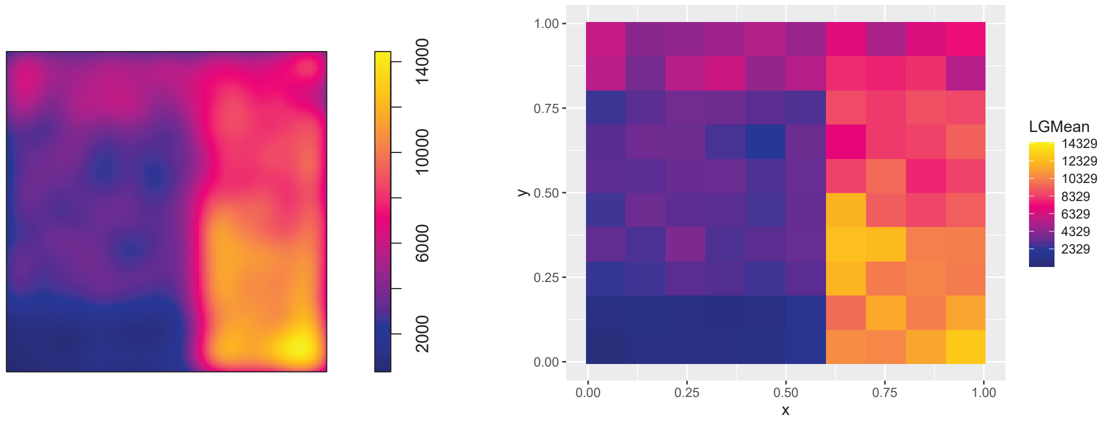


(a) Original Intensity (b) Realization of Process consisting of 5579 points



(c) Posterior Mean (d) Posterior Median

Figure 3.2: Original Intensity, posterior mean and posterior median for 4 trees.



(a) Kernel Smoothing with $h = 0.039$.

(b) LGCP

Figure 3.3: Kernel estimator and inference with spatial log-Gaussian Cox processes.

| Proposed BART Algorithm | | | | | | | |
|---|---|---|---|---|---|---|---|
| Number of trees | AAE for Posterior Mean | AAE for Posterior Median | RISE for Posterior Mean | RISE for Posterior Median | $D_g$ | $D_l$ | $D_{LPO}$ |
| 3 | 224.1 | 230.3 | 419.2 | 453.7 | 87227.2 | 87232.3 | 505.1 |
| 4 | 208.7 | 213 | 410.2 | 447.9 | 87223.7 | 87230.5 | 491.2 |
| 5 | 216.8 | 212.9 | 389.5 | 410.9 | 87211.6 | 87220.6 | 406 |
| 6 | 228.9 | 221.9 | 395.8 | 412.8 | 87197.5 | 87214.7 | 463.9 |

Table 3.7: Average Absolute Error, Root Integrated Square Error and diagnostics for various trees for the data in Figure 3.2.

| Kernel Smoothing | | |
|---|---|---|
| Bandwidth (sigma) | AAE | RISE |
| 0.027 | 763.8 | 1041.3 |
| 0.038 | 662.7 | 956.8 |
| 0.039 (LCV) | 655.9 | 954.7 |
| 0.047 | 636.7 | 960.6 |
| 0.067 | 672.8 | 1042.5 |

Table 3.8: Average Absolute Error and Root Integrated Square Error for fixed bandwidth estimators for the data in Figure 3.2.

| Inference with spatial log-Gaussian Cox processes | | |
|---|---|---|
| grid | AAE | RISE |
| $10 \times 10$ | 568.4 | 750.9 |
| $20 \times 20$ | 678 | 953 |

Table 3.9: Average Absolute Error and Root Integrated Square Error with LGCP for the data in Figure 3.2.

### 3.3.3 Inhomogeneous three-dimensional Poisson process with Gaussian intensity

Our first example for multidimensional intensities is a three-dimensional Poisson process with intensity $\lambda(x) = 500e^{\mathrm{x}^T\mathrm{x}}$ for x $\in [0,1)^3$. We generated a realization of 1616 points via thinning. We run 3 parallel chains of the same length for 100000 iterations for 3-10 trees and 30000 iterations for 12 trees. Tables 3.10 and 3.11 illustrate the statistics of our algorithm and kernel smoothing. Figures 3.4 and 3.5 show our estimators and the kernel estimator with $h$=0.073 for 8 Trees and 10 Trees with fixed third dimension (x[3]) at 0.4 and 0.8, respectively.

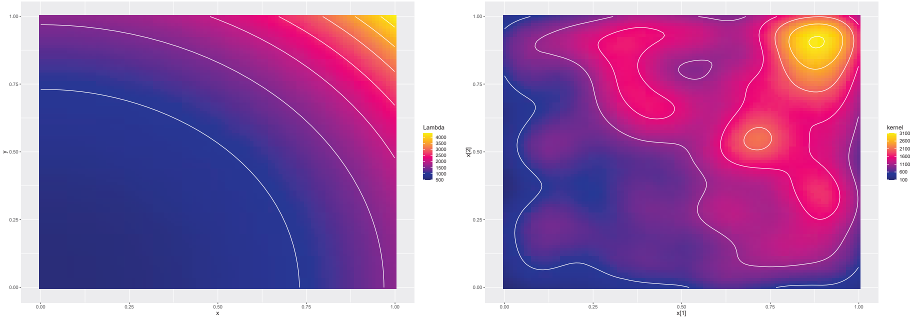
The AIC diagnostics $D_g$, $D_l$ and $D_{LPO}$ get their highest values with 4 trees, respectively. We observe that the diagnostic $D_l$ slightly differs between 4 and 8 trees. The diagnostic $D_g$ is similar between 4 and 5 trees. The estimate of the average logarithm of Poisson process likelihood does not change significantly from 4 trees to 12 trees. Specifically, we observe its maximum equal to 10536.3 at 12 trees, while its minimum to 10531.9 at 4 trees. In addition, the estimated average number of leaves in a tree of an ensemble is about 3 for $4-12$ trees. That explains why we observe higher values of diagnostics for a small number of trees. The metrics $AAE$ and $RISE$ are optimised with 12 trees. However, it should be noted that only small variations in the metrics are seen between 4 and 12 trees. The diagnostics provide evidence that increasing the number of trees does not improve the fit of the proposed model.

| Proposed BART Algorithm | | | | | | | |
|---|---|---|---|---|---|---|---|
| Number of trees | AAE for Mean | AAE for Median | RISE for Mean | RISE for Median | $D_g$ | $D_l$ | $D_{LPO}$ |
| 4 | 247.6 | 254.9 | 360.7 | 376.3 | 20993.7 | 21040.6 | -1409 |
| 5 | 250.2 | 258.3 | 364.3 | 380.4 | 20992.1 | 21039.8 | -1492 |
| 6 | 247.6 | 254.7 | 360.8 | 375.4 | 20979.5 | 21038.6 | -1529 |
| 8 | 234.8 | 239.4 | 341 | 352.4 | 20938.6 | 21032.8 | -1515 |
| 10 | 226.8 | 229.4 | 330.5 | 338.5 | 20883 | 21026.9 | -1539 |
| 12 | 221.6 | 222.3 | 320.6 | 326.4 | 20810.4 | 21020.7 | -1609 |

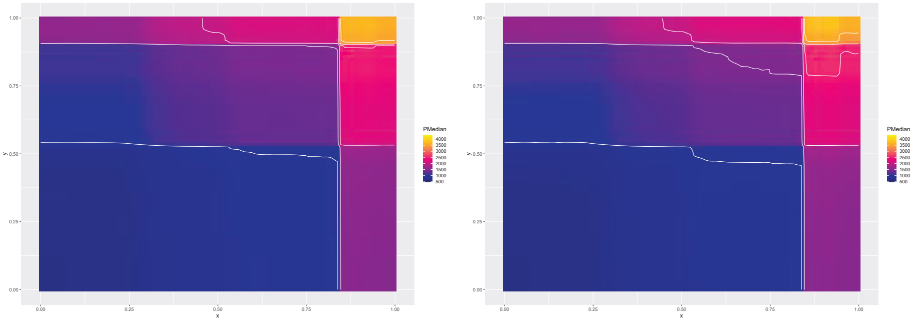
Table 3.10: Average Absolute Error, Root Integrated Square Error and diagnostics for various number of trees.

| Kernel Smoothing | | |
|---|---|---|
| $h$ | AAE | RISE |
| 0.02 | 1483.3 | 2105.1 |
| 0.053 | 480.8 | 667.5 |
| 0.073 (LCV) | 415.86 | 645.16 |
| 0.08 | 417.7 | 661.4 |
| 0.085 | 423.2 | 676.2 |
| 0.1 | 450.3 | 727.6 |
| 0.3 | 890.4 | 1236 |

Table 3.11: Average Absolute Error and Root Integrated Square Error for various isotropic variance matrices.

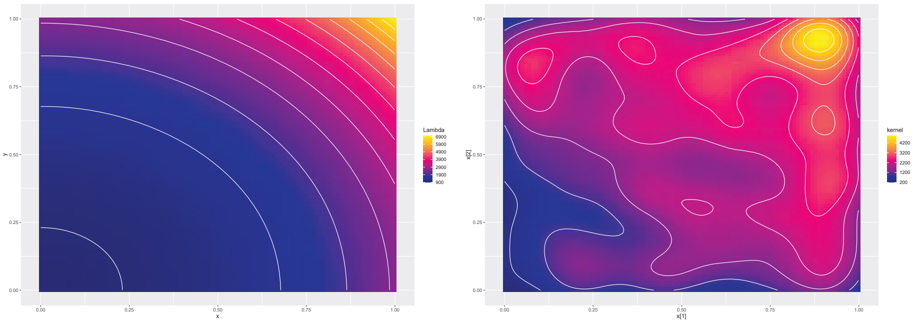


(a) Original Intensity (b) Kernel Estimator with $h$=0.073

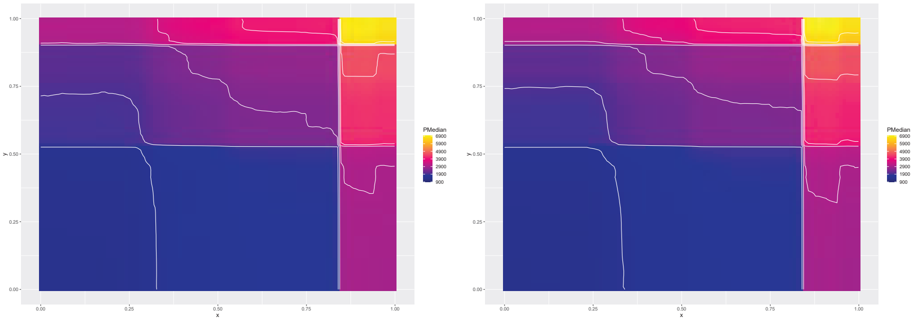


(c) Posterior Median for 8 Trees (d) Posterior Median for 10 Trees

Figure 3.4: Kernel estimator and Posterior Median for 8 and 10 Trees with x[3] = 0.4.



(a) Original Intensity (b) Kernel Estimator with $h$=0.073



(c) Posterior Median for 8 Trees (d) Posterior Median for 10 Trees

Figure 3.5: Kernel Estimator and Posterior Median for 8 and 10 Trees with x[3] = 0.8.

### 3.3.4 Inhomogeneous five dimensional Poisson process with sparsity assumption

Here, we demonstrate the performance of our algorithm to detect the dimensions that contribute most in the intensity of $s \in \mathcal{S}$ in a noisy environment. Consider a five dimensional inhomogeneous Poisson process with intensity function of $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5) \in [0, 1)^5$ depending on 3 of 5 dimensions:

$$\begin{aligned}\lambda(\mathbf{x}) = {} & (2\mathbb{1}\,(x_1 < 0.2) + 10\mathbb{1}\,(x_1 \geq 0.2)) * (3\mathbb{1}\,(x_2 < 0.5) + 15\mathbb{1}\,(x_2 \geq 0.5)) \\ & * (3\mathbb{1}\,(x_3 < 0.8) + 30\mathbb{1}\,(x_3 \geq 0.8))\end{aligned}$$

We generate a realization of 669 points via thinning. We run 3 parallel chains of the same length for 100000 iterations for 4-8 trees, 50000 iterations for 10 trees, 30000 iterations for 12 trees and 10000 iterations for 15 trees. The convergence criterion is smaller than 1.1 for the majority of testing points over all numbers of trees considered.

Table 3.12 shows the metrics and diagnostics $D_g$ and $D_l$ of the estimated intensity over various numbers of trees. The diagnostics $D_g$ and $D_l$ obtain their highest values with 4 trees, and the diagnostic $D_l$ shows only small differences between 4 and 5 trees. We note that (i) the average number of leaves in a tree of the ensemble is about 2.2 for 4-5 trees, and (ii) the estimated logarithm of Poisson process likelihood for 4 and 5 trees are 4271.5 and 4271.8, respectively. The diagnostic $D_{LPO}$ gets its highest value with 5 trees. The $p$-thinning approach confirms the diagnostics, and indicates that increasing the number of trees does not improve the fit of the proposed model to the data.

Table 3.16 demonstrates the frequency of times we meet each dimension in the decision rules of a tree. Table 3.15 shows how likely each dimension is to be involved in the root's decision rule. The results illustrate that the important covariates $x_1$, $x_2$ and $x_3$ are more likely to be involved in the decision rules of a tree than the noisy dimensions $x_4$ and $x_5$. That indicates the algorithm prioritizes the dimensions that contribute most to the intensity. Figure 3.6 shows that the mean of the posterior marginal intensities are similar to the expected marginal intensities given that $\{x_i\}_{i=1}^{5}$ are uniform independent covariates.

Tables 3.12, 3.13 and 3.14 show that our algorithm outperforms kernel smoothing and the maximum likelihood approach considering linear conditional intensity as expected. The ability of our method to identify important features demonstrates an important advantage over other procedures.

| Proposed BART Algorithm | | | | | | | |
|---|---|---|---|---|---|---|---|
| Number of trees | AAE for Mean | AAE for Median | RISE for Mean | RISE for Median | $D_g$ | $D_l$ | $D_{LPO}$ |
| 4 | 48.36 | 45.47 | 159.95 | 170.35 | 8510.1 | 8525.4 | -485.9 |
| 5 | 49.18 | 44.54 | 158.82 | 169.07 | 8486.1 | 8520.9 | -467.1 |
| 6 | 50.59 | 45.05 | 161.36 | 170.61 | 8462.1 | 8519 | -477.4 |
| 8 | 56.06 | 47.94 | 162.56 | 164.46 | 8349 | 8511.8 | -490.8 |
| 10 | 61.55 | 52.23 | 169.72 | 166.62 | 8141.8 | 8505.5 | -503.1 |
| 12 | 67.01 | 57.06 | 180.53 | 175.23 | 7774.7 | 8499.8 | -522.2 |
| 15 | 75 | 65.06 | 192.88 | 181.03 | 6813.1 | 8490 | -500.2 |

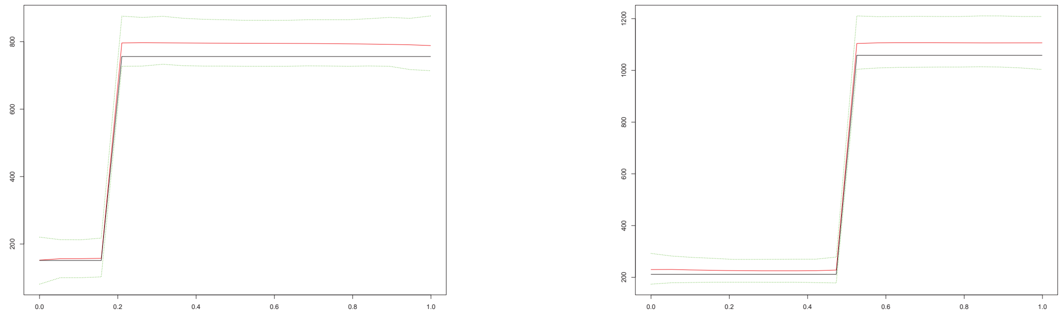
Table 3.12: Average Absolute Error, Root Integrated Square Error and diagnostics for various number of trees in the case of Inhomogeneous five dimensional Poisson process with sparsity assumption.

| Kernel Smoothing | | |
|---|---|---|
| Bandwidth (sigma) | AAE | RISE |
| 0.121 (LCV) | 407.1 | 888.1 |

Table 3.13: Average Absolute Error and Root Integrated Square Error for fixed bandwidth estimators in the case of Inhomogeneous five dimensional Poisson process with sparsity assumption.

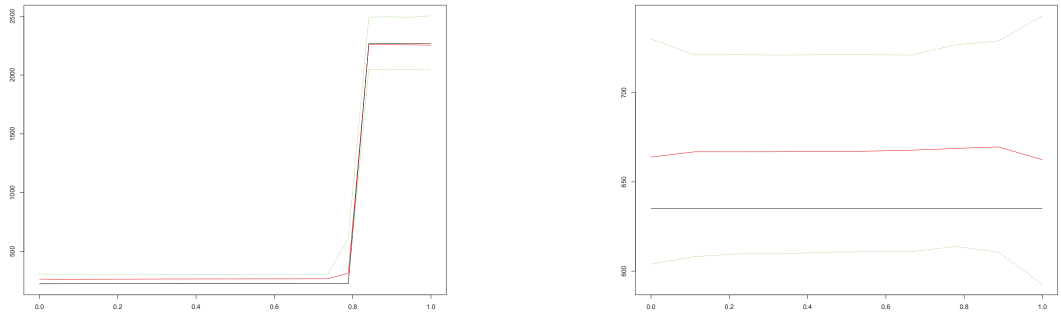
| Linear conditional intensity | |
|---|---|
| AAE | RISE |
| 654.2 | 1076.5 |

Table 3.14: Average Absolute Error and Root Integrated Square Error for linear conditional intensity in the case of Inhomogeneous five dimensional Poisson process with sparsity assumption.



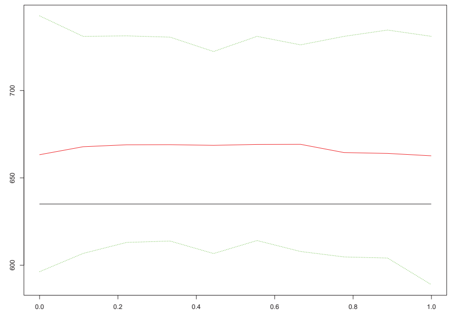
(a) The posterior mean of $\lambda(x_1)$ (red line; 95% CI (green line)) and the true expected value of $\lambda(x_1)$ (black line).

(b) The posterior mean of $\lambda(x_2)$ (red line; 95% CI (green line)) and the true expected value of $\lambda(x_2)$ (black line).



(c) The posterior mean of $\lambda(x_3)$ (red line; 95% CI (green line)) and the true expected value of $\lambda(x_3)$ (black line).

(d) The posterior mean of $\lambda(x_4)$ (red line; 95% CI (green line)) and the true expected value of $\lambda(x_4)$ (black line).



(e) The posterior mean of $\lambda(x_5)$ (red line; 95% CI (green line)) and the true expected value of $\lambda(x_5)$ (black line).

Figure 3.6: Posterior marginal intensities considering 4 trees.

| Proposed BART Algorithm | | | | | |
|---|---|---|---|---|---|
| Number of trees | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
| 4 | 0.31 | 0.29 | 0.34 | 0.03 | 0.03 |
| 5 | 0.35 | 0.29 | 0.26 | 0.05 | 0.06 |

Table 3.15: How likely each dimension is to be involved in the root's decision rule.

| Proposed BART Algorithm | | | | | |
|---|---|---|---|---|---|
| Number of trees | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
| 4 | 0.35 | 0.36 | 0.37 | 0.06 | 0.07 |
| 5 | 0.39 | 0.34 | 0.37 | 0.09 | 0.10 |

Table 3.16: The frequency of times we meet each dimension in the decision rules of a tree.

| Proposed BART Algorithm | | | |
|---|---|---|---|
| Number of trees | $N_s = 1$ | $N_s = 32$ | $N_s = 243$ |
| 4 | 5.40 | 0.99 | 0.71 |
| 5 | 5.40 | 1 | 0.71 |
| 6 | 5.42 | 1 | 0.71 |
| 8 | 5.42 | 1 | 0.71 |
| 10 | 5.42 | 1 | 0.71 |
| 15 | 5.43 | 1 | 0.71 |

Table 3.17: The average RPS on testing points over 7 different splits of the original data set in the case of Inhomogeneous five dimensional Poisson process with sparsity assumption.

| Proposed BART Algorithm | | | |
|---|---|---|---|
| Number of trees | $N_s = 1$ | $N_s = 32$ | $N_s = 243$ |
| 4 | 0.64 | 0.95 | 1 |
| 5 | 0.64 | 0.95 | 1 |
| 6 | 0.65 | 0.95 | 1 |
| 8 | 0.65 | 0.96 | 1.01 |
| 10 | 0.65 | 0.96 | 1.01 |
| 15 | 0.65 | 0.96 | 1.01 |

Table 3.18: The average RMCS on testing points over 7 different splits of the original data set in the case of Inhomogeneous five dimensional Poisson process with sparsity assumption.

## 3.4 Intensity estimation for Real Data

In this section, we first apply our algorithm to real data sets when modelled as realizations of inhomogeneous Poisson processes in one and two dimensions. To assess the performance of our algorithm, we break the domain $[0, 1)^d$ into equal volume subareas $\{S_i\}_{i=1}^{N_S}$ and consider a set $\{z_i\}$ by uniformly sampling points in the domain $[0, 1)^d$. We compute the AAE of the estimated expected number of points falling into each of the subareas :

$$\mathrm{AAE}(\widehat{N}) = \frac{1}{N_S} \sum_{i=1}^{N_S} |\widehat{N}(S_i) - N(S_i)| \tag{3.17}$$

and Root Integrated Square Error (RISE):

$$\mathrm{RISE}(\widehat{N}) = \left( \frac{1}{N_S} \sum_{i=1}^{N_S} (\widehat{N}(S_i) - N(S_i))^2 \right)^{1/2}, \tag{3.18}$$

where $N(S_i)$ is the actual number of points in $S_i$ and

$$\widehat{N}(S_i) = \int_{S_i} \widehat{\lambda}(s) ds \simeq \frac{1}{N_{S_i}} \sum_{z_j \in S_i} |S_i| \widehat{\lambda}(z_j) \tag{3.19}$$

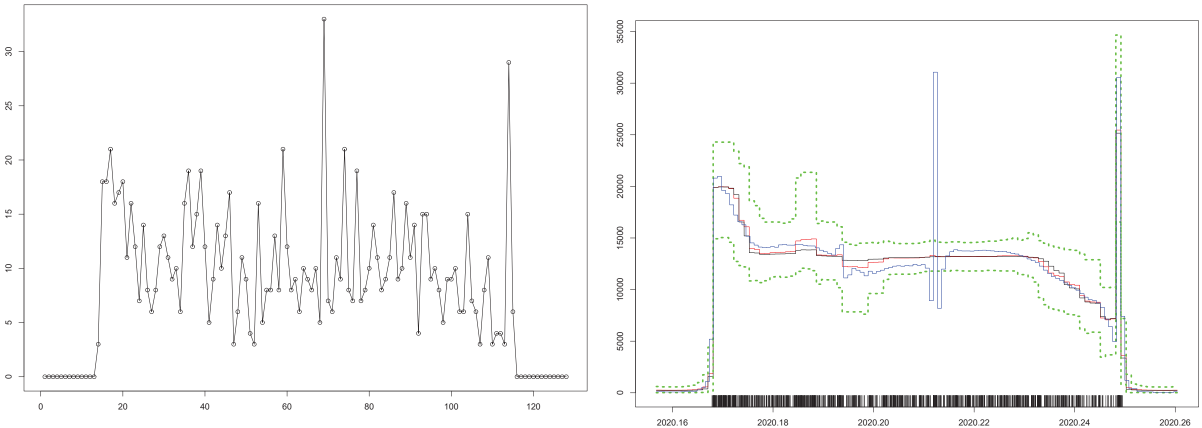
with $N_{S_i}$ being the number of testing points $\{z_j\}$ falling in $S_i$. We apply the metrics AAE and RISE to compare our intensity estimates of one dimensional processes with those obtained by applying the Haar-Fisz algorithm for one dimensional data; and with kernel estimators for two-dimensional data. We observe that our algorithm, the Haar-Fisz algorithm and the kernel smoothing lead to similar results. As expected, the reconstructions of the intensity function are less smooth than those derived with kernel smoothing. The kernel estimator, as well as the bandwidth value given by likelihood cross-validation were computed using the **R** package **spatstat** [Baddeley and Turner, 2005]. We provide more simulation results in Appendix A.3.

### 3.4.1 Earthquakes Data

This data set is available online from the Earthquake Hazards Program [Murray and Svarc, 2017] and consists of the times of 1088 earthquakes from 2-3-2020 to 1-4-2020. We consider the period from 27-2-2020 to 5-4-2020 to avoid edges. We run 3 parallel chains of the same length for 100000 iterations for 3-10 trees. The convergence criteria included in the supplementary material indicate that the considered chains have converged.

Figure 3.7 presents the Posterior Mean and the Posterior Median for 5 Trees, as well as the intensity estimate of the Haar-Fisz algorithm applied to the counts in 128 consecutive intervals of equal length. The deterministic discretized intensity of the **R** package **haarfisz** is divided by the duration of an interval. The differences between both algorithms are due to different assumptions; the Haar-Fisz algorithm considers the aggregated counts into disjoint subintervals of the domain, while the proposed algorithm the times of individual events. The most noticeable difference is observed between 2020.212 and 2020.213 (69th interval) where we see a jump in earthquakes from 5 to 33 and again to 7. The Haar-Fisz algorithm detects that peak as we feed it with that information, while the proposed algorithm does not indicate a sharp rise in the intensity in that period, treating it as an outlier. The intensity estimate of the Haar-Fisz algorithm applied in 64 consecutive intervals is closer to the proposed algorithm (see Figure 3.8), as expected. Similar to coarser binning, the proposed algorithm is less prone to overfitting to spikes in the data, which get filtered out. The estimated AAE and RISE demonstrate good performance compared to the Haar-Fisz method. The simulation results illustrate that our algorithm can track the varying intensity of earthquakes.

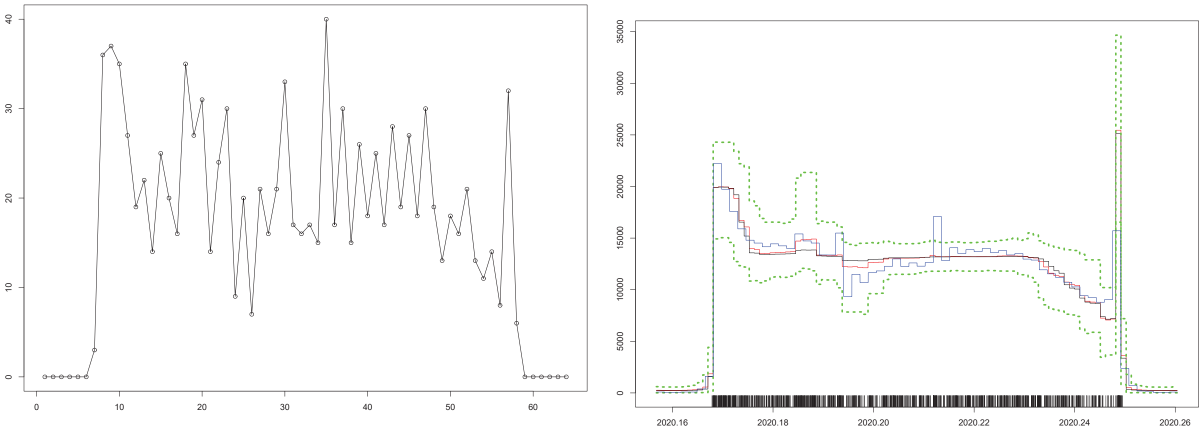
The diagnostics $D_g$, $D_l$ and $D_{LPO}$ obtain their highest values at 9, 3 and 8 trees, respectively. The AIC diagnostics values between 3 and 9 trees show only small variations, we choose 5 trees for the analysis, noting that the results will not vary significantly for other choices of $m$ in this region.



(a) Aggregated Earthquakes in 128 consecutive intervals

(b) 5 Trees and Haar-Fisz Algorithm

Figure 3.7: Earthquakes Data: The posterior mean (red curve), the posterior median (black curve), the 95% hdi interval of the estimated intensity illustrated by the dotted green lines and the intensity estimator of the Haar-Fisz Algorithm illustrated by the blue line. The rug plot on the bottom displays the event times.



(a) Aggregated Earthquakes in 64 consecutive intervals

(b) 5 Trees and Haar-Fisz Algorithm

Figure 3.8: Earthquakes Data: The posterior mean (red curve), the posterior median (black curve), the 95% hdi interval of the estimated intensity illustrated by the dotted green lines and the intensity estimator of the Haar-Fisz Algorithm illustrated by the blue line. The rug plot on the bottom displays the event times.

| Proposed BART Algorithm | | | | | | | |
|---|---|---|---|---|---|---|---|
| Number of trees | AAE for Posterior Mean | AAE for Posterior Median | RISE for Posterior Mean | RISE for Posterior Median | $D_g$ | $D_l$ | $D_{LPO}$ |
| 3 | 93.8 | 94.1 | 106.9 | 107.1 | 13570.1 | 13565.4 | -1194.7 |
| 4 | 94 | 94.1 | 106.8 | 107 | 13570.6 | 13563.6 | -1163.8 |
| 5 | 93.6 | 94 | 106.7 | 107 | 13570.1 | 13560.7 | -1150.5 |
| 6 | 93.8 | 94 | 106.9 | 107 | 13571.6 | 13559.6 | -1169.7 |
| 8 | 93.5 | 94 | 106.6 | 107.1 | 13571.8 | 13554.9 | -1140.2 |
| 9 | 93.4 | 94 | 106.7 | 107.3 | 13572.1 | 13552.6 | -1192.5 |
| 10 | 93.4 | 94 | 106.8 | 107.3 | 13571.9 | 13549.8 | -1184.4 |

Table 3.19: Average Absolute Error, Root Integrated Square Error and diagnostics for the data in Fig. 3.7.

| Haar-Fisz Algorithm | | |
|---|---|---|
| Subintervals | AAE | RMSE |
| 128 | 94.1 | 107.8 |
| 64 | 94 | 107 |

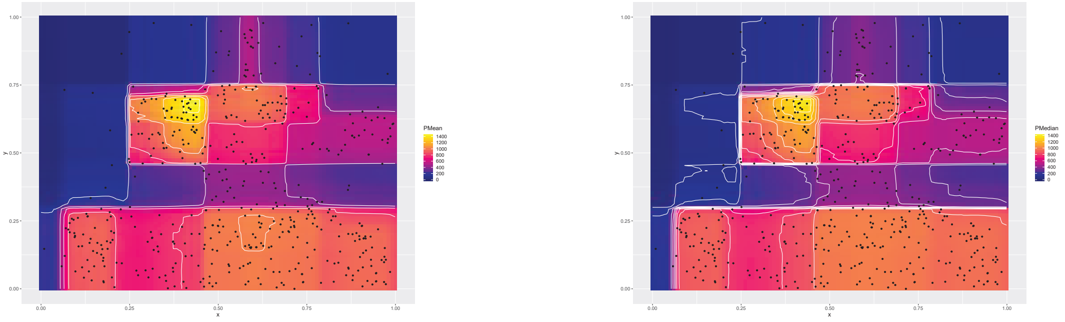
Table 3.20: Average Absolute Error and Root Mean Square Error for Haar-Fisz estimator for the data in Fig. 3.7

## 3.4.2 Lansing Data

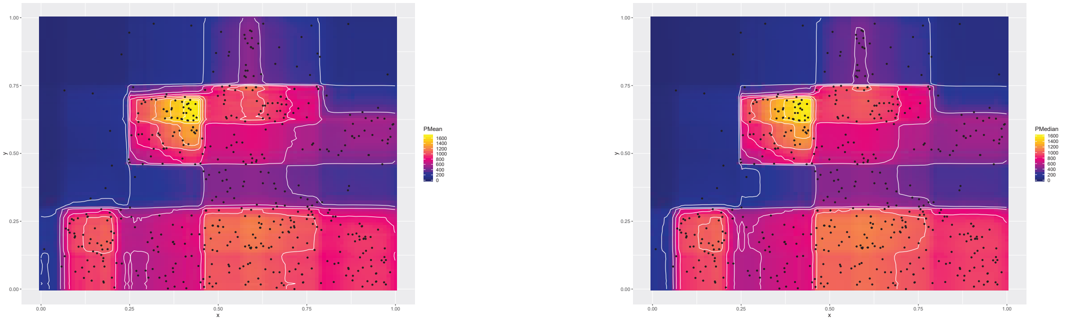
The **lansing** data set included in the **R** package **spatstat** describes the locations of different types of trees in the Lansing woods forest. Our attention is restricted to the locations of 514 maples that are presented with dots in Figures 3.9-3.10. We run 3 parallel chains of the same length for 200000 iterations for 3-10 trees and 100000 iterations for 12 trees. The diagnostic criteria included in the supplementary material indicate that the considered chains have converged for the majority of testing points.

We compare our algorithm to a fixed bandwidth estimator using a Gaussian kernel. Our algorithm and the kernel estimator are consistent in the overall structure. The differences are due to the different nature of the methods. Given the tree locations, our algorithm recovers the spatial pattern of trees as rectangular regions of different intensities (Fig. 3.9), whereas the kernel method produces a continuum with more localized peaks in space. As expected, the kernel estimator presented in Figure 3.10 consists of smoother subregions with various intensities. Tables 3.21-3.23 show that our algorithm is competitive to kernel smoothing with fixed bandwidth chosen with likelihood cross-validation. In contrast to our method, kernel methods are highly sensitive to parameter (bandwidth) choice.

The diagnostics $D_g$ and $D_l$ obtain their highest values at 3 and 4 trees, respectively.



(a) Posterior Mean for 5 Trees (b) Posterior Median for 5 Trees



(c) Posterior Mean for 10 Trees (d) Posterior Median for 10 Trees

Figure 3.9: Posterior Mean and Posterior Median for 5 and 10 Trees

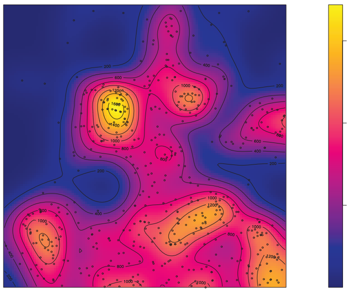


Figure 3.10: Fixed-bandwidth chosen using likelihood cross-validation.

| | Proposed BART Algorithm | | | | | |
|---|---|---|---|---|---|---|
| Number of trees | AAE for Posterior Mean | AAE for Posterior Median | RMSE for Posterior Mean | RMSE for Posterior Median | $D_g$ | $D_l$ |
| 3 | 1.3 | 1.2 | 1.7 | 1.8 | 5686.5 | 5705.3 |
| 4 | 1.2 | 1.2 | 1.7 | 1.8 | 5683.8 | 5709.5 |
| 5 | 1.2 | 1.2 | 1.7 | 1.7 | 5672.4 | 5705.4 |
| 7 | 1.2 | 1.2 | 1.7 | 1.71 | 5643.5 | 5702 |
| 8 | 1.2 | 1.2 | 1.7 | 1.7 | 5634 | 5707.2 |
| 9 | 1.2 | 1.2 | 1.7 | 1.7 | 5614.3 | 5698.1 |
| 10 | 1.2 | 1.2 | 1.6 | 1.7 | 5596.6 | 5699.8 |
| 12 | 1.2 | 1.2 | 1.7 | 1.7 | 5558.2 | 5692.5 |

Table 3.21: Average Absolute Error, Root Integrated Square Error with $N_S = 225$ and diagnostics for the data in Fig. 3.9.

| | Proposed BART Algorithm | | | |
|---|---|---|---|---|
| Number of trees | AAE for Posterior Mean | AAE for Posterior Median | RMSE for Posterior Mean | RMSE for Posterior Median |
| 3 | 0.9 | 0.9 | 1.3 | 1.3 |
| 4 | 0.9 | 0.9 | 1.2 | 1.3 |
| 5 | 0.9 | 0.9 | 1.2 | 1.3 |
| 7 | 0.9 | 0.9 | 1.2 | 1.2 |
| 8 | 0.9 | 0.9 | 1.2 | 1.2 |
| 9 | 0.9 | 0.9 | 1.2 | 1.2 |
| 10 | 0.9 | 0.9 | 1.2 | 1.2 |
| 12 | 0.9 | 0.9 | 1.2 | 1.2 |

Table 3.22: Average Absolute Error and Root Integrated Square Error with $N_S = 400$ for the data in Fig. 3.9.

| Kernel Smoothing | | |
|---|---|---|
| Bandwidth (sigma) | AAE | RISE |
| 0.05 (LCV) for $N_S = 225$ | 1.03 | 1.42 |
| 0.05 (LCV) for $N_S = 400$ | 0.82 | 1.13 |

Table 3.23: Average Absolute Error and Root Integrated Square Error for fixed bandwidth estimators for data in Fig. 3.10.

## 3.5 Discussion

In this chapter, we have studied how the Bayesian Additive Regression Trees (BART) model can be applied to estimating the intensity of Poisson processes. The BART framework provides a flexible non-parametric approach to capturing non-linear and additive effects in the underlying functional form of the intensity. Our numerical experiments show that our algorithm provides good approximations of the intensity with ensembles of less than 10 trees. This enables our algorithm to detect the dimensions contributing most to the intensity. The ability of our method to identify important features demonstrates an important advantage over other procedures.

Our approach enables full posterior inference of the intensity in a non-parametric regression setting. In addition, the method extends easily to higher dimensional settings. The simulation study on synthetic data sets show that our algorithm can detect change points and provides good estimates of the intensity via either the posterior mean or the posterior median. Our algorithm is competitive with the Haar-Fisz algorithm and kernel methods in one and two dimensions and inference using spatial log-Gaussian Cox processes. The strength of our method is its performance in higher dimensions, and we demonstrate that it outperforms the kernel approach for multidimensional intensities. We also demonstrate that our inference for the intensity is consistent with the variability of the rate of events in real and synthetic data. The convergence criteria included in Appendices A.3-A.6 indicate good convergence of the considered chains. We ran each chain for at least 100000 iterations to increase our confidence in the results. However, our algorithm works well with considerably fewer iterations (around 10000).

The BART model assumes independence of the underlying tree structure. The alternative method of [Sardy and Tseng, 2004] makes use of a locally dependent Markov Random Field, and one way of extending our model in this direction is to consider neighbouring intensities following Chipman et al. [2021].

Our method has only considered the standard priors commonly used in BART procedures, an interesting avenue of future research would be to implement different prior assumptions. In addition, we have fixed the parameters for the Galton-Watson prior on the trees, and further work on sensitivities to hyperparameter selection and alternative methods for inference of the hyperparameters is of interest. Currently, our model is limited to non-homogeneous Poisson process and we believe the flexibility of the BART approach could be extended to more general point processes.

# Chapter 4

# Axis-alignment effect

In Chapter 3, we have proposed a Metropolis-Hastings within block Gibbs sampler (Algorithm 4) for inferring the parameters $\{\Lambda_h, T_h\}_{h=1}^m$, given a finite realization of an inhomogeneous Poisson process. A drawback of our proposed algorithm is its axis-alignment nature; the imposed prior on the trees is a stochastic process that considers axis-aligned cuts, resulting in partitioning the domain into hyper-rectangular regions. The simulation studies showed that points close to jumps are estimated with less reliability, which is expected due to the axis-alignment effect. Ge et al. [2019]; Fan et al. [2016] emphasize that the axis-aligned nature of the decision boundaries restricts the model's flexibility, which may affect the ability to capture interdimensional dependencies in the domain. Recently, many studies including [Ge et al., 2019; Fan et al., 2016; Tomita et al., 2020; Rainforth and Wood, 2015; Rodriguez et al., 2006; Blaser and Fryzlewicz, 2021, 2016] have emerged proposing methods for dealing with the axis-alignment effects.

Bayesian non-parametric methods have been developed to allow more flexible non-axis aligned partitioning. Fan et al. [2016] introduced a non-self-consistent stochastic partition process, named the Ostomachion process (OP), which produces polygonal partitions on the unit square. To address missing properties, Fan et al. [2018] introduced a stochastic partition on an arbitrary two-dimensional convex polygon, named a Binary Space Partitioning (BSP) Tree process. The BSP-Forest, which is an extension of the BSP-Trees procedure to a higher dimensional space with all but two cutting hyperplanes parallel to dimensions, was introduced by Fan et al. [2019]. The Random Tessellation Process (RTP), which considers non-axis-aligned cuts and divides the space into polytopes, was proposed by Ge et al. [2019]. Ge et al. [2019] also proposed a random forest of RTPs, named Random Tessellation Forests (RTFs), by training RTPs independent of each other. Recently, Maia et al. [2022] considered non-axis aligned cuts of two-dimensional spaces by introducing two new moves for learning the structure of trees in the standard BART algorithm [Chipman et al., 2010].

Alternative non-Bayesian methods considering feature rotations from structured ([Rodriguez et al., 2006]) to unstructured ([Blaser and Fryzlewicz, 2021, 2016]) before building the tree ensemble improved classification performance. Blaser and Fryzlewicz [2021] emphasize that the majority of rotations do not enhance the performance of the out-of-sample classifier. To construct simple and compact base learners for a particular classification task, Blaser and Fryzlewicz [2021] proposed a technique that favours the most helpful rotations.

Following Blaser and Fryzlewicz [2016], we run the proposed algorithm in Chapter 3 using various rotated versions of a finite realization of an inhomogeneous Poisson process to remedy the axis-alignment nature of the proposed algorithm. Specifically, we estimate the intensity at each point as the average mean of the estimates given by tiny ensembles of trees using the rotations with the higher estimated log-likelihood, which we name helpful rotations in line with Blaser and Fryzlewicz [2016]. The simulation analysis in two dimensions demonstrates the advantages of considering helpful rotations in intensities with non-axis-aligned boundaries. Complementing the algorithm using random rotations can outperform the kernel approach for two-dimensional intensities.

The outline of the chapter is as follows. Section 4.1 introduces our approach for investigating the advantages of considering various rotated versions of the domain. Section 4.2 describes an ad-hoc method for estimating the intensity. Sections 4.3 and 4.4 present the application of the new approach to synthetic data and real data sets, respectively. Section 4.5 provides our conclusions and future work.

## 4.1 Model

Consider an inhomogeneous Poisson process defined on a $d$-dimensional domain $S \subset \mathbb{R}^d$, $d > 1$, with intensity $\lambda : S \to \mathbb{R}^+$. We extend the model 3.1-3.3 introduced in Chapter 3 by considering a rotated version of the domain $S$.

To estimate the intensity of the inhomogeneous Poisson process, we use an ensemble of $m$ trees, $T = \{T_h\}_{h=1}^m$. Each tree $T_h = \{\Omega_{ht}\}_{t=1}^{b_h}$ is generated on the randomly rotated domain $S\mathrm{R}^{\mathrm{T}}$ independent of the other trees and associated with the rotation matrix R uniformly sampled over all feasible rotations of $\mathbb{R}^d$, where $b_h$ is the number of terminal nodes in the corresponding tree $T_h$. Each leaf node $t$ corresponds to one of the subregions of $T_h$, $\Omega_{ht}$, which has an associated parameter $\lambda_{ht}$, and hence each tree $T_h$ has an associated vector of leaf intensities $\Lambda_h = (\lambda_{h1}, \lambda_{h2}, \ldots, \lambda_{hb_h})$.

We model the intensity of $\boldsymbol{s} \in S$ as follows,

$$\mathrm{R} \in \mathbb{R}^{\mathrm{d}\times\mathrm{d}} \sim \text{ uniformly sampled over all feasible rotations } \mathbb{R}^d \ , \tag{4.1}$$

$$\lambda(\boldsymbol{s}) = \prod_{h=1}^{m}\prod_{t=1}^{b_h} \lambda_{ht}^{I(\mathrm{Rs}\in\Omega_{\mathrm{ht}})}, \tag{4.2}$$

$$T_h \sim \text{heterogeneous Galton-Watson process for a partition of the rotated domain} \tag{4.3}$$

$$\lambda_{ht}|T_h \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta),\ t = 1, \ldots, b_h,\ h = 1, \ldots, m \tag{4.4}$$

Given a fixed number of trees, $m$, the parameters of the model are thus the regression trees $T = \{T_h\}_{h=1}^m$, and their corresponding intensities $\Lambda = \{\Lambda_h\}_{h=1}^m$ and the rotation matrix R.

## 4.2 Ad-hoc Inference Algorithm using helpful rotations

Given a finite realization of an inhomogeneous Poisson process $\boldsymbol{s} \in S \subset \mathbb{R}^d$, we introduce an ad-hoc method for estimating the intensity.

Following Blaser and Fryzlewicz [2016], we run Algorithm 4 introduced in Chapter 3 using $M$ various rotated versions of $\boldsymbol{s}$ (see Algorithm 6) with normalized each dimension of the rotated domain from $(0,1)$. We use the index $r$ to indicate the multiple rotations with $r = 1, \ldots, M$. We estimate the log-likelihod associated to the rotation matrix $\mathrm{R_r}$, as follows,

$$\hat{l}_r = \log \left( \frac{1}{N_w} \sum_{w=1}^{N_w} P\left( \boldsymbol{s} \middle| \left( T_r^{(w)}, \Lambda_r^{(w)} \right) \right) \right)$$

where $(T_r^{(w)}, \Lambda_r^{(w)})$ is the sequence of draws after the burn-in period with $T_r^{(w)} = \{T_{rh}^{(w)}\}_{h=1}^{b_{rh}}$ and $\Lambda_r^{(w)} = \{\{\lambda_{rht}^{(w)}\}_{t=1}^{b_{rh}}\}_{h=1}^{m}$.

Let $v_1, v_2, .., v_M$ be a permutation of $\{1, .., M\}$ s.t. $\hat{l}_{v_1} \geq \hat{l}_{v_2} \geq \ldots \geq \hat{l}_{v_M}$. We estimate the intensity at each point $\boldsymbol{s}$ as the average mean of the estimates over the higher ranked rotations in estimated log-likelihood, named helpful rotations,

$$\lambda(\boldsymbol{s}) = \frac{1}{|\mathcal{S}_H|} \sum_{r \in \mathcal{S}_H} \lambda_r(\boldsymbol{s})$$

where

$$\lambda_r(\boldsymbol{s}) = \frac{1}{N_w} \sum_{w=1}^{N_w} \prod_{h=1}^{m} \prod_{t=1}^{b_{rh}} {\lambda_{rht}^{(w)}}^{\mathbb{1}\left( R_r \boldsymbol{s} \in \Omega_{rht}^{(w)} \right)} \tag{4.5}$$

and $\mathcal{S}_H = \{v_1, .., v_k\}$ is the set of $k$ helpful rotations.

**Algorithm 6** Ad-hoc Method

```
for $r = 1, .., M$ do
    $\mathrm{R_r} \sim$ uniformly sampled over all feasible rotations
    Sample $\left(T_r^{(w)}, \Lambda_r^{(w)}\right) \sim P\left(T_r, \Lambda_r | \boldsymbol{s}, \mathrm{R_r}\right)$ applying Algorithm 4
    Estimate the log-likelihood, $\hat{l}_r$, applying Equation 4.5
end for
```

## 4.3 Simulation Analysis

We carried out a simulation study on synthetic data to illustrate the performance of using helpful rotations to estimate the intensity of two-dimensional inhomogeneous Poisson processes. To assess the performance of our algorithm, we apply the metrics AAE and RISE defined in Section 3.3. We compare the results of our algorithm with fixed-bandwidth estimators by choosing a Gaussian kernel and finding the bandwidth via likelihood cross-validation (LCV).

|  | RISE | AAE |
|---|---|---|
| kernel (LCV) | 108.1 | 80.5 |
| unrotated | 119.6 | 90.8 |
| 20 helpful rotations | 97.2 | 69.9 |
| 5 helpful rotations | 88.1 | 58.9 |
| 1 helpful rotation | 89.7 | 58.1 |

Table 4.1: Average Absolute Error and Root Integrated Square Error for our estimated intensity with and without (unrotated) helpful rotations, and fixed bandwidth estimators for the scenario of Inhomogeneous Poisson process in a disk with one diagonal boundary illustrated in Fig. 4.1

|  | RISE | AAE |
|---|---|---|
| kernel (LCV) | 167.8 | 104.7 |
| unrotated | 208.2 | 129.5 |
| 20 helpful rotations | 148.7 | 86.4 |
| 5 helpful rotations | 141.6 | 79.8 |
| 1 helpful rotation | 155.1 | 74.3 |

Table 4.2: Average Absolute Error and Root Integrated Square Error for our estimated intensity with (rotated) and without (unrotated) helpful rotations, and fixed bandwidth estimators for the scenario of Inhomogeneous Poisson process in a disk with two diagonal boundaries illustrated in Fig. 4.2

### 4.3.1 Inhomogeneous Poisson processes in a disk

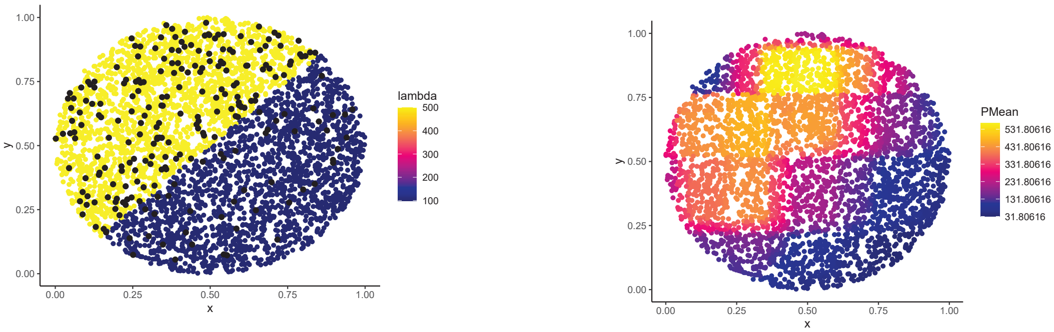
In the simulation scenarios, we consider that the under consideration domain is a disk with the centre located at $(0.5, 0.5)$ and a radius equal to 0.5. We rotate the point patterns around the centre of the disk. The rotation matrix is defined by

$$\mathrm{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

with $\theta$ being uniformly sampled between 0 and $2\pi$. The testing points are a set $\{z_i\}$ defined by uniformly sampling points in the disk.

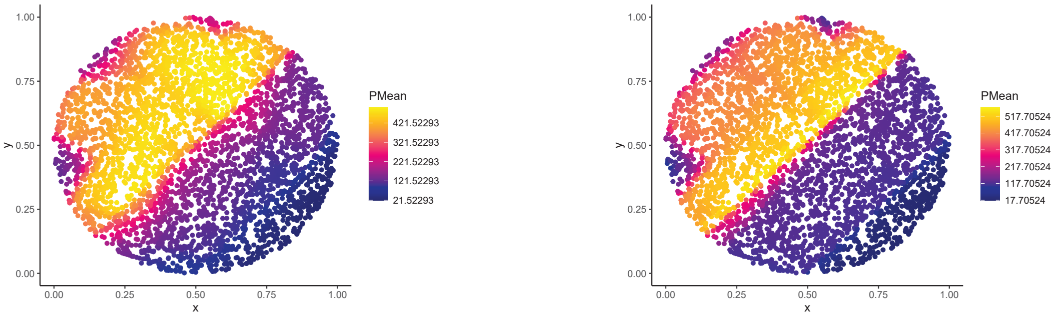
We run Algorithm 6 using 150 random rotations for inhomogeneous Poisson processes, illustrated in Figures 4.1-4.2. We run 3 parallel chains of the same length for 20000 iterations for 5 trees for each rotation. The criterion indicates convergence of the simulating chains for all randomly chosen rotations for both scenarios.

Figures 4.3-4.4 illustrate the twenty helpful rotations in both scenarios. As we expected, the rotations with the higher estimated log-likelihood are those that align segments of intensity boundaries with the axes. Figures 4.1-4.2 illustrate that complementing Algorithm 4 with helpful rotations captures the diagonal nature of intensity boundaries with ensembles of just five trees. Considering helpful rotations improves the metrics and outperforms the kernel estimator (see Tables 4.1-4.2).



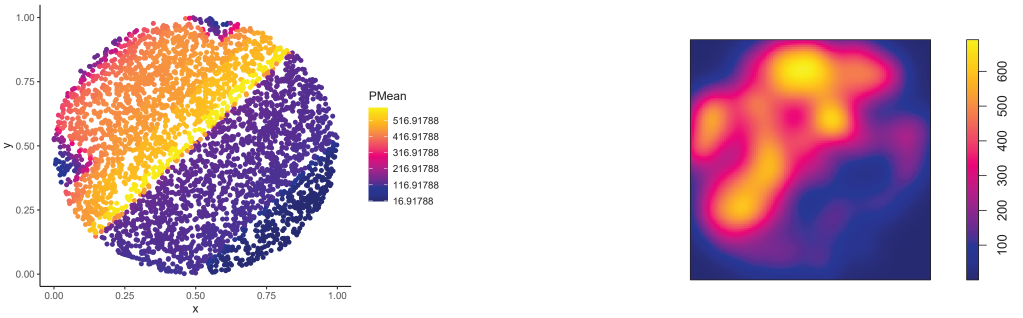
(a) Realization of process presented by black dots and original intensity

(b) unrotated



(c) 20 helpful rotations

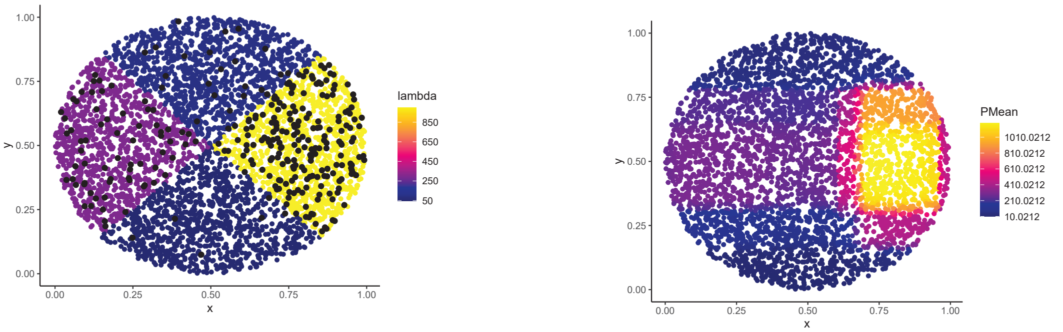
(d) 5 helpful rotations



(e) 1 helpful rotation

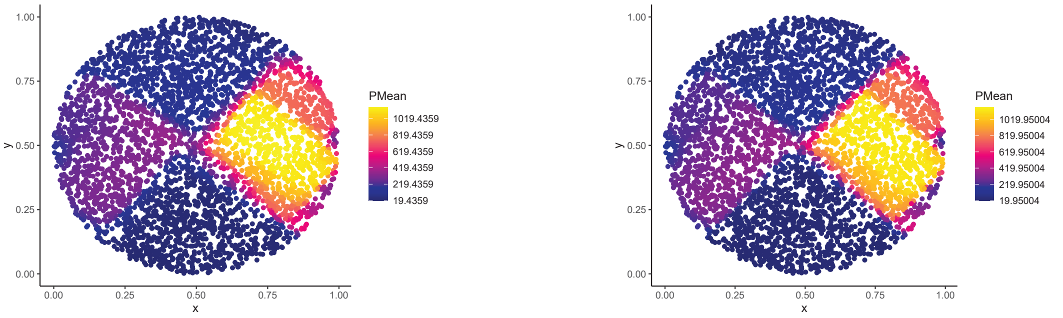
(f) Kernel Estimator

Figure 4.1: Original and estimated intensity for 5 Trees with and without (unrotated) helpful rotations for the scenario of Inhomogeneous Poisson process in a disk with one diagonal boundary.



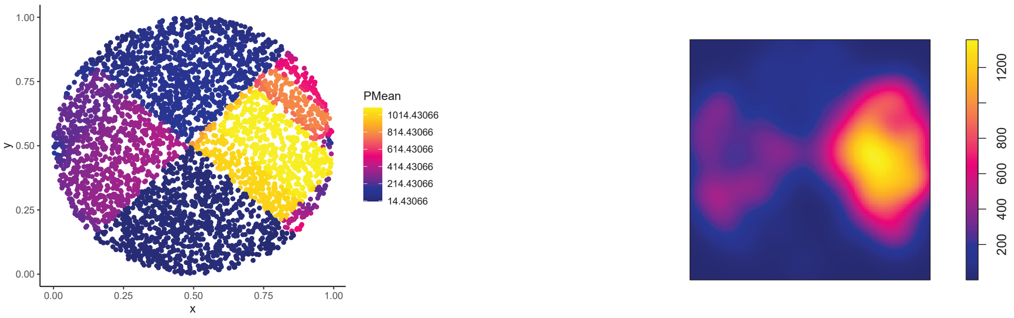
(a) Realization of process presented by black dots and original intensity

(b) unrotated



(c) 20 helpful rotations

(d) 5 helpful rotations



(e) 1 helpful rotation

(f) Kernel Estimator

Figure 4.2: Original and estimated intensity for 5 Trees with and without (unrotated) helpful rotations for the scenario of Inhomogeneous Poisson process in a disk with two diagonal boundaries.

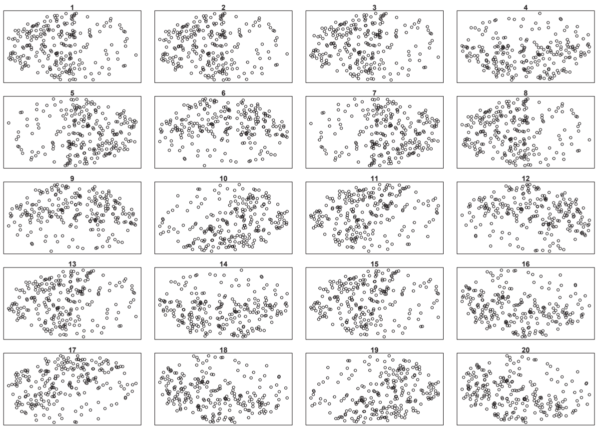


Figure 4.3: The twenty helpful rotations of the point pattern of Inhomogeneous Poisson process in a disk with one diagonal boundary illustrated in Figure 4.1. The top left is the best rotation, and the bottom right is the $20^{th}$.

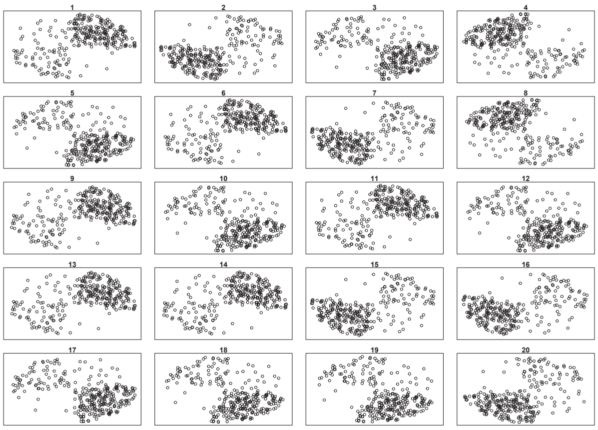


Figure 4.4: The twenty helpful rotations of the point pattern of Inhomogeneous Poisson process in a disk with two diagonal boundaries illustrated in Figure 4.2. The top left is the best rotation, and the bottom right is the $20^{th}$.

| | RISE | AAE |
|---|---|---|
| kernel (LCV) | 166.2 | 111.4 |
| unrotated | 186.6 | 117.1 |
| 20 helpful rotations | 165.5 | 112 |
| 5 helpful rotations | 150.7 | 96.6 |
| 1 helpful rotation | 153.9 | 95.3 |

Table 4.3: Average Absolute Error and Root Integrated Square Error for our estimated intensity with and without (unrotated) helpful rotations using ensembles of 5 trees, and fixed bandwidth estimators for Inhomogeneous Poisson process in a square illustrated in Fig. 4.5

| | RISE | AAE |
|---|---|---|
| kernel (LCV) | 166.2 | 111.4 |
| unrotated | 190.6 | 123.3 |
| 20 helpful rotations | 164.3 | 111.2 |
| 5 helpful rotations | 151.4 | 96.1 |
| 1 helpful rotation | 156.1 | 100.7 |

Table 4.4: Average Absolute Error and Root Integrated Square Error for our estimated intensity with and without (unrotated) helpful rotations using ensembles of 7 trees, and fixed bandwidth estimators for Inhomogeneous Poisson process in a square illustrated in Fig. 4.6

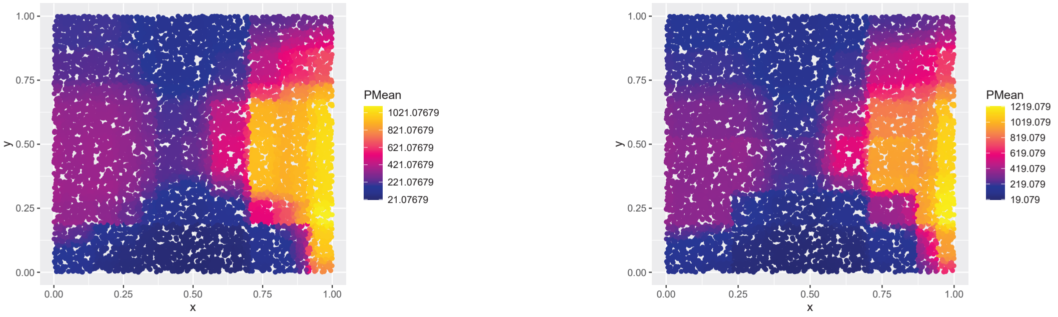
### 4.3.2 Inhomogeneous Poisson process in a square

We run Algorithm 6 using 100 random rotations for a two-dimensional Poisson process in a unit square, illustrated in Figure 4.6, considering ensembles of five and seven trees. We run three parallel chains of the same length for 20000 iterations for each rotation. The testing points are a set of uniformly sampling 10000 points in the unit square. The convergence criterion indicates convergence of the simulation chains for more than 60% and 75% of testing points for all rotations for five and seven trees, respectively. We observe that increasing the number of trees improves the convergence of chains for the rotated domains, which is under investigation.

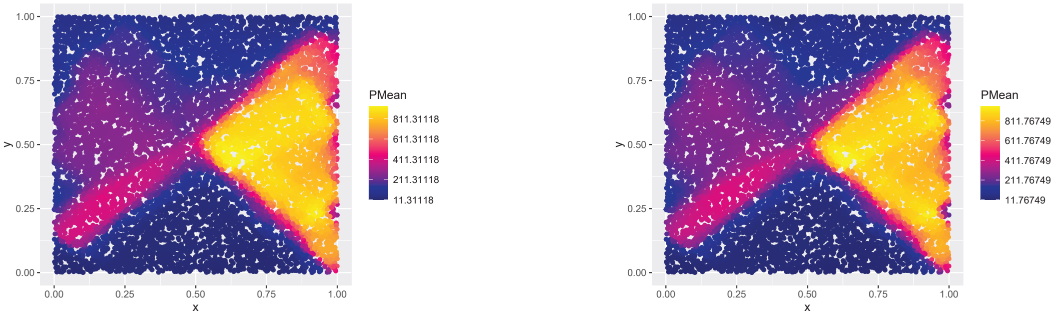
We illustrate the twenty helpful rotations in Figures 4.7-4.8 for ensembles of five and seven trees, respectively. The rotations that line up intensity boundary segments with the axes are those with the highest calculated log-likelihood. Using helpful rotations captures the diagonal nature of intensity boundaries, improves the metrics, and wins kernel smoothing with edge correction (See Figures 4.5-4.6 and Tables 4.3-4.4).

## 4.4 Intensity estimation for Real Data

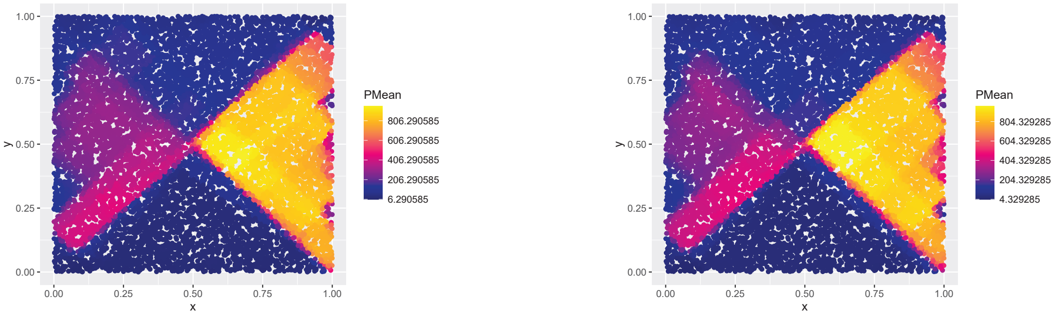
We apply the rotation approach considering 100 random rotations to the locations of 514 maples introduced in Subsection 3.4.2, and presented with dots in Figure 4.9. We run three parallel chains of the same length for 20000 iterations for five trees.



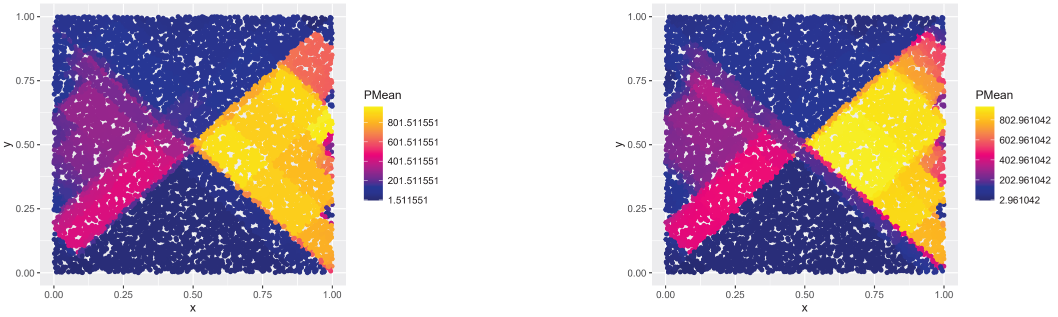
(a) unrotated with 5 trees (b) unrotated with 7 trees



(c) 20 helpful rotations with 5 trees (d) 20 helpful rotations with 7 trees

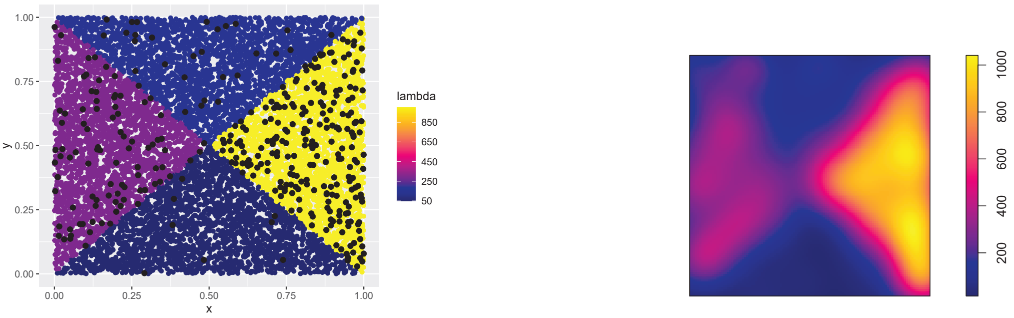


(e) 5 helpful rotations with 5 trees (f) 5 helpful rotations with 7 trees



(g) 1 helpful rotation with 5 trees (h) 1 helpful rotation with 7 trees

Figure 4.5: Estimated intensity of Inhomogeneous Poisson process in a square with diagonal boundaries for 5 and 7 Trees with and without (unrotated) helpful rotations.



(a) Original Intensity and realization of process (black dots)

(b) Kernel Estimator

Figure 4.6: Original intensity of Inhomogeneous Poisson process in a square and kerrnel estimator.

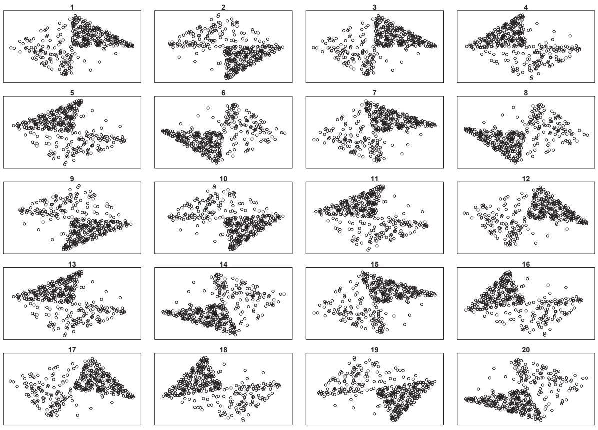


Figure 4.7: The twenty helpful rotations of the point pattern of Inhomogeneous Poisson process in a square illustrated in Figure 4.6, using ensembles of 5 trees. The top left is the best rotation, and the bottom right is the $20^{th}$ best rotation.

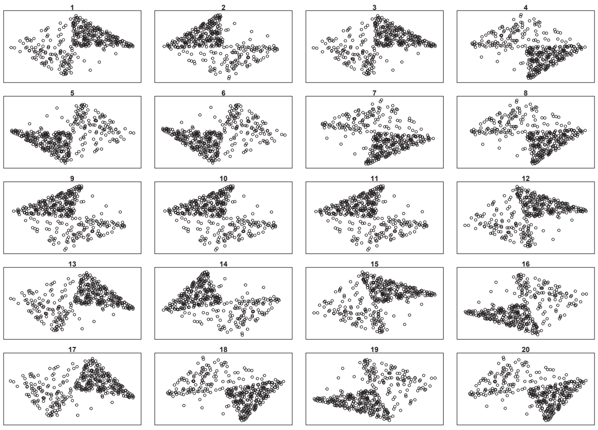


Figure 4.8: The twenty helpful rotations of the point pattern of Inhomogeneous Poisson process in a square illustrated in Figure 4.6, using ensembles of 7 trees. The top left is the best rotation, and the bottom right is the $20^{th}$ best rotation.

|  | RMSE | AAE |
|---|---|---|
| kernel (LCV) | 1.41 | 1.03 |
| unrotated | 1.7 | 1.24 |
| 20 helpful rotations | 1.64 | 1.2 |
| 5 helpful rotations | 1.66 | 1.24 |
| 10 helpful rotation | 1.65 | 1.21 |

Table 4.5: Average Absolute Error and Root Integrated Square Error for our estimated intensity with and without (unrotated) helpful rotations using ensembles of 5 trees, and fixed bandwidth estimators with $N_S = 225$ for maples illustrated in Fig. 4.9

|  | RMSE | AAE |
|---|---|---|
| kernel (LCV) | 1.13 | 0.82 |
| unrotated | 1.24 | 0.9 |
| 20 helpful rotations | 1.22 | 0.88 |
| 5 helpful rotations | 1.23 | 0.88 |
| 10 helpful rotation | 1.23 | 0.88 |

Table 4.6: Average Absolute Error and Root Integrated Square Error for our estimated intensity with and without (unrotated) helpful rotations using ensembles of 5 trees, and fixed bandwidth estimators with $N_S = 400$ for maples illustrated in Fig. 4.9

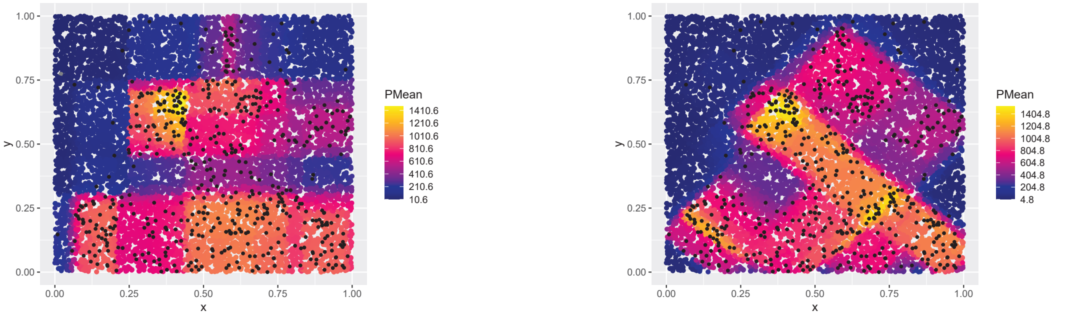
The diagnostic criteria indicate convergence of the chains for the majority of testing points.

We compare our intensity estimates with kernel estimators applying a residual analysis with metrics AAE and RMSE, following the process described in Section 3.4. Figure 4.9 and Tables 4.5-4.6 show that complementing Algorithm 4 with helpful rotations derives estimators with smoother boundaries competitive to kernel fixed-bandwidth estimator and slightly improves the metrics. The point pattern does not indicate a natural rotation that aligns all the cuts with the axes.

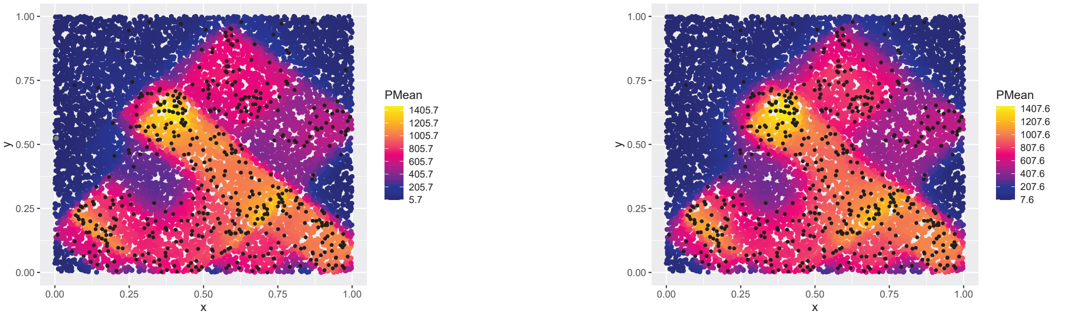
## 4.5 Discussion

This chapter introduced a novel method to remedy the axis-alignment effect. We illustrate the effectiveness of the method in training Algorithm 4 using randomly rotated versions of finite realizations of two-dimensional inhomogeneous Poisson processes. The simulation analysis on synthetic data illustrates that using up to 20 helpful rotations captures the diagonal nature of boundaries, improves the metrics, and outperforms the kernel estimators. The simulation analysis using the Lansing data set demonstrates the random approach makes also smoother the boundaries and derives estimates similar to kernel smoothing. Future work includes investigating the rotation approach into higher-dimensional intensities.

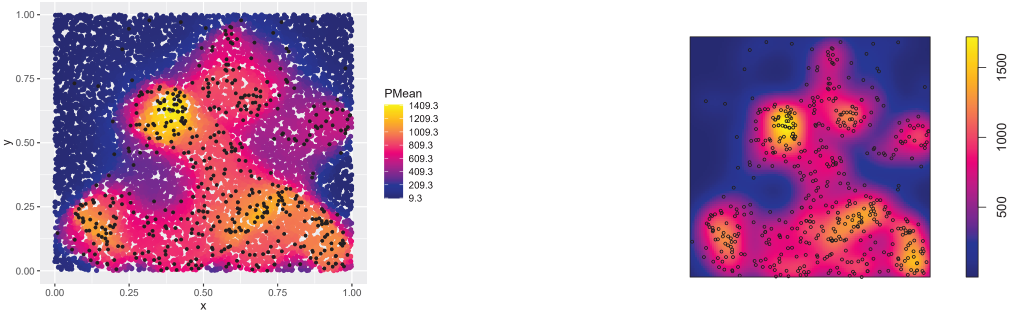
Further work in two dimensions includes the Bayesian inference of the rotation



(a) unrotated (b) 5 helpful rotations



(c) 10 helpful rotations (d) 20 helpful rotations



(e) 30 helpful rotations (f) kernel smoothing

Figure 4.9: Estimate intensity with and without (unrotated) helpful rotations, and kernel estimator for maples.

matrix

$$\mathrm{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

That is, to infer the parameters of the model $(\Lambda, T, \mathrm{R})$ by sampling from the posterior $P(\Lambda, T, \mathrm{R}|\mathrm{s})$. To do that, we implement a Metropolis-Hastings within a block Gibbs sampler, which requires $m$ successive draws from $T_h, \Lambda_h|T_{(h)}, \Lambda_{(h)}, \mathrm{R}, \mathrm{s}$ and one from $\mathrm{R}|\mathrm{T}, \Lambda, \mathrm{s}$, where $T_{(h)} = \{T_j\}_{j=1,j\neq h}^{m}$ and $\Lambda_{(h)} = \{\Lambda_j\}_{j=1,j\neq h}^{m}$. To sample from $P(\mathrm{R}|\mathrm{T}, \Lambda, \mathrm{s})$, we implement a Metropolis-Hastings Algorithm (Algorithm 7). The transition kernel randomly selects the angle $\theta$, such that $\theta \sim \mathrm{U}\left(\theta^{(v)} - \delta, \theta^{(v)} + \delta\right)$, with $\theta^{(v)}$ being the angle associated to the current state $v$ of the chain and $\delta$ a tuning factor, and then rotates the realization $\boldsymbol{s}$. Further work also includes allowing trees of the ensemble to be associated with different rotation matrices and investigating the approach into higher-dimensional intensities.

Barr and Schoenberg [2010] estimated the intensity of an inhomogeneous planar Poisson process constructing the Voronoi tessellation of a point pattern. They compared the performance of the Voronoi estimator to kernel estimators demonstrating the advantages of using Voronoi tessellations in some circumstances. Further work includes considering adaptations of Voronoi tessellation as imposing prior on trees.

**Algorithm 7** Metropolis-Hastings Algorithm for sampling from the posterior $P\left(R|T, \Lambda, \boldsymbol{s}\right)$

```
Sample $\theta^* \sim \mathrm{U}\left(\theta^{(v)} - \delta, \theta^{(v)} + \delta\right)$
Set $\theta^{(v+1)} = \theta^*$ with probability
    $\alpha(\theta^{(v)}, \theta^*) = \min\left\{1, \frac{P(\theta^{(v)}|\theta^*)}{P(\theta^*|\theta^{(v)})} \frac{P(\boldsymbol{s}|\theta^*, T, \Lambda)}{P(\boldsymbol{s}|\theta^{(v)}, T, \Lambda)}\right\}$
Otherwise, set $\theta^{(v+1)} = \theta^{(v)}$.
```

# Chapter 5

# Using a latent Hawkes process for epidemiological modelling

The novel coronavirus disease (COVID-19) has been declared a Global Health Emergency of International Concern with over 557 million cases and 6.36 million deaths as of 3 August 2022 according to the World Health Organization. In the absence of vaccines, countries followed mitigation strategies or countermeasures to prevent the rapid spread of COVID-19, such as social distancing, quarantine, mask wearing, and lock-downs.

A large number of studies have been carried out to understand the spread of COVID-19, forecast new cases and when the peak of the pandemic will occur, and investigate "what-if-scenarios". For example, [Ferguson et al., 2020] presented the results of epidemiological modelling looking at a variety of nonpharmaceutical interventions. Several compartmental models [Zou et al., 2020; Chen et al., 2020; Wangping et al., 2020; Roques et al., 2020] using ordinary differential equations (ODE) have been proposed for modelling the spread of COVID-19. Various models using Hawkes processes [Garetto et al., 2021; Kresin et al.; Escobar, 2020; Chiang et al., 2021; Browning et al., 2021; Koyama et al., 2021; Bertozzi et al., 2020], widely used to model contagion patterns, have been introduced as an alternative to ODE models. Others have used a Poisson autoregression model of the daily new observed cases [Agosto and Giudici, 2020] and a Bayesian model linking the infection cycle to observed deaths[Flaxman et al., 2020].

We introduce a novel epidemic model using a latent Hawkes process [Laub et al., 2015] with temporal covariates for modelling the infections. Unlike other Hawkes models, we model the reported cases via a probability distribution with a mean driven by the underlying Hawkes process.

The outline of the chapter is as follows. Section 5.1 introduces the related work, and Section 5.2 our approach for modelling the spread of an epidemic and estimating the intensity and the reproduction number. Section 5.3 presents the proposed inference algorithm for the latent cases and the reproduction number. Sections 5.4 and 5.5 are applications of our proposed algorithm to synthetic and real data sets.

## 5.1 Related work

The Hawkes process is a well known self-exciting process in which the intensity function depends on all previous events assuming infinite population that allow for parametric or nonparametric estimation of the reproduction number (that is the expected number of infections triggered per infected individual). Hawkes processes have been widely used in numerous applications such as social media, criminology and earthquake modelling. In this section, we present the application of the Hawkes processes in the modelling of COVID-19.

First, we briefly review basic compartmental models and their connection with Hawkes process and COVID. The Susceptible-Infected-Recovered (SIR) and Susceptible-Exposed-Infected-Recovered (SEIR) models are the two basic compartmental epidemic models for modelling the spread of infectious disease [Jones, 2007; Zou et al., 2020]. The SIR model defines three classes of individuals: those susceptible to infection (S), those currently infected (I) and those recovered (R). The SEIR model involves an additional compartment (E) that models the exposed individuals without having obvious symptoms. For many diseases, including COVID-19, there is an incubation period during which exposed individuals to the virus may not be as contagious as the infectious individuals. A variant of the SEIR model called SuEIR was introduced by Zou et al. [2020] for modelling and forecasting the spread of COVID. The SuEIR compared to SEIR has an additional compartment (u) that models the unreported cases. Estimates based on compartmental models can be unreliable as they are highly sensitive to initial conditions and parameters such as transmission and recovery rates Escobar [2020].

A stochastic formulation of SIR called Stochastic SIR [Allen, 2008] is a point process having events that are either the recovery times or the infection times of individuals with exponentially distributed recovery times. Rizoiu et al. [2018] introduced the SIR-Hawkes process (also known as HawkesN), which is a generalization of the Hawkes process concerning finite population. They showed that the conditional intensity of the SIR-Hawkes process with no background events and exponential infectious period distribution is identical to the expected conditional intensity of Stochastic SIR with respect to the recovery period distribution. The Hawkes process with gamma infectious period distribution can approximate stage compartment models if the average waiting times in the compartments are independent exponential distributed [Lloyd, 2001; Chiang et al., 2021].

Kresin et al. claim that although the SEIR model is mostly used for COVID modelling compared to the Hawkes process, a Hawkes model offers more accurate forecasts. Specifically, they suggest a SEIR-Hawkes model in which the intensity of newly exposed cases is a function of infection times and size of the population. Chiang et al. [2021] introduced a Hawkes process model of COVID-19 that estimates the intensity of cases and the reproduction number. The reported cases are modelled via a Hawkes process. The reproduction number is estimated via a Poisson regression with spatial-temporal covariates including mobility indices and demographic features. Based on the branching nature of the Hawkes process, Escobar [2020] derived a simple expression for the intensities of reported and unreported COVID-19 cases. The key to this model is that at the beginning of a generation the infectious will either (1) be registered, (2) not be registered but continue being contagious, or (3) recover with fixed probabilities. However, we believe that the probability of remaining contagious and not being registered infectious should be a decreasing function of time and not fixed.

Garetto et al. [2021] proposed a modulated marked Hawkes process for modelling the spread of COVID-19 under the impact of countermeasures. Each mark corresponds to a different class of infectious individuals with specific kernel functions. Three classes of infectious are considered: symptomatic, asymptomatic and super-spreader, for obtaining the average intensity function and the average total number of points up to a specific time. Symptomatic people are those who will develop evident symptoms and by extension they will be quarantined. Asymptomatic people are those who will not develop strong enough symptoms to be quarantined. Super-spreaders are individuals who exert a high infection rate but do not get quarantined. The model estimates the reproduction number taking into account the amount of recourses employed by the health service to discover the infected population, the countermeasures, as well as the stages that all infectious go through: random incubation time, presymptomatic period, random disease period and random residual phase.

Koyama et al. [2021] developed a discrete-time Hawkes model for estimating the temporally changing reproduction number, and hence detecting the change points via assuming a negative binomial distribution for the daily cases. Some analysis such as [Browning et al., 2021; Triambak and Mahapatra, 2021] examined the daily death data to avoid the issues raised from the reported cases. Browning et al. [2021] modelled the reported daily deaths using a discrete-time Hawkes process, where the cases are assumed Poisson distributed. They considered one fixed change point that breaks the period of analysis into two phases: the initial period where the virus is spreading rapidly and the period after the introduction of preventative measures. The model provides accurate predictions for short-time intervals.

In this chapter, we introduce a novel epidemic model using a latent Hawkes process [Laub et al., 2015] with temporal covariates for modelling the infections. We model the reported cases via a probability distribution with a mean driven by the underlying Hawkes process. We propose a Kernel Density Particle Filter (KDPF) [Sheinson et al., 2014; Liu and West, 2001] for inference of both latent cases and reproduction number and predicting the new cases in the near future. It is feasible to employ particle filter type algorithms, like the KDPF, because the computational effort is linear to the number of infections. Modelling the infections via a Hawkes process allows us to estimate by whom an infected individual was infected. We demonstrate the performance of the proposed algorithm on synthetic data and COVID-19 reported cases in various local authorities in the UK.

## 5.2 Model

We introduce a novel epidemic model using a latent Hawkes process of unobserved infections that then trigger a process of reported infection cases.

We focus on an infinite homogeneous population and restrict our attention to an epidemic process over a horizon $[T_0, T)$, $T_0 < T$, in which we assume immunity to re-infection that is a reasonable assumption over the time scales we consider. We break the horizon $[T_0, T)$ into $k$ subintervals $\mathcal{T}_j = [T_{j-1}, T_j)$ for $j = 1, .., k$ with $T_k = T$. We assume that the epidemic is triggered by a set of infectious individuals at the beginning of the process, the times of their infections denoted by a finite set $\mathcal{H}_0$.

The epidemic process is seen as a counting process $N(t)$ with a set of jump times $\mathcal{T}^N = \{t_0 < t_1 < t_2 < \ldots\}$ and intensity given by

$$\lambda^N(t) = \sum_{t_i \in h_t^0} R(t) h(t - t_i)$$

for $t > 0$ with $h_t^0 = \{t_i | t_i < t\} \cup \mathcal{H}_0$ [Laub et al., 2015; Shelton et al., 2018] being the set of all infection events prior to time $t$. The kernel $h(t - t_i)$ represents the relative infectiousness at time $t$ of an infection at time $t_i$. Under the assumption that the transition kernel $h$ is a probability density function with non-negative real-valued support: $h : [0, \infty) \to [0, \infty)$ and $\int\limits_0^\infty h(s) ds = 1$, $R(t)$ represents the instantaneous reproduction number that is the average number of newly infected people that each infected individual would infect if the conditions, such as interventions and control measures for restriction of epidemic, remained as they were at time $t$ [Cori et al., 2013].

It is natural to see the reported infections as a counting process $M(t)$ with a set of jump times $\mathcal{T}^M = \{\tau_1 < \tau_2 < \ldots < \tau_m\}$ and intensity of observed cases at time $\tau$ as a function of the times of infection up to time $\tau$, namely

$$\lambda^M(\tau) = \sum_{t_i \in h_\tau^0} \beta g(\tau - t_i) \tag{5.1}$$

for $\tau > 0$, where $\beta$ is the expected number of observed cases at each time $\tau$ (also known as ascertainment rate). The transition kernel $g(\tau - t_i)$ represents the relative delay between the infection at time $t_i$ and the onset of symptoms at $\tau$. Similar to the transition kernel of latent cases $h$, we specify the transition kernel of observed cases $g$ to be a probability density function with non-negative real-valued support.

Given that the reported cases are usually given on a weekly or daily basis, we model the observed cases falling in $\mathcal{T}_n$, denoted by $Y_n$, via a distribution $G$ having mean $\mu_n$ equal to the expected observed cases in $\mathcal{T}_n$ given by

$$\mu_n = \beta \sum_{t_w \in [0, T_n)} \int\limits_{\max(t_w, T_{n-1})}^{T_n} g(s - t_w) ds.$$

The usual options of $G$ are Negative Binomial (NB) [Koyama et al., 2021; Stocks et al., 2020] and Poisson distribution [Browning et al., 2021; Cori et al., 2013]. We model the reproduction number $R(t)$ as a stepwise function having as many weights as the number of subintervals, that is,

$$R(t) = \prod_{n=1}^{k} R_n^{\mathbb{1}(t \in \mathcal{T}_n)},$$

where $\{R_n\}$ is assumed to be a Markov process. Usually, a random walk on a logarithmic scale [Storvik et al., 2022] or a normal scale [Koyama et al., 2021] is imposed as a prior on the weights $\{R_n\}$.

The model is described by the equations:

$$\lambda^N(t) = R(t) \sum_{t_i \in h_t^0} h(t - t_i),\ t \in [T_0, T) \tag{5.2}$$

$$Y_n \sim G \text{ with mean } E(Y_n) = \mu_n,\ n = 1, .., k \tag{5.3}$$

$$R(t) = \prod_{n=1}^{k} R_n^{\mathbb{1}\{t \in \mathcal{T}_n\}},\ t \in [T_0, T) \tag{5.4}$$

$$\{R_n\}_{n=1}^{k} \text{ is a Markov process} \tag{5.5}$$

$$\mu_n = \beta \sum_{t_w \in [0, T_n)} \int_{\max(t_w, T_{n-1})}^{T_n} g(s - t_w) ds,\ n = 1, .., k\ . \tag{5.6}$$

## 5.3 Inference algorithm

Given a set of observed infections, we seek to infer the counting process $N(t)$ and the reproduction number $R(t)$.

The proposed epidemic model described by the equations 5.2-5.6 is seen as a state-space model with a latent state process $\{X_n : 1\ \leq n \leq k\}$ and an observed process $\{Y_n : 1 \leq n \leq k\}$. Each hidden state $X_n$ consists of the weight $R_n$ associated to $\mathcal{T}_n$ and the set of latent cases $S_n^N$ falling into $\mathcal{T}_n$. The time-constant parameters are the parameters associated with the distribution $G$ and the prior imposed on the weights $\{R_n\}_{n=1}^{k}$. We apply a KDPF for inferring the counting process $N(t)$, the weights $\{R_n\}_{n=1}^{k}$, and the time-constant parameters.

We focus on illustrating the performance of our model on COVID-19. As the COVID-19 reported cases are subject to erroneous observation and for the data we observe the sample variance is larger than the sample mean, we model the observed cases $Y_n$ via a negative binomial distribution (NB) with mean $\mu_n$ and dispersion $v > 0$. We use the following form of the negative binomial distribution

$$P(Y_n|\mu_n, v) = \frac{\Gamma(Y_n + v^{-1})}{Y_n!\Gamma(v^{-1})} \left(\frac{1}{1 + v\mu_n}\right)^{\frac{1}{v}} \left(\frac{v\mu_n}{v\mu_n + 1}\right)^{Y_n}$$

with mean $E(Y_n) = \mu_n$ and variance $\mathrm{var}(Y_n) = \mu_n(1 + v\mu_n)$. Before we discuss the KDPF, we define the transition kernels of the observed and latent cases and the prior on weights $\{R_n\}_{n=1}^{k}$ for the COVID-19. We also suggest a simple method to initialize $\mathcal{H}_0$.

**Transition Kernels** The dynamics of latent and observed cases are determined by the generation interval (GI) and incubation period (IP) [Fine, 2003]. The generation interval is the time interval between the time of infection of the infector (the primary case) and that of the infectee (the secondary case generated by the primary case).

The incubation period is the time interval between the infection and the onset of symptoms in a specific case. Zhao et al. [2021] assume that the GI and IP follow a gamma distribution. They infer that the mean and SD of GI are equal at 6.7 days and 1.8 days and those of IP at 6.8 and 4.1 days by using a maximum likelihood estimation approach and contact tracing data of COVID-19 cases. We follow the same assumption for the GI (namely, the transition kernel of latent cases is a gamma density with a mean at 6.7 days and SD of 1.8 days). We model the time interval between the observed time and actual time of infection as a gamma density with a mean at 8.8 days and SD of 4.1 days (that is, the transition kernel of observed cases is a gamma density having mean equal at 8.8 days and SD of 4.1 days). For the transition kernel of the observed events, we adopt the values inferred by Zhao et al. [2021] for IP with a slightly increased mean to consider the necessary time for conducting a test against COVID-19. Figure 5.1 illustrates the transition kernels. These kernels are a sensible choice to approximately replicate wait times with the aforementioned expectations and variances.

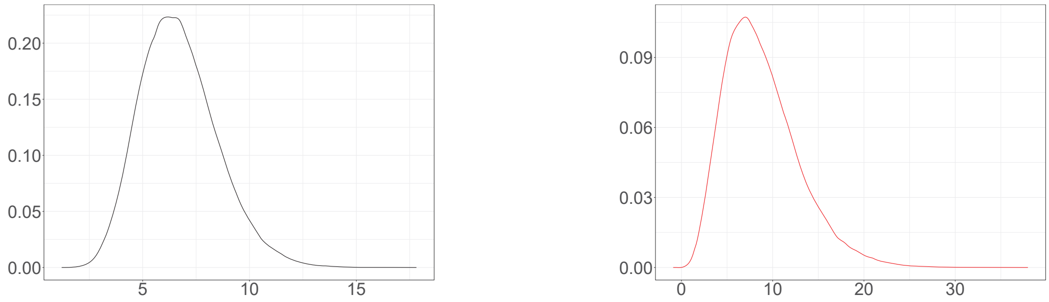


Figure 5.1: The generation interval (GI) (black curve) and the period between observed and actual infection times (red curve).

**Set of infectious at the beginning of the process, $\mathcal{H}_0$** We adopt a heuristic approach to initialize $\mathcal{H}_0$. The transition kernel of latent cases illustrated in Figure 5.1 shows that a latent case at $t_w$ can influence the latent intensity at $t$ if $t_w$ has occurred at most 21 days before $t$. Otherwise, the influence of $t_w$ is negligible. Therefore, as the history of the process, we consider the latent cases of 21 days/3 weeks before the beginning of the process. The transition kernel of observed cases shown in Figure 5.1 demonstrates that an event is most likely to be observed seven days after the actual infection time. Considering the observed cases are daily, we initialize the history of latent case, $\mathcal{H}_0$ by uniformly spreading on the day $-i$ the number of cases occurred on the day $(-i+7)$ times $1/\beta$. In simulation analysis, we propose initialization of $\mathcal{H}_0$ when we deal with weekly reported cases.

**Imposed prior on weights $\{\mathbf{R_n}\}_{\mathbf{n=1}}^{\mathbf{k}}$** A geometric random walk (RW) is imposed as prior on weights $\{R_n\}_{n=1}^{k}$ :

$$\begin{aligned} \log R_n &= \log R_{n-1} + \log \epsilon_n, \quad \epsilon_n \sim \text{Gamma}(d,d),\ n = 2,..,k \\ R_1 &\sim \text{Uniform}(\alpha, b). \end{aligned}$$

We assume that the noise of RW $\epsilon_n$ has a gamma distribution with equal shape and rate at $d$. This induces that the weight $R_n$ is gamma distributed with a mean equal to $R_{n-1}$ and standard deviation $R_{n-1}/\sqrt{d}$. The stronger fluctuations in the observed data, the more flexible modelling we need. Smaller values of $d$ have higher standard deviation and lead to a wider range of possible values of $R_n$ increasing the flexibility of the model.

### 5.3.1 Kernel Density Particle Filter

We apply a KDPF (Algorithm 9) for inferring the counting process $N(t)$, the weights $\{R_n\}_{n=1}^{k}$, and the time-constant parameters. The time-constant parameters for modelling COVID-19 infections are the shape $d$ of the noise $\epsilon_n$ and the dispersion parameter $v$.

The KDPF builds on the auxiliary particle filter (APF) [Pitt and Shephard, 1999; Kantas et al., 2015; Doucet et al., 2009] by adding small random perturbations to all the parameter particles to reduce the sample degeneracy by modelling the time-constant parameters as random quantities and their posterior via a mixture of normal distributions. We assume independence among the time-constant parameters, and, following Sheinson et al. [2014], we use logarithms for the time-constant parameters, as they have positive support:

$$p(\log d_{n+1}, \log v_{n+1} | Y_{1:(n+1)}) = p(\log d_{n+1} | Y_{1:(n+1)}) p(\log v_{n+1} | Y_{1:(n+1)}).$$

The posteriors $p(\log d_{n+1} | Y_{1:(n+1)})$ and $p(\log v_{n+1} | Y_{1:(n+1)})$ are smoothly approximated via a mixture of normal distributions weighted by the sample weights $w_{jn}$ given by

$$\begin{aligned} p(\log d_{n+1} | Y_{1:(n+1)}) &\approx \sum_{j=1}^{N} \omega_{jn} \mathcal{N}(\log d_{n+1} | m_{j,dn}^{(L)}, h^2 V_{nd}^{(L)}) \\ p(\log v_{n+1} | Y_{1:(n+1)}) &\approx \sum_{j=1}^{N} \omega_{jn} \mathcal{N}(\log v_{n+1} | m_{j,vn}^{(L)}, h^2 V_{nv}^{(L)}), \end{aligned}$$

where $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian pdf with mean $\mu$ and variance $\sigma^2$. The KDPF uses a tuning parameter $\Delta \in (0, 1]$ and two quantities as a function of that parameter: $h^2 = 1 - ((3\Delta - 1)/(2/\Delta))^2$ and $a^2 = 1 - h^2$. The parameter $\Delta$ is typically taken to be between 0.95 and 0.99 for reducing the chance of degeneracy [Sheinson et al., 2014; West, 1993].

The mean values and the variances of the posteriors of time-constant parameters are defined as follows[West, 1993; Sheinson et al., 2014]:

$$
\begin{aligned}
m^{(L)}_{j,dn} &= a\log d_{jn} + (1-a)\bar{d}_{Ln}, \quad \bar{d}_{Ln} = \sum_{j=1}^{N} w_{jn}\log d_{jn} \\
m^{(L)}_{j,vn} &= a\log v_{jn} + (1-a)\bar{v}_{Ln}, \quad \bar{v}_{Ln} = \sum_{j=1}^{N} w_{jn}\log v_{jn} \\
V^{(L)}_{nv} &= \frac{V_1}{V_1^2 - V_2}\sum_{j=1}^{N}\omega_{jn}(\log v_{jn} - \bar{v}_{Ln})^2 \\
V^{(L)}_{nd} &= \frac{V_1}{V_1^2 - V_2}\sum_{j=1}^{N}\omega_{jn}(\log d_{jn} - \bar{d}_{Ln})^2,
\end{aligned}
$$

with $V_1 = \sum_{j=1}^{N} w_{jn}$ and $V_2 = \sum_{j=1}^{N} w_{jn}^2$.

Following Sheinson et al. [2014], we define the initial densities of parameters $d$ and $v$ to be log-normal:

$$
\begin{aligned}
\log d &\sim \mathcal{N}(\mu_d, \sigma_d^2),\ \mu_d = \frac{\log(d_{max}) + \log(d_{min})}{2},\ \sigma_d = \frac{\log(d_{max}) - \log(d_{min})}{8} \\
\log v &\sim \mathcal{N}(\mu_v, \sigma_v^2),\ \mu_v = \frac{\log(v_{max}) + \log(v_{min})}{2},\ \sigma_v = \frac{\log(v_{max}) - \log(v_{min})}{8}
\end{aligned}
$$

considering that $d_{min} \le d \le d_{max}$ and $v_{min} \le v \le v_{max}$ . The transition densities of the time-constant parameters are given by

$$
\begin{aligned}
p(\log d_{n+1}|\log d_n) &\sim \mathcal{N}(\log d_{n+1}|a\log d_n + (1-a)\bar{d}_{Ln}, h^2 V^{(L)}_{nd}) \\
p(\log v_{n+1}|\log v_n) &\sim \mathcal{N}(\log v_{n+1}|a\log v_n + (1-a)\bar{v}_{Ln}, h^2 V^{(L)}_{nv}).
\end{aligned}
$$

The initial density of the hidden process is given by

$$
f(x_1|\mathcal{H}_0) = \mathrm{U}\,(R_1; \alpha, b)\, P(S_1^N|R_1, \mathcal{H}_0),
$$

while the transition density is given by

$$
f(x_n|x_{1:(n-1)}, \mathcal{H}_0, d, v) = P(S_n^N|R_n, S^N_{1:(n-1)}, \mathcal{H}_0)P(R_n|R_{n-1}, d).
$$

$U(R_1; \alpha, b)$ denotes that $R_1$ is uniformly distributed within the interval $[\alpha, b]$.

**Sampling the latent cases** We sample the latent cases $S_n^N$ falling into the subinterval $\mathcal{T}_n$ by applying Algorithm 8, which is a simulation procedure based on the branching structure of the Hawkes process [Laub et al., 2015]. The proposed algorithm is a superposition of Poisson processes in the interval $\mathcal{T}_n$; the descendants of each latent event at $t_i$ form an inhomogeneous Poisson process with intensity

$$
\lambda_i(t) = R_n h(t - t_i)
$$

for $t > t_i$ and $t \in [T_{n-1}, T_n)$. This induces that:

**Algorithm 8** Sample $S_n^N | S_{1:(n-1)}^N,\ \mathcal{H}_0, R_n$

```
Input: $S_{1:(n-1)}^N$, $\mathcal{H}_0$, $R_n$
Initialize an empty queue: $Q_t$.
$Q_t = \{S_v^N\}_{v=n-\eta}^{n-1}$ with $\eta$ being the number of former subintervals we consider (the value of $\eta$ is determined by the transition kernel of latent cases).
while $Q_t$ is not empty do
    Remove the first element $t_i$ from $Q_t$.
    Draw the number of events $n_i$ triggered by an event at $t_i$ from a Poisson distribution with parameter $\lambda = R_n \int_{\max(t_i,T_{n-1})}^{T_n} h(s - t_i)ds$ that is the average number of offsprings generated by an event at $t_i$ in $\mathcal{T}_n$.
    Generate $n_i$ events from the truncated distribution $h(t)$ in $[\max(t_i, T_{n-1}), T_n)$, and add the new elements to $S_n^N$ and the back of queue $Q_t$.
end while
Return $S_n^N$.
```

- The number of events $n_i$ triggered by an event at $t_i$ in the interval $\mathcal{T}_n$ is Poisson distributed with parameter

$$\lambda = R_n \int_{\max(t_i,T_{n-1})}^{T_n} h(s - t_i)ds.$$

- The arrival times of the $n_i$ descendants are $t_i + E_i$ with $E_i$ being iid random variables with pdf the truncated distribution $h(t)$ in $[\max(t_i, T_{n-1}), T_n)$.

The computational cost of Algorithm 8 is linear to the number of infections falling into $(\eta + 1)$ consecutive subintervals, that is $O\left(\sum_{v=n-\eta}^{n} |S_v^N|\right)$, with $\eta$ being the number of former subintervals that influence the latent cases falling into $\mathcal{T}_n$ determined by the transition kernel of latent cases. The $O$-notation denotes the asymptotic upper bound [Cormen et al., 2022]

**Who infected whom** The Hawkes process is an excellent option for modelling the evolution of an epidemic due to its mutually exciting nature, making it feasible to estimate by whom an infected individual was infected. Bertozzi et al. [2020] describe how we can infer the primary infection $i$ that triggered a secondary infection $j$ using a self-exciting branching process. The parent of each infection $j$ falling into $\mathcal{T}_z$ is assumed to be sampled from a multinomial distribution parameterized by $\pi_j$, where $\pi_j = \{\pi_{ji}\}_{i \in h_j}$ with

$$\pi_{ji} = \frac{h(t_j - t_i)}{\sum_{t_w \in h^0_{t_j^P}} h(t_j - t_w)}$$

being the probability of secondary infection $j$ having been caused by primary infection $i$, where $h_j^P = \{i : t_i | t_i \in \cup_{v=z-\eta}^{z} \mathcal{T}_v,\ t_i < t_j\}$, $h_j = \{i : t_i \in h_j^P\}$ and

$\eta$ the number of former subintervals that influence the latent cases falling into $\mathcal{T}_z$ determined by the transmission kernel of latent cases ($\eta = 21$ days for COVID-19). Alternatively, by recording the parent of each latent infection at step 7 of Algorithm 8, the proposed model can show the branching structure of the process. This approach increases the computational complexity of the algorithm, as more memory units will be required.

**Computational complexity** The computational cost of each propagation step (steps 7 and 12 of Algorithm 9) at state (interval) $n$ is equal to the cost of Algorithm 8 times the number of particles (N), that is $O\left(N \sum_{v=n-\eta}^{n} |S_v^N|\right)$. The cost of finding weights (steps 8 and 13 of Algorithm 9) at state (interval) $n$ is also $O\left(N \sum_{v=n-\eta}^{n} |S_v^N|\right)$. Hence, the computational cost of Algorithm 9 over all states (intervals) is $O\left(N\eta|\mathcal{T}^N|\right)$. $S_n^N$ is the set of latent cases falling into subinterval $\mathcal{T}_n$ and $\mathcal{T}^N = \cup_n S_n^N$. The algorithm is easily parallelized over $N$.

**Model complexity** The set of parameters for inference includes the two time-constant parameters, $d$ governing the variability of the noise in the reproduction number and $v$ the dispersion parameter of the observed counts, the latent process, $\{S_n^N\}_{n=1}^k$, and the steps of instantaneous reproduction number, $\{R_n\}_{n=1}^k$. There is a set of model parameters, including the ascertainment rate $\beta$, the transition kernels of latent and observed cases, which we consider as given. The set of infectious at the beginning of the process, $\mathcal{H}_0$, is initialized applying the heuristic approach described above. We rely on the Bayesian paradigm for regularizing the parameters for inference.

**Fixed-lag smoothing densities** Resampling results in replicating samples, and in the long run results in a lack of diversity called particle degeneracy [Endo et al., 2019]. As the resampling step leads to path degeneracy, it is difficult to obtain a good approximation of the smoothing density $p(x_{1:T}|y_{1:T})$ for large $T$ via SMC. Therefore, we use SMC to sample from the fixed-lag smoothing densities with lag $L$ that is a reasonable approach as the reported cases after $n$+28 days do not bring any additional information about the infections of the day $n$. We apply the multinomial resampling step when the Effective Sample Size (ESS) is less than the 80% of the number of particles, to avoid unnecessary resampling steps.

## 5.4 Simulation Analysis

We carried out a simulation study on synthetic data to illustrate the performance of the KDPF (Algorithm 9) for inferring the intensity of latent cases, the reproduction number and the time-constant parameters.

Two different scenarios illustrated in Figure 5.2 were simulated as follows:

**Algorithm 9** Kernel density particle filter

```
1: Initialize the parameters $\{\theta_{j1}\}_{j=1}^{N}$, $\theta_{j1} = (d_{j1}, v_{j1})$ with $d_{min} \leq d \leq d_{max}$ and $v_{min} \leq v \leq v_{max}$:
    for $j$ in $1 : N$ do
      $\log d_{j1} = \mathcal{N}(\mu_d, \sigma_d^2)$ with $\mu_d = \frac{\log d_{max} + \log d_{min}}{2}$ and $\sigma_d = \frac{\log d_{max} - \log d_{min}}{8}$
      $\log v_{j1} = \mathcal{N}(\mu_v, \sigma_v^2)$ with $\mu_v = \frac{\log v_{max} + \log v_{min}}{2}$ and $\sigma_v = \frac{\log v_{max} - \log v_{min}}{8}$
    end for
2: Sample $N$ particles $\{X_{j1}\}_{j=1}^{N}$, $X_{j1} = \left(R_{j1},\ S_{j1}^{N}\right)$:
    for $j$ in $1 : N$ do
      $R_{j1} \sim \text{Uniform}(\alpha, b)$
      $S_{j1}^{N} \sim P\left(S_1^N | R_{j1}, \mathcal{H}_0\right)$ using Algorithm 8
    end for
3: Find the weights, $\tilde{w}_1 = \{\tilde{w}_{j1}\}_{j=1}^{N}$:
    for $j$ in $1 : N$ do
      $\tilde{w}_{j1} = P\left(Y_1 | S_{j1}^N, \beta, \mathcal{H}_0, v_{j1}\right)$
    end for
4: Normalize the weights, $w_1 = \{w_{j1}\}_{j=1}^{N}$:
    for $j$ in $1 : N$ do
      $w_{j1} = \frac{\tilde{w}_{j1}}{\sum_{j=1}^{N} \tilde{w}_{j1}}$
    end for
5: for $n = 1, .., k$ do
6:     for $j$ in $1 : N$ do
         $m_{j,dn}^{(L)} = a \log d_{jn} + (1-a)\bar{d}_{Ln}$,    $\bar{d}_{Ln} = \sum_{j=1}^{N} w_{jn} \log d_{jn}$
         $m_{j,vn}^{(L)} = a \log v_{jn} + (1-a)\bar{v}_{Ln}$,    $\bar{v}_{Ln} = \sum_{j=1}^{N} w_{jn} \log v_{jn}$
         $m_{j,dn} = a d_{jn} + (1-a)\bar{d}_n$,    $\bar{d}_n = \sum_{j=1}^{N} w_{jn} d_{jn}$
         $m_{j,vn} = a v_{jn} + (1-a)\bar{v}_n$,    $\bar{v}_n = \sum_{j=1}^{N} w_{jn} v_{jn}$
       end for
7:     For each particle $j$, we calculate an estimate of $X_{j,n+1}$ called $\tilde{X}_{j,n+1}$ by drawing a sample from $P(X_{n+1} | X_n, \mathcal{H}_0)$:
       for $j$ in $1 : N$ do
         $\tilde{R}_{j,n+1} \sim P\left(R_{n+1} | R_{jn}, m_{j,dn}\right)$
         $\tilde{S}_{j,n+1}^{N} \sim P\left(S_{n+1}^N | S_{j,1:n}^N, \tilde{R}_{j,n+1}, \mathcal{H}_0\right)$ using Algorithm 8
       end for
8:     Find the auxiliary weights, $\tilde{g}_{n+1} = \{\tilde{g}_{j,n+1}\}_{j=1}^{N}$:
       for $j$ in $1 : N$ do
         $\tilde{g}_{j,n+1} = g_{jn} w_{jn} P\left(Y_{n+1} | S_{j,1:n}^N, \tilde{S}_{j,n+1}^N, \beta, \mathcal{H}_0, m_{j,vn}\right)$
       end for
9:     Normalize the auxiliary weights, $g_{n+1} = \{g_{j,n+1}\}_{j=1}^{N}$:
       for $j$ in $1 : N$ do
         $g_{j,n+1} = \frac{\tilde{g}_{j,n+1}}{\sum_{j=1}^{N} \tilde{g}_{j,n+1}}$
       end for
10:    if $\left(ESS(g_{n+1}) = 1 / \sum_{j=1}^{N} g_{j,n+1}^2 < 0.8N\right)$ then resample and form $N$ equally weighted particles, $\bar{X}_{1:n} = \{\bar{X}_{1:n}^i\}_{i=1}^{N}$:
       for $j$ in $1 : N$ do
         (i) sample index $i_j$ from a multinomial distribution with probabilities $g_{n+1}$
         (ii) $\bar{X}_{1:n}^j = X_{i_j,1:n}$
         (iii) $g_{j,n+1} = 1$
       end for
       end if
11:    Regenerate the fixed parameters:
       for $j$ in $1 : N$ do
         $\log v_{j,n+1} \sim \mathcal{N}(m_{i_j,vn}^{(L)}, h^2 V_{nv}^{(L)})$
         $\log d_{j,n+1} \sim \mathcal{N}(m_{i_j,dn}^{(L)}, h^2 V_{nd}^{(L)})$
       end for
       where $V_{nv}^{(L)}$ is the weighted variance of $\{\log v_{jn}\}_{n=1}^{N}$ and $V_{nd}^{(L)}$ the weighted variance of $\{\log d_{jn}\}_{n=1}^{N}$.
12:    Using $\bar{X}_{1:n}$ propagate:
       for $j$ in $1 : N$ do
         $R_{j,n+1} \sim P(R_{n+1} | R_{jn}, d_{j,n+1})$
         $S_{j,n+1}^N \sim P\left(S_{n+1}^N | S_{j,1:n}^N, R_{j,n+1}, \mathcal{H}_0\right)$ using Algorithm 8
         Set $X_{j,1:n+1} = (\bar{X}_{1:n}^j, \left(R_{j,n+1}, S_{j,n+1}^N)\right)$
       end for
13:    Find the weights, $\tilde{w}_{n+1} = \{\tilde{w}_{j,n+1}\}_{j=1}^{N}$:
       for $j$ in $1 : N$ do
         $\tilde{w}_{j,n+1} = \frac{P\left(Y_{n+1} | S_{j,1:n+1}^N, \beta, \mathcal{H}_0, v_{j,n+1}\right)}{P\left(Y_{n+1} | S_{j,1:n}^N, \tilde{S}_{i_j,n+1}^N, \beta, \mathcal{H}_0, m_{i_j,vn}\right)}$
       end for
14:    Normalize the weights, $w_{n+1} = \{w_{j,n+1}\}_{j=1}^{N}$:
       for $j$ in $1 : N$ do
         $w_{j,n+1} = \frac{\tilde{w}_{j,n+1}}{\sum_{j=1}^{N} \tilde{w}_{j,n+1}}$
       end for
15:    To draw a sample from $P\left(X_{1:n+1} | Y_{1:n+1}\right)$. We do resampling with weights $\{w_{j,n+1}\}_{j=1}^{N}\}$ if resampling was performed at step 10. Otherwise, we do resampling with weights $k_{j,n+1} \propto \tilde{w}_{j,n+1} g_{j,n+1}$.
16: end for
```

- **Scenario A**: The process is triggered by 1745 infectious and the times of their infections, $\mathcal{H}_0$ , are uniformly allocated in 3 weeks ($[0, 21)$) with a day being the time unit. We generate weekly latent and observed cases according to the model equations 5.2-5.6 for weeks 1-20 ($[21, 161)$) given $\mathcal{H}_0$, $v = 0.014$, $d = 14.44$, $\beta = 0.5$ and $R_1 = 1.79$. We are interested in inferring the latent cases in weeks $4 - 19$ with $\mathcal{H}_0$ being the set of times of latent infections in weeks 1-3. Using the generated observed cases in weeks 2-4, we estimate the latent infections in weeks 1-3 as follows: The latent cases in the week $i$ are equal to the number of observed events in the week $(i + 1)$ times $1/\beta$, and are spread uniformly in $[(i-1)\times 7+21, i\times 7+21)$ for $1 \leq i \leq 3$. We assume $\alpha = 1$, $b = 2$, $d_{min} = 10$, $d_{max} = 20$, $v_{min} = 0.0001$ and $v_{max} = 0.5$. The ground truth is characterized by $\mathcal{H}_0$ consisting of 4855 seeds, while the estimated seeds are 4228. The observed cases in weeks 4-20 are 17540 (see Figure 5.2).

- **Scenario B**: The process is triggered by 1176 seeds and generated as described above. We assume $d = 15.28$, $v = 0.001$, $\beta = 0.5$, $R_1 = 1.51$, $d_{min} = 10$, $d_{max} = 20$, $v_{min} = 0.001$, $v_{max} = 0.5$, $\alpha = 1$ and $b = 2$. The observed cases in weeks 4-20 are 15448 (see Figure 5.2).

Appendix B contains a third scenario, in which we consider daily latent and reported infections. Here, we focus on Scenarios A and B.

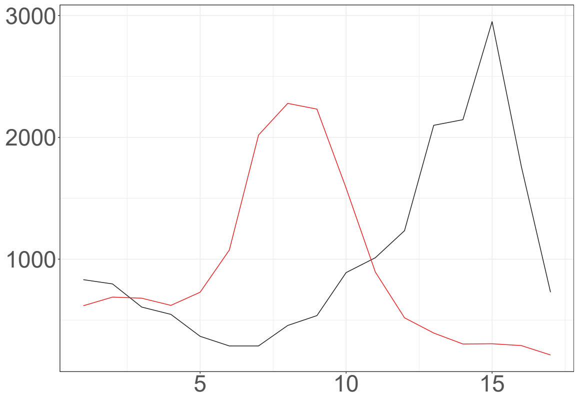
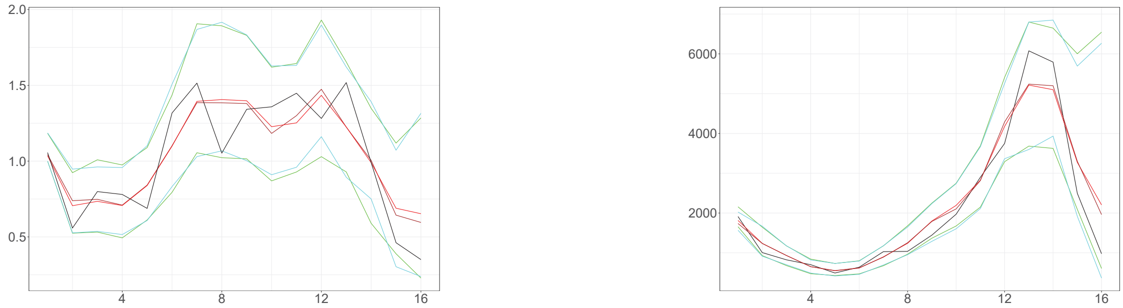


Figure 5.2: Weekly Observed Data (Scenario A (black line); Scenario B (red line)) plotted against time.

We deal with 16 hidden states $\{X_n\}_{n=1}^{16}$. Each state $X_n$ is associated with the latent cases falling during the week $\mathcal{T}_n$ and the parameter $R_n$ associated with that week. We infer the latent intensity $\lambda^N(t)$ and the weights $\{R_n\}_{n=1}^{16}$ as well as the weekly latent cases via the particle sample derived by drawing samples from the smoothing density with lag equal to 4.

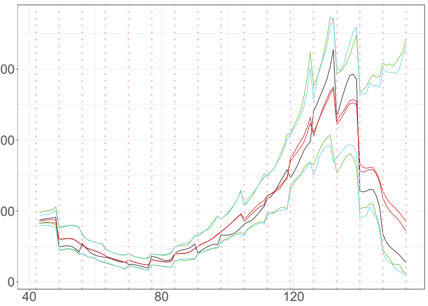
Figures 5.3 - 5.4 illustrate the estimated latent intensity, the estimated weekly hidden cases and the estimated weights of the reproduction number for both scenarios using 40000 particles. We note that the 99% Credible Intervals (CIs) of the time-constant parameters include the actual values of the parameters (see Figure 5.5). Figure 5.6 shows the Effective Sample Size (ESS) for both scenarios using

40000 particles. The simulation analysis shows that the KDPF approaches well the ground truth.



(a) $\{R_n\}_{n=1}^{16}$

(b) Weekly latent cases.



(c) Intensity of latent cases.

Figure 5.3: The true (black line) and the estimated weighs $\{R_n\}_{n=1}^{16}$, weekly latent cases and latent intensity (with estimated seeds (posterior median (brown line) ; 99% CI (cyan line)), and true seeds (posterior median (red line) ; 99% CI (green line))) for Scenario A plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.

To confirm the convergence of posterior estimates of weights and weekly hidden cases concerning the number of particles ($N$), we find the associated Monte Carlo Standard Errors (MCSEs) that give a sense of the variability of particle mean per state. The MCSEs of the average of posterior means of weights and weekly latent cases are given by

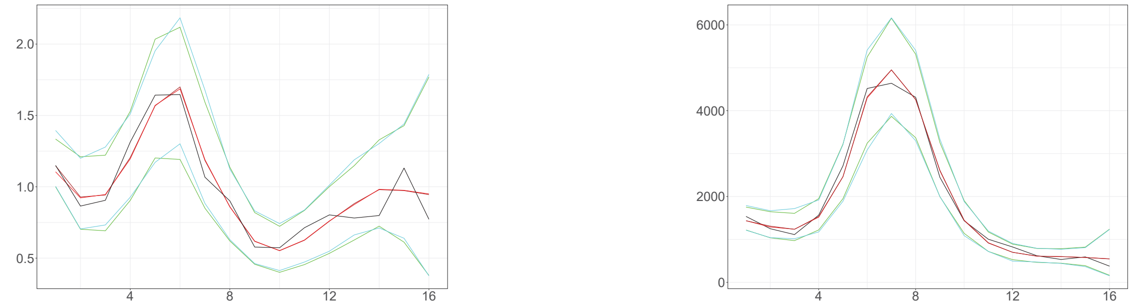
$$\text{MCSE}(R) = \frac{1}{16}\sum_{i=1}^{16}\left(\frac{\text{var}(R_i)}{N}\right)^{1/2}$$

and

$$\text{MCSE}(Y) = \frac{1}{16}\sum_{i=1}^{16}\left(\frac{\text{var}(Y_i)}{N}\right)^{1/2}$$

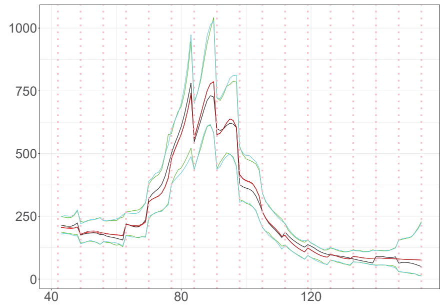
where $\text{var}(z)$ is the variance of $z$ and $Y_i$ the aggregate latent cases in $i_{th}$ week. The MCSE verifies the convergence of posterior estimates concerning the number of particles (see Tables 5.1 - 5.2).

Finally, we compare the performance of the KDPF (Algorithm 9), APF (Algorithm 10), bootstrap filter (BF) (Algorithm 11) and particle marginal Metropolis-Hastings (PMMH) (Algorithm 12) [Andrieu et al., 2010] for inferring the latent



(a) $\{R_n\}_{n=1}^{16}$

(b) Weekly latent cases.



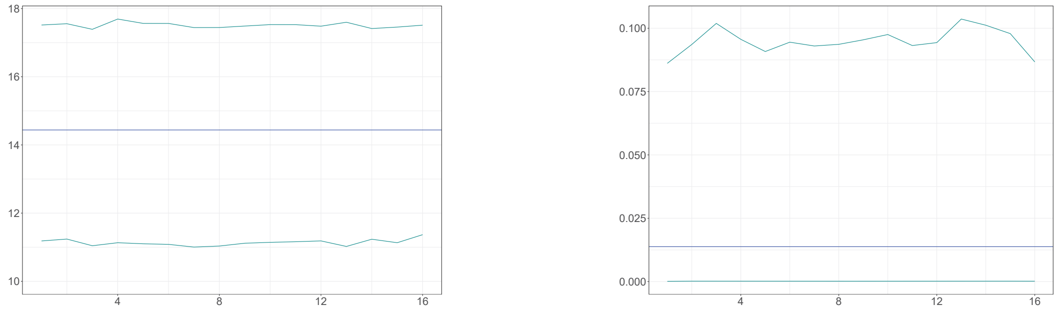
(c) Intensity of latent cases.

Figure 5.4: The true (black line) and the estimated weighs $\{R_n\}_{n=1}^{16}$, weekly latent cases and latent intensity (with estimated seeds (posterior median (brown line) ; 99% CI (cyan line)), and true seeds (posterior median (red line) ; 99% CI (green line))) for Scenario B plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.

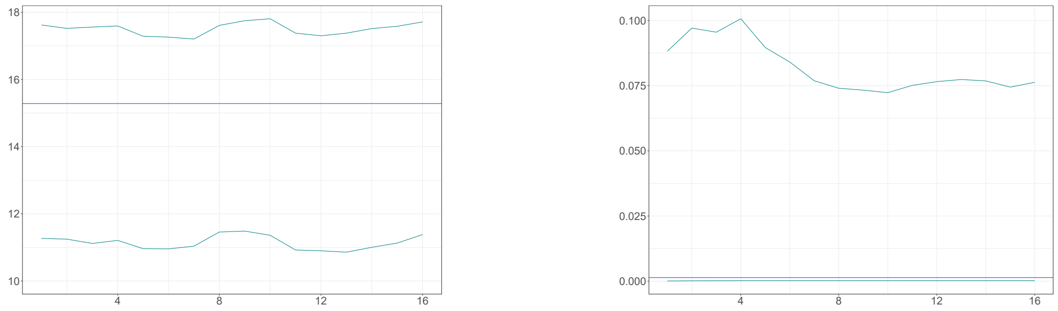
intensity $\lambda^N(t)$ and the reproduction number $R(t)$ illustrated in a new simulation scenario C (see Figure 5.7). Scenario C concerns a process triggered by 661 infectious and generated similar to scenario A assuming that $\alpha = 0.5$, $b = 2$, $d = 15.11$, $v = 0.01$, $R_1 = 1.57$, $d_{min} = 10$, $d_{max} = 20$, $v_{min} = 0.0001$ and $v_{max} = 0.5$. The time-constant parameters $d$ and $v$ are known for BF and APF. We used 10000 iterations of the PMMH sampler with a burn-in of 5000 iterations. We use APF using 50 particles as an SMC sampler. The average acceptance ratio is about 0.1844 resulting in a Markov chain that mixes well. For the KDPF, $\Delta$ was set to 0.99.

| Convergence of the posterior estimates | | | |
|---|---|---|---|
| MCSE | $N = 20000$ | $N = 30000$ | $N = 40000$ |
| $R$ | 0.000626 | 0.000662 | 0.000668 |
| $Y$ | 1.617828 | 1.554482 | 1.529521 |

Table 5.1: MCSEs of posterior means of weights and weekly hidden cases in Scenario A.

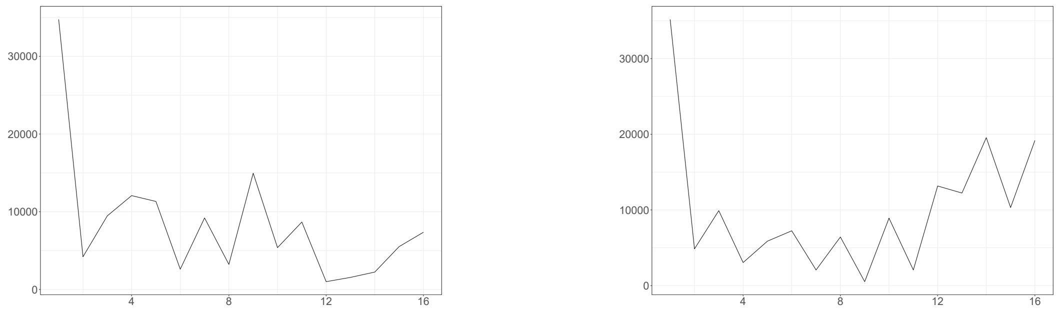


(a) d for Scenario A. (b) v for Scenario A.



(c) d for Scenario B. (d) v for Scenario B.

Figure 5.5: The 99% CIs of time-constant parameters for both scenarios plotted against subintervals.



(a) Scenario A. (b) Scenario B.

Figure 5.6: The Effective Sample Size (ESS) for both scenarios plotted against subintervals.

We find the Average Absolute Error (AAE) of the computed estimates:

$$\text{AAE}(\hat{\lambda}^N) = \frac{1}{N_t} \sum_{i=1}^{N_t} |\hat{\lambda}^N(x_i) - \lambda^N(x_i)|$$

$$\text{AAE}(\hat{R}) = \frac{1}{16} \sum_{i=1}^{16} |\hat{R}_i - R_i|$$

| Convergence of the posterior estimates | | | |
|---|---|---|---|
| MCSE | $N = 20000$ | $N = 30000$ | $N = 40000$ |
| $R$ | 0.000637 | 0.000655 | 0.000629 |
| $Y$ | 0.997832 | 0.993364 | 0.978517 |

Table 5.2: MCSEs of posterior means of weights and weekly hidden cases in Scenario B.

| Scenario C | | | | |
|---|---|---|---|---|
| Filter | AAE($\hat{R}$) | RMSE($\hat{R}$) | AAE($\hat{\lambda}^N$) | RMSE($\hat{\lambda}^N$) |
| KDPF | 0.16 | 0.22 | 25.19 | 73.04 |
| APF | 0.16 | 0.21 | 24.79 | 71.56 |
| BF | 0.15 | 0.21 | 24.7 | 71.43 |
| PMMH | 0.16 | 0.22 | 24.87 | 71.55 |

Table 5.3: Average Absolute Error and Root Mean Square Error for the reproduction number and latent intensity in scenario C.

and the Root Mean Square Error (RMSE):

$$\text{RMSE}(\hat{\lambda}^N) = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\Big(\hat{\lambda}^N(x_i) - \lambda^N(x_i)\Big)^2}$$

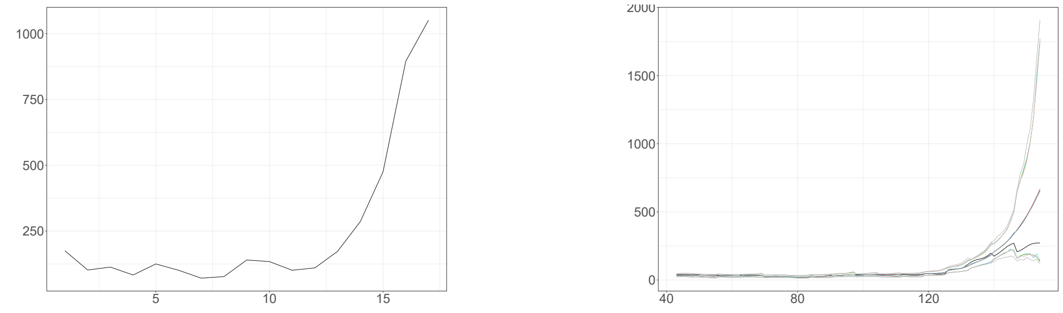
$$\text{RMSE}(\hat{R}) = \sqrt{\frac{1}{16}\sum_{i=1}^{16}(\hat{R}_i - R_i)^2},$$

where $N_t$ is the number of test time points $x_i$ randomly chosen in the time-horizon we consider, $\hat{\lambda}^N(x_i)$ and $\lambda^N(x_i)$ the estimated via posterior median and true intensity at time $x_i$, $\hat{R}_i$ and $R_i$ the estimated via posterior median and true reproduction number in the $i_{th}$ week.

Table 5.3 shows the errors related to KDPF, APF, BF and PMMH for scenario C. The errors associated with KDPF are comparable to those obtained using BF and APF for which the time-constant parameters are known. The performance of KDPF compares well with PMMH, having the advantage that it is a more computationally efficient algorithm than PMMH.

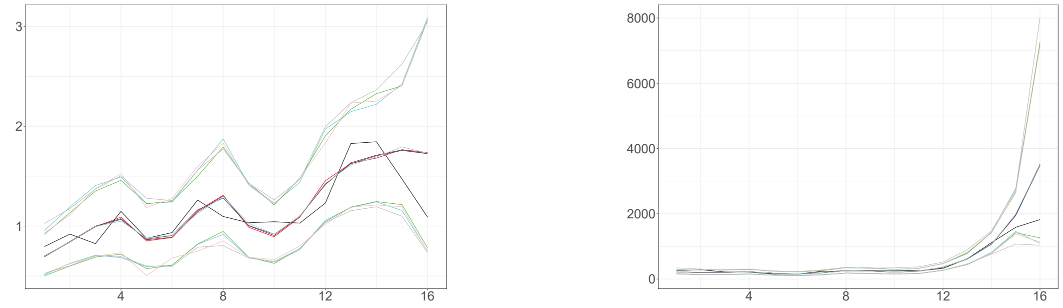
## 5.5 Real Data

We apply the KDPF (Algorithm 9) to real cases in the local authorities: Leicester (4/9/2021 - 24/12/2021) [GOV.UK, 2022d], Kingston upon Thames (11/12/2021 - 8/4/2022) [GOV.UK, 2022c] and Ashford (19/12/2021 - 9/4/2022) GOV.UK [2022a]. Figure 5.8 illustrates the daily and weekly observed cases in the local authorities. We deal with 16 hidden states $\{X_n\}_{n=1}^{16}$ and 16 subintervals $\{\mathcal{T}_n\}_{n=1}^{16}$; each subinterval corresponds to the duration of one week. We infer the latent intensity $\lambda^N(t)$,



(a) The weekly observed cases.

(b) The estimated intensity of latent cases.



(c) The estimated weights $\{R_n\}_{n=1}^{16}$.

(d) The estimated weekly hidden cases.

Figure 5.7: The weekly observed cases, the estimated intensity, the estimated reproduction number, the estimated weekly hidden cases using KDPF( median (red line) ; 99% CI (cyan line)), using APF (median (brown line); 99% CI (green dashed lines)), using PMMH (median (aquamarine line); 99% CI (grey dashed lines)), using BF ( median (blue line); 99% CI (pink dashed lines)) and the true values (black line) in scenario C plotted against time.

**Algorithm 10** Auxiliary particle filter

```
Sample $N$ particles $\{X_{j1}\}_{j=1}^{N}$, $X_{j1} = (R_{j1}, S_{j1}^{N})$:
  for $j$ in $1 : N$ do
    $R_{j1} \sim \mu(R_1)$
    $S_{j1}^{N} \sim P(S_1^N | R_{j1}, \mathcal{H}_0)$
  end for
Find the weights, $\tilde{w}_1 = \{\tilde{w}_{j1}\}_{j=1}^{N}$:
  for $j$ in $1 : N$ do
    $\tilde{w}_{j1} = P(Y_1 | S_{j1}^{N}, \beta, \mathcal{H}_0, v)$
  end for
Normalize the weights, $w_1 = \{w_{j1}\}_{j=1}^{N}$:
  for $j$ in $1 : N$ do
    $w_{j1} = \frac{\tilde{w}_{j1}}{\sum_{j=1}^{N} \tilde{w}_{j1}}$
  end for
Estimate $P(Y_1) : \hat{P}(Y_1) = \frac{1}{N} \sum_{j=1}^{N} \tilde{w}_{j1}$
for $n = 1, .., k$ do
    For each particle $j$, we calculate an estimate of $X_{j,n+1}$ called $\tilde{X}_{j,n+1}$ by drawing a sample from $P(X_{n+1} | X_n, \mathcal{H}_0, d)$:
      for $j$ in $1 : N$ do
        $\tilde{R}_{j,n+1} \sim P(R_{n+1} | R_{jn}, d)$
        $\tilde{S}_{j,n+1}^{N} \sim P(S_{n+1}^{N} | S_{j,1:n}^{N}, \tilde{R}_{j,n+1}, \mathcal{H}_0)$
      end for
    Find the auxiliary weights, $\tilde{g}_{n+1} = \{\tilde{g}_{j,n+1}\}_{j=1}^{N}$:
      for $j$ in $1 : N$ do
        $\tilde{g}_{j,n+1} = g_{jn} w_{jn} P(Y_{n+1} | S_{j,1:n}^{N}, \tilde{S}_{j,n+1}^{N}, \beta, \mathcal{H}_0, v)$
      end for
    Normalize the auxiliary weights, $g_{n+1} = \{g_{j,n+1}\}_{j=1}^{N}$:
      for $j$ in $1 : N$ do
        $g_{j,n+1} = \frac{\tilde{g}_{j,n+1}}{\sum_{j=1}^{N} \tilde{g}_{j,n+1}}$
      end for
    Resample and form $N$ equally weighted particles, $\bar{X}_{1:n} = \{\bar{X}_{1:n}^{i}\}_{i=1}^{N}$:
      for $j$ in $1 : N$ do
        (i) sample index $i_j$ from a multinomial distribution with probabilities $g_{n+1}$
        (ii) $\bar{X}_{1:n}^{j} = X_{i_j,1:n}$
        (iii) $g_{j,n+1} = 1$
      end for
    end if
    Using $\bar{X}_{1:n}$ propagate:
      for $j$ in $1 : N$ do
        $R_{j,n+1} \sim P(R_{n+1} | R_{jn}, d)$
        $S_{j,n+1}^{N} \sim P(S_{n+1}^{N} | S_{j,1:n}^{N}, R_{j,n+1}, \mathcal{H}_0)$
        Set $X_{j,1:n+1} = (\bar{X}_{1:n}^{j}, (R_{j,n+1}, S_{j,n+1}^{N}))$
      end for
    Find the weights, $\tilde{w}_{n+1} = \{\tilde{w}_{j,n+1}\}_{j=1}^{N}$:
      for $j$ in $1 : N$ do
        $\tilde{w}_{j,n+1} = \frac{P(Y_{n+1} | S_{j,1:n+1}^{N}, \beta, \mathcal{H}_0, v)}{P(Y_{n+1} | S_{i_j,1:n}^{N}, \tilde{S}_{i_j,n+1}^{N}, \beta, \mathcal{H}_0, v)}$
      end for
    Normalize the weights, $w_{n+1} = \{w_{j,n+1}\}_{j=1}^{N}$:
      for $j$ in $1 : N$ do
        $w_{j,n+1} = \frac{\tilde{w}_{j,n+1}}{\sum_{j=1}^{N} \tilde{w}_{j,n+1}}$
      end for
    Estimate $P(Y_{n+1} | Y_{1:n}) : \hat{P}(Y_{n+1} | Y_{1:n}) = \Big( \sum_{i=1}^{N} w_{i,n} P(Y_{n+1} | S_{i,1:n}^{N}, \tilde{S}_{i,n+1}^{N}, \beta, \mathcal{H}_0, v) \Big) \Big( \frac{1}{N} \sum_{i=1}^{N} \tilde{w}_{i,n+1} \Big)$.
end for
Estimate the marginal likelihood $P(Y_{1:k})$: $\hat{P}(Y_{1:k}) = \hat{P}(Y_1) \prod_{n=2}^{k} \hat{P}(Y_n | Y_{1:n-1})$.
```

**Algorithm 11** Bootstrap filter

```
1: Sample $N$ particles $\{X_{j1}\}_{j=1}^{N}$, $X_{j1} = (R_{j1}, S_{j1}^{N})$:
     for $j$ in $1 : N$ do
       $R_{j1} \sim \mu(R_1)$
       $S_{j1}^{N} \sim P(S_1^N | R_{j1}, \mathcal{H}_0)$
     end for
2: Find the weights, $\tilde{w}_1 = \{\tilde{w}_{j1}\}_{j=1}^{N}$:
     for $j$ in $1 : N$ do
       $\tilde{w}_{j1} = P(Y_1 | S_{j1}^{N}, \beta, \mathcal{H}_0, v)$
     end for
3: Normalize the weights, $w_1 = \{w_{j1}\}_{j=1}^{N}$:
     for $j$ in $1 : N$ do
       $w_{j1} = \frac{\tilde{w}_{j1}}{\sum\limits_{j=1}^{N} \tilde{w}_{j1}}$
     end for
4: Estimate $P(Y_1) : \hat{P}(Y_1) = \frac{1}{N} \sum\limits_{j=1}^{N} \tilde{w}_{j1}$
5: for $n = 2, .., k$ do
6:     Resample and form $N$ equally weighted particles, $\bar{X}_{1:n-1} = \{\bar{X}_{1:n-1}^{i}\}_{i=1}^{N}$:
         for $j$ in $1 : N$ do
           (i) sample index $i_j$ from a multinomial distribution with probabilities $w_{n-1}$
           (ii) $\bar{X}_{1:n-1}^{j} = X_{i_j,1:n-1}$
           (iii) $w_{j,n-1} = 1$
         end for
       end if
7:     Using $\bar{X}_{1:n-1}$ propagate:
         for $j$ in $1 : N$ do
           $R_{j,n} \sim P(R_n | R_{j,n-1}, d)$
           $S_{j,n}^{N} \sim P(S_n^N | S_{j,1:n-1}^{N}, R_{j,n}, \mathcal{H}_0)$
           Set $X_{j,1:n} = (\bar{X}_{1:n-1}^{j}, (R_{j,n}, S_{j,n}^{N}))$
         end for
8:     Find the weights, $\tilde{w}_n = \{\tilde{w}_{j,n}\}_{j=1}^{N}$:
         for $j$ in $1 : N$ do
           $\tilde{w}_{j,n} = w_{j,n-1} P(Y_n | S_{j,1:n}^{N}, \beta, \mathcal{H}_0, v)$
         end for
9:     Normalize the weights, $w_n = \{w_{j,n}\}_{j=1}^{N}$:
         for $j$ in $1 : N$ do
           $w_{j,n} = \frac{\tilde{w}_{j,n}}{\sum\limits_{j=1}^{N} \tilde{w}_{j,n}}$
         end for
10:    Estimate $P(Y_n | Y_{1:n-1}) : \hat{P}(Y_n | Y_{1:n-1}) = \frac{1}{N} \sum\limits_{j=1}^{N} \tilde{w}_{j,n}$.
11: end for
12: Resample and form $N$ equally weighted particles, $\bar{X}_k = \{\bar{X}_k^i\}_{i=1}^{N}$:
      for $j$ in $1 : N$ do
        (i) sample index $i_j$ from a multinomial distribution with probabilities $w_k$
        (ii) $\bar{X}_k^j = X_{i_j,k}$
        (iii) $w_{j,k} = 1$
      end for
    end if
13: Estimate the marginal likelihood $P(Y_{1:k})$: $\hat{P}(Y_{1:k}) = \hat{P}(Y_1) \prod\limits_{n=2}^{k} \hat{P}(Y_n | Y_{1:n-1})$.
```

**Algorithm 12** Particle marginal Metropolis-Hastings sampler

```
Step 1: Initialization,$i = 0$,
    (a) $\log d_0 \sim \mathcal{N}(\mu_d, \sigma_d^2)$, $\log v_0 \sim \mathcal{N}(\mu_v, \sigma_v^2)$
    (b) - run a SMC algorithm targetting $P(x_{1:n}|Y_{1:n}, \beta, \mathcal{H}_0, d_0, v_0)$
        -Sample $X_{1:n}(0) \sim \hat{P}(\cdot|Y_{1:n}, \beta, \mathcal{H}_0, d_0, v_0)$
        -Let $\hat{P}(Y_{1:n}|d_0, v_0)$ denote the marginal likelihood estimate.
Step 2: for iteration $i \geq 1$
    (a) $\log d^* \sim \mathcal{N}(\log d_{i-1}, \sigma_{d*})$, $\sigma_{d*} = \frac{|\log d_{i-1} - \log d_{min}|}{4}$,
        $\log v^* \sim \mathcal{N}(\log v_{i-1}, \sigma_{v*})$, $\sigma_{v*} = \frac{|\log v_{i-1} - \log v_{min}|}{4}$
    (b) - run a SMC algorithm targetting $P(x_{1:n}|Y_{1:n}, \beta, \mathcal{H}_0, d^*, v^*)$
        -Sample $X^*_{1:n} \sim \hat{P}(\cdot|Y_{1:n}, \beta, \mathcal{H}_0, d^*, v^*)$
        -Let $\hat{P}(Y_{1:n}|d^*, v^*)$ denote the marginal likelihood estimate.
    (c) wp min $\left(1, \frac{\hat{P}(Y_{1:n}|d^*, v^*)}{\hat{P}(Y_{1:n}|d_{i-1}, v_{i-1})} \frac{P(d^*)}{P(d_{i-1})} \frac{P(v^*)}{P(v_{i-1})} \frac{P(d_{i-1}|d^*)}{P(d^*|d_{i-1})} \frac{P(v_{i-1}|v^*)}{P(v^*|v_{i-1})}\right)$
        set $d_i = d^*$, $v_i = v^*$, $X_{1:n}(i) = X^*_{1:n}$, $\hat{P}(Y_{1:n}|d_i, v_i) = \hat{P}(Y_{1:n}|d^*, v^*)$. Otherwise,
    $d_i = d_{i-1}$, $v_i = v_{i-1}$, $X_{1:n}(i) = X_{1:n}(i-1)$, $\hat{P}(Y_{1:n}|d_i, v_i) = \hat{P}(Y_{1:n}|d_{i-1}, v_{i-1})$.
```

the reproduction number $R(t)$, and the weekly and daily latent cases via the particle sample derived by drawing samples from the smoothing density with lag equal to 4. We demonstrate that the proposed model can be applied to predict the new observed cases in near future.

We assume that the initial reproduction number during the first week is uniformly distributed over the interval from 0.5 to 2. Our initialization takes into account the 90% Confidence Interval published from the government in the UK: 0.9-1.1 on 4/9/2021 and 11/12/2021, 1-1.2 on 19/12/2021 [GOV.UK, 2022b]. We also assume $d_{min} = 1$, $d_{max} = 10$, $v_{min} = 0.0001$ and $v_{max} = 0.5$.

Figures 5.9 - 5.12 show the estimated latent and observed intensity, the estimated weekly and daily hidden cases, the estimated reproduction number, the time-constant parameters, and the ESS in the local authorities. We illustrate the intensity of observed cases, approximating via equation 5.1. We note that the estimated latent intensity and the estimated weekly latent cases are in agreement with the weekly observed cases. According to the analysis, the instantaneous reproduction number $R(t)$ depicts the pandemic's development and capture alterations. For the COVID-19 pandemic, there is a maximum delay of 21 days between the reported and actual infection times, which provides information regarding the progression of the epidemic. As a result, estimates have become more uncertain towards the end of the horizon.

To assess the performance of our algorithm, we compute the mean absolute percentage error (MAPE) of the computed estimate of weekly observed cases (see Algorithm 13) :

$$\text{MAPE}(\hat{Y}) = \frac{1}{16} \sum_{i=1}^{16} \frac{|\hat{Y}_i - Y_i|}{Y_i},$$

where $Y_i$ and $\hat{Y}_i$ are the true and estimated weekly observed cases via the posterior median in week $i$, respectively. The metric is 1.46%, 1.08% and 2.09% for Ashford, Leicester and Kingston upon Thames. Figure 5.13 shows the estimated weekly ob-

**Algorithm 13** Estimating the weekly observed cases

```
for $n = 1, .., 16$ do
    for $j = 1, .., N$ do
        Calculate the mean of observed cases in the interval $\mathcal{T}_n$ denoted by $\mu_{j,n}$.
        $Y_{j,n} \sim \text{NB}(\mu_{j,n}, v_{jn})$.
    end for
    Use the sample $\{Y_{j,n}\}_{j=1}^{N}$ to find the posterior median and the 95% CI of the estimated observed cases in $\mathcal{T}_n$.
end for
```

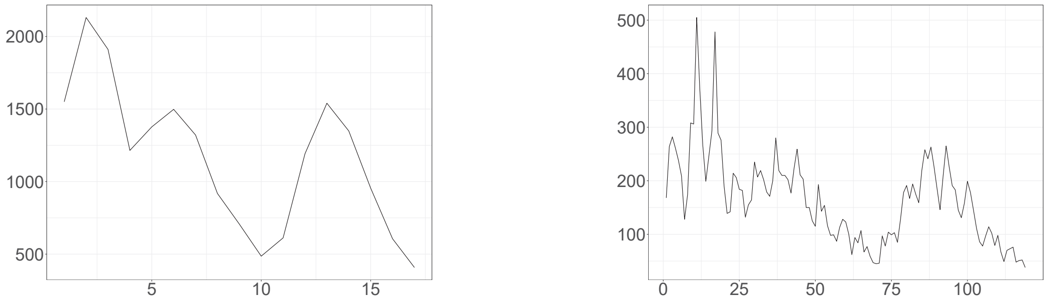
served cases for the local authorities. The analysis demonstrates that our algorithm provides a good approximation of the weekly reported cases.

We compare the proposed algorithm with two methods of estimating the reproduction number. The method suggested by Cori et al. [2013] (EpiEstim) estimates the reproduction number from incidence time series using a Bayesian framework with a gamma distributed prior imposed on the reproduction number. An alternative method suggested by Koyama et al. [2021] is a state-space method for estimating the daily reproduction number from a time series of reported infections using a random walk prior to the reproduction number and log-normal distribution as the distribution of the serial interval (SI). We assume that the mean and standard deviation of the SI distribution is at 6.9 days and 5.6 days following Zhao et al. [2021]. We apply EpiEstim by using the gamma and log-normal distribution as the distribution of SI. Both choices lead to identical results.

Figure 5.14 shows the weekly average of daily estimates of the reproduction number via posterior median derived by the method of Koyama et al. [2021] and the posterior medians of $R(t)$ given by EpiEstim and the proposed algorithm following the course of the pandemic. The method of Koyama et al. [2021] and EpiEstim provide similar estimates to those of Algorithm 9 most of the time. Koyama et al. [2021] and EpiEstim do not build a delay between reported and actual infection time in their models, which is why there are variations in their estimations compared to our algorithm. Therefore, the reproduction number given by EpiEstim responds later to changes compared to our estimation. Koyama et al. [2021] shows a bit less of a time lag, which we conjecture to be due to it working with daily reproduction numbers and cases (which are being shown averaged in Figure 5.14). In the first week, the estimates of Koyama et al. [2021] and EpiEstim have essential higher values than one of the proposed algorithm due to different assumptions about the initialization of the epidemic.

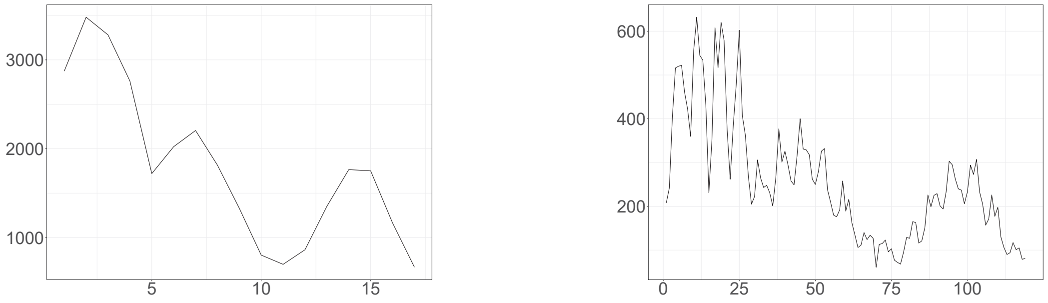
We also compare the estimated rate of latent cases $\lambda^N(t)$ and observed cases $\lambda^M(t)$ with the estimated daily number of events derived by Koyama et al. [2021]. Figure 5.15 shows that the expected daily number of events is almost identical to $\lambda^M(t)$ and in agreement with $\lambda^N(t)$ after the end of the 3rd week. The differences in the first three weeks are due to different initializations of the methods.

**Forecasting** Using the proposed model, it is also possible to predict the number of new observed cases in the near future, by fitting the model with data up to week $\mathcal{T}_k$ and forecasting cases for the week $\mathcal{T}_{k+1}$ using Algorithm 14. To analyse the performance of this algorithm, we conduct a rolling-window analysis and predict the observed cases in weeks $\mathcal{T}_{11} - \mathcal{T}_{17}$. Table 5.4 shows the estimated numbers in the local authorities by applying Algorithm 14 and the method introduced by Koyama



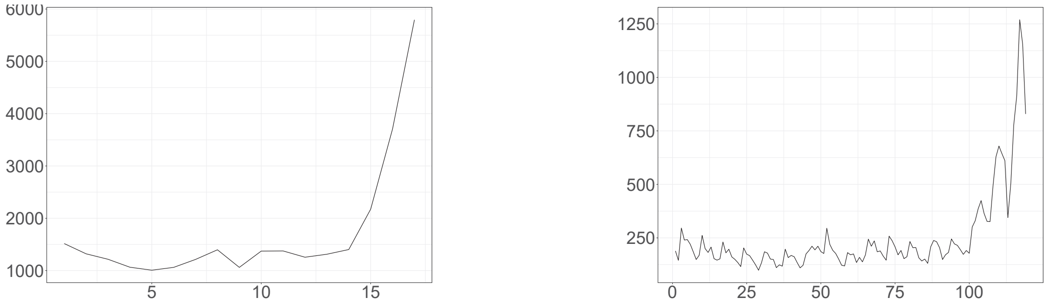
(a) The weekly observed cases in Ashford.

(b) The daily observed cases in Ashford.



(c) The weekly observed cases in Kingston upon Thames.

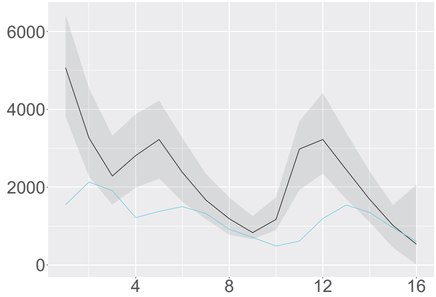
(d) The daily observed cases in Kingston upon Thames.

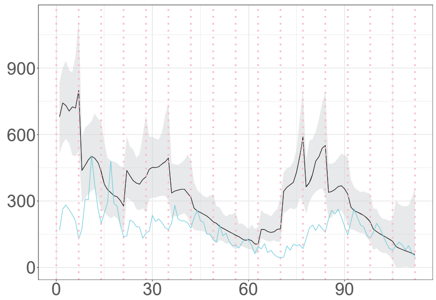


(e) The weekly observed cases in Leicester.

(f) The daily observed cases in Leicester.

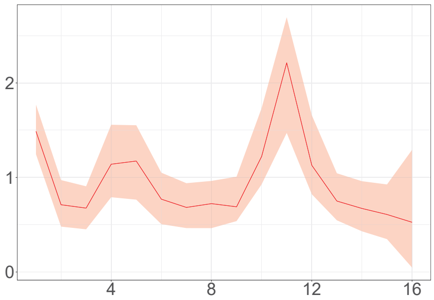
Figure 5.8: The daily and weekly observed infections in local authorities plotted against time.

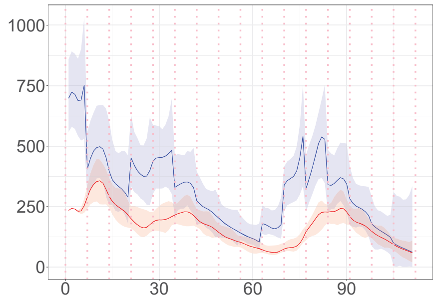




(a) The estimated weekly latent cases (posterior median (black line) ; 99% CI (ribbon)), and the weekly observed cases (cyan line).

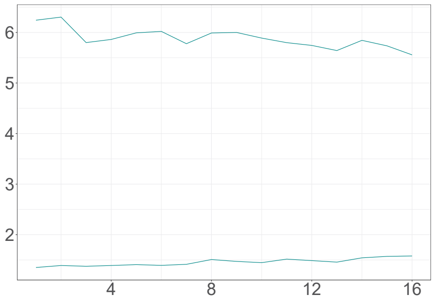
(b) The estimated daily latent cases (posterior median (black line) ; 99% CI (ribbon)), and the daily observed cases (cyan line).

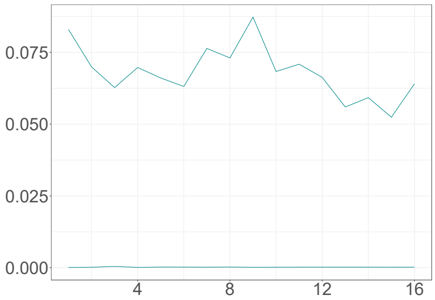




(c) The estimated reproduction number (posterior median (red line); 95% CI (ribbon)).

(d) The estimated intensity of latent cases (posterior median (blue line) ; 99% CI (ribbon)) and the estimated intensity of observed cases (posterior median (red line) ; 99% CI (ribbon)).





(e) The 99% CI of $d$.

(f) The 99% CI of $v$.

Figure 5.9: The weekly and daily latent cases, the reproduction number, the latent and observed intensity and the 99% CIs of time-constant parameters in Ashford plotted against time. The time interval between two successive pink vertical dashed lines corresponds to a week.

et al. [2021], assuming that the reproduction number remains at the value obtained for the last day. Table 5.5 shows the metrics MAPE and AAE of the estimated cases via posterior median. The empirical coverage probability of our 80% CIs is about 86%. Our estimates are similar to those given by Koyama et al. [2021] most of the time.

| **Authority** | **Week** | **Mean** | **Median** | 80% **CIs** | **True** | **Koyama et al.** |
|---|---|---|---|---|---|---|
| Ashford | $\mathcal{T}_{11}$ | 334 | 301 | (149, 452) | 612 | 355 |
| | $\mathcal{T}_{12}$ | 902 | 748 | (266, 1196) | 1192 | 899 |
| | $\mathcal{T}_{13}$ | 2568 | 2317 | (821,3548) | 1540 | 1959 |
| | $\mathcal{T}_{14}$ | 1934 | 1711 | (756, 2650) | 1349 | 1228 |
| | $\mathcal{T}_{15}$ | 1018 | 917 | (465, 1351) | 955 | 1018 |
| | $\mathcal{T}_{16}$ | 626 | 563 | (287, 835) | 606 | 600 |
| | $\mathcal{T}_{17}$ | 375 | 337 | (164, 502) | 408 | 419 |
| Leicester | $\mathcal{T}_{11}$ | 1551 | 1347 | (527, 2171) | 1376 | 1488 |
| | $\mathcal{T}_{12}$ | 1437 | 1264 | (585, 1950) | 1256 | 1393 |
| | $\mathcal{T}_{13}$ | 1095 | 1000 | (508,1480) | 1313 | 1153 |
| | $\mathcal{T}_{14}$ | 1682 | 1551 | (653, 2339) | 1403 | 1333 |
| | $\mathcal{T}_{15}$ | 1505 | 1356 | (606, 2022) | 2174 | 1461 |
| | $\mathcal{T}_{16}$ | 3478 | 3094 | (1253, 4834) | 3708 | 3563 |
| | $\mathcal{T}_{17}$ | 6641 | 6134 | (2699, 9233) | 5794 | 1692 |
| Kingston | $\mathcal{T}_{11}$ | 478 | 424 | (164, 661) | 699 | 624 |
| | $\mathcal{T}_{12}$ | 638 | 571 | (253, 868) | 863 | 504 |
| | $\mathcal{T}_{13}$ | 1148 | 1046 | (441,1586) | 1352 | 1151 |
| | $\mathcal{T}_{14}$ | 2102 | 1865 | (719, 2895) | 1765 | 1959 |
| | $\mathcal{T}_{15}$ | 2298 | 2001 | (871, 3082) | 1751 | 2095 |
| | $\mathcal{T}_{16}$ | 1643 | 1472 | (707,2265) | 1165 | 1530 |
| | $\mathcal{T}_{17}$ | 693 | 624 | (320, 931) | 667 | 722 |

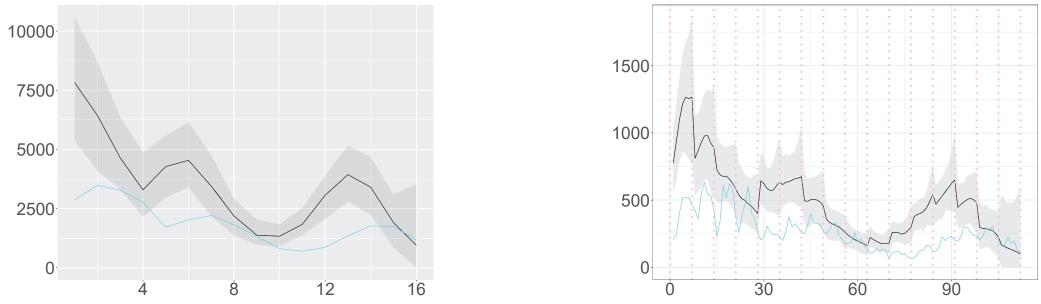
Table 5.4: The true number of infections in $\mathcal{T}_{11} - \mathcal{T}_{17}$, the posterior median, the posterior mean and the 80% CIs of the new infections projected by Algorithm 14, and the predicted infections given by the method of Koyama et al. [2021] using a rolling-window analysis in local authorities.

| **Authority** | **MAPE(Alg. 14)** | **AAE(Alg. 14)** | **MAPE(Koyama)** | **AAE(Koyama)** |
|---|---|---|---|---|
| Kingston | 0.21 | 224.7 | 0.20 | 227.6 |
| Leicester | 0.14 | 324.3 | 0.21 | 777 |
| Ashford | 0.28 | 292.3 | 0.1 | 167.1 |

Table 5.5: The MAPE and AAE of estimated observed cases over next week projected by Algorithm 14 and the method of Koyama et al. [2021] using a rolling-window analysis in local authorities .

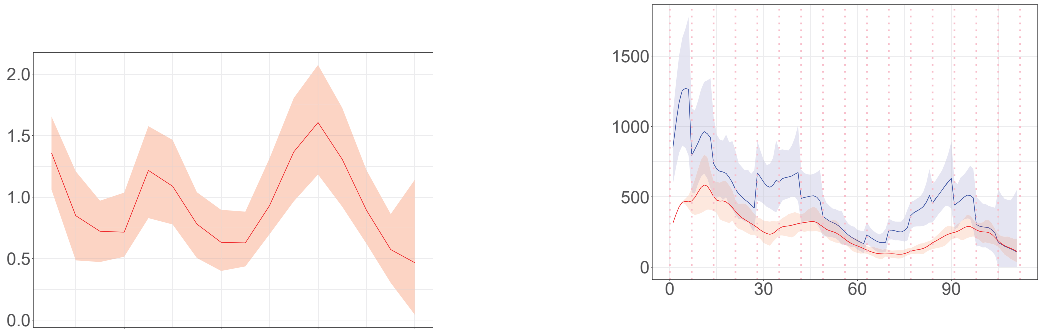
**Algorithm 14** Predicting the new observed cases in week $\mathcal{T}_{k+1}$

```
for $j = 1, .., N$ do
    $R_{j,k+1} \sim P(R_{k+1} | R_{j,k}, d_{jk})$
    $S^N_{j,k+1} \sim P(S^N_{k+1} | S^N_{j,1:k}, R_{j,k+1}, \mathcal{H}_0)$
    Calculate the mean of observed cases in the interval denoted by $\mu_{j,k+1}$.
    $Y_{j,k+1} \sim \mathrm{NB}(\mu_{j,k+1}, v_{jk})$.
end for
Use the sample $\{Y_{j,k+1}\}_{j=1}^N$ to find the posterior mean, the posterior median and the 80% CI of the estimated observed cases in $\mathcal{T}_{k+1}$.
```



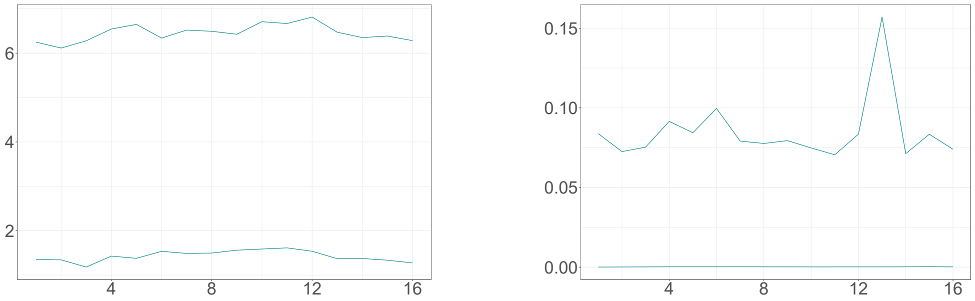
(a) The estimated weekly latent cases (posterior median (black line) ; 99% CI (ribbon)), and the weekly observed cases (cyan line).

(b) The estimated daily latent cases (posterior median (black line) ; 99% CI (ribbon)), and the daily observed cases (cyan line).



(c) The estimated reproduction number (posterior median (red line); 95% CI (ribbon)).

(d) The estimated intensity of latent cases (posterior median (blue line) ; 99% CI (ribbon)) and the estimated intensity of observed cases (posterior median (red line) ; 99% CI (ribbon)).



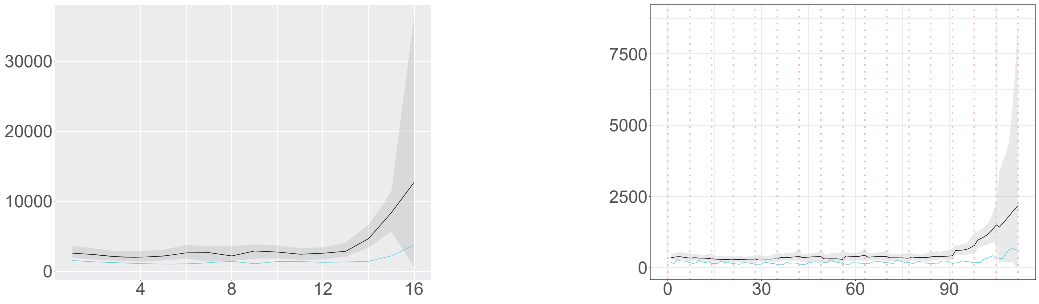
(e) The 99% CI of $d$.

(f) The 99% CI of $v$.

Figure 5.10: The weekly and daily latent cases, the reproduction number, the latent and observed intensity and the 99% CIs of time-constant parameters in Kingston upon Thames plotted against time. The time interval between two successive pink vertical dashed lines corresponds to a week.

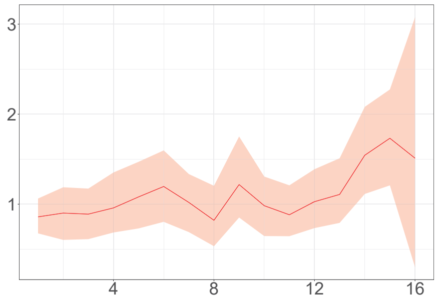
## 5.6 Discussion

In this chapter, we introduce a novel epidemic model using a latent Hawkes process with temporal covariates. Unlike other Hawkes models, we model the infections via a Hawkes process and the aggregated reported cases via a probability distribution $G$ with a mean driven by the underlying Hawkes process. The usual options of $G$ are

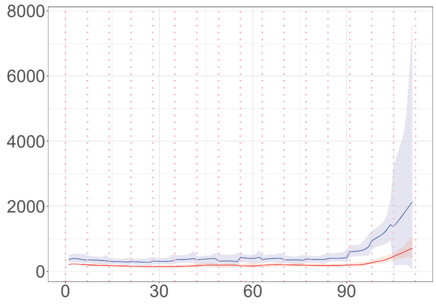


(a) The estimated weekly latent cases (posterior median (black line) ; 99% CI (ribbon)), and the weekly observed cases (cyan line).

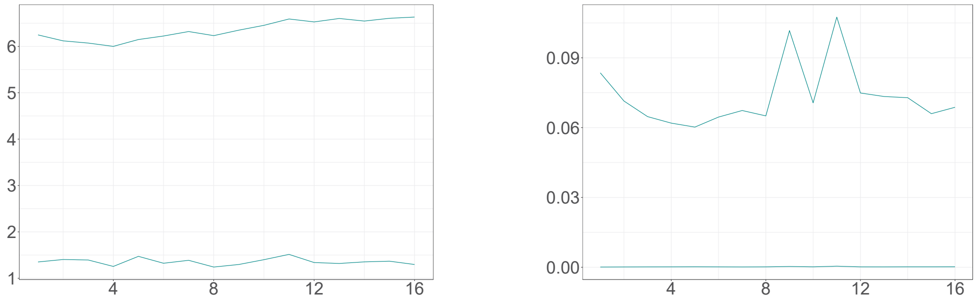
(b) The estimated daily latent cases (posterior median (black line) ; 99% CI (ribbon)), and the daily observed cases (cyan line).



(c) The estimated reproduction number (posterior median (red line); 95% CI (ribbon)).



(d) The estimated intensity of latent cases (posterior median (blue line) ; 99% CI (ribbon)) and the estimated intensity of observed cases (posterior median (red line) ; 99% CI (ribbon)).



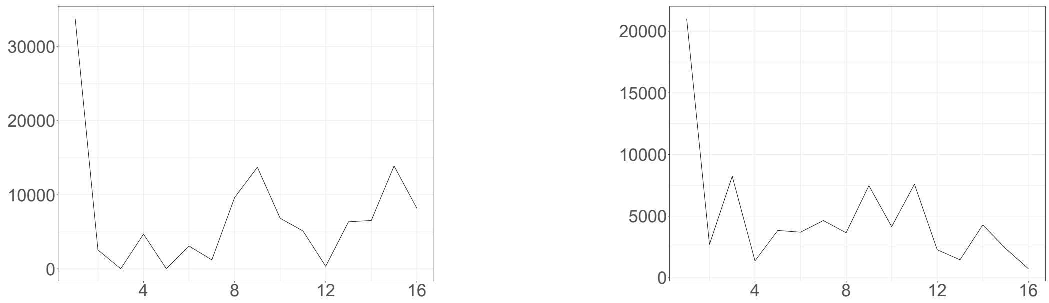
(e) The 99% CI of $d$.

(f) The 99% CI of $v$.

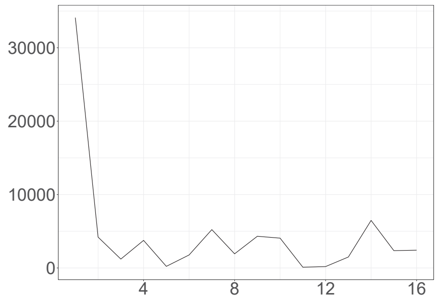
Figure 5.11: The weekly and daily latent cases, the reproduction number, the latent and observed intensity and the 99% CIs of time-constant parameters in Leicester plotted against time. The time interval between two successive pink vertical dashed lines corresponds to a week.

Negative Binomial and Poisson distribution. We propose a KDPF for inferring the latent cases and the instantaneous reproduction number and for predicting the new observed cases in the near future. We demonstrate the performance of the proposed algorithm on COVID-19.

The analysis of synthetic data shows that KDPF compares well with PMMH, having the advantage that it is a more computationally efficient algorithm than

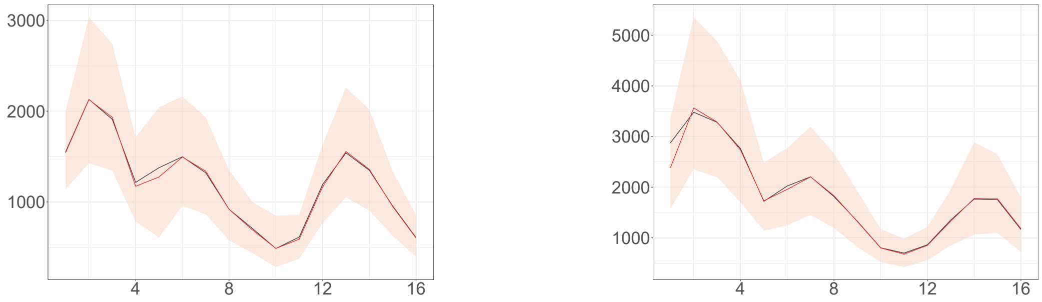


(a) Ashford (b) Kingston upon Thames

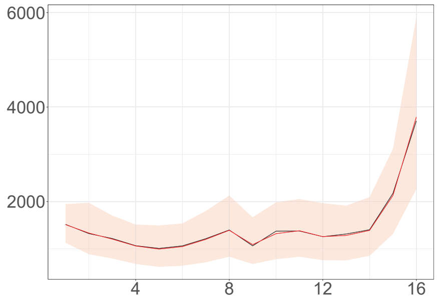


(c) Leicester

Figure 5.12: The Effective Sample Size (ESS) in local authorities plotted against subintervals.

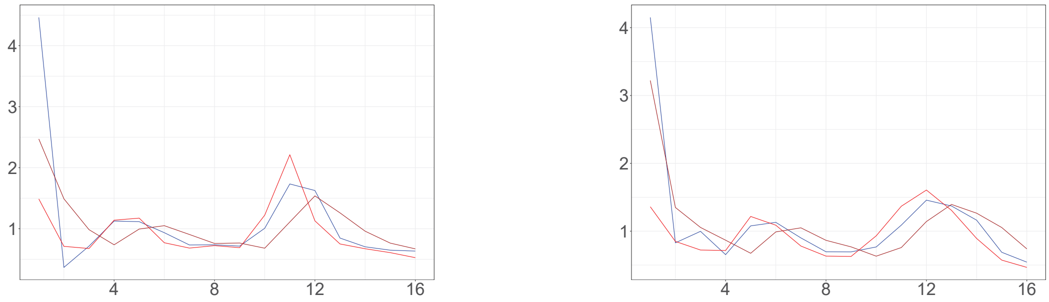


(a) Ashford (b) Kingston upon Thames



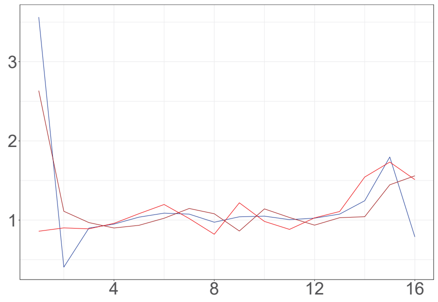
(c) Leicester

Figure 5.13: The estimated weekly observed cases (posterior median (red line) ; 99% CI (ribbon)) and the actual cases (black line) plotted against time.



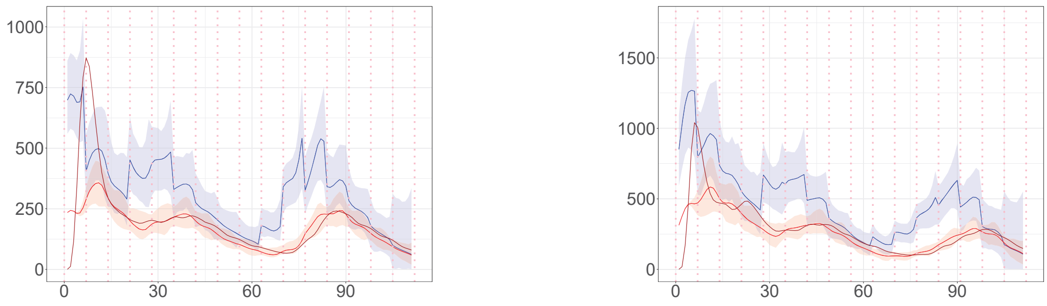
(a) Ashford

(b) Kingston upon Thames

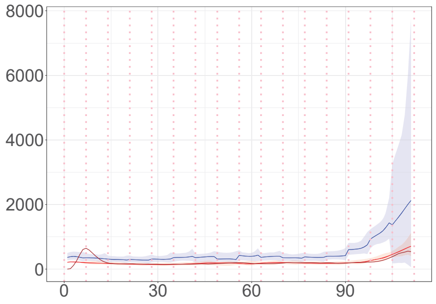


(c) Leicester

Figure 5.14: The weekly average of daily estimates of the reproduction number via posterior median derived by the method of Koyama et al. [2021] (blue line) and the posterior medians of $R(t)$ given by EpiEstim (brown line) and our proposed algorithm (red line) plotted against time.



(a) Ashford (b) Kingston upon Thames



(c) Leicester

Figure 5.15: The estimated daily number of events derived by the method of Koyama et al. [2021] (brown line), the estimated latent intensity (posterior median (blue line) ; 99% CI (ribbon)) and the estimated intensity of observed cases (posterior median (red line) ; 99% CI (ribbon)) plotted against time.

PMMH. We also demonstrate that our predicted new cases, and our inference for the latent intensity, the daily and weekly hidden cases are consistent with the observed cases in various local authorities in the UK. The simulation analysis shows that the proposed algorithm provides comparable estimates of the observed case fluctuations compared to the method introduced by Koyama et al. [2021]. The method of Koyama et al. [2021] and EpiEstim provide similar estimates of the reproduction number to the proposed algorithm.

The simulation analysis shows that working with daily reported infections leads to better ESS using a smaller number of particles, as the data spikes are reduced. According to Cori et al. [2013], the estimates of the instantaneous reproduction number are expected to be affected by the selection of the time window size. Large sizes result in more smoothing and reductions in statistical noise, whereas small sizes result in faster detection of transmission changes and more statistical noise. Cori et al. [2013] suggests an appropriate way of choosing the time window size. We have selected a weekly time window to analyse the real data in line with Cori et al. [2013].

Uncovering disease dynamics and tracing how and by whom an infected individual was infected is challenging due to unobservable transmission routes [Yang and Zha, 2013; Kim et al., 2019]. Modelling the infections via a Hawkes process allows us to model infection dynamics.

Isham and Medley [1996]; Wallinga et al. [1999]; Farrington et al. [2001] contend that it is necessary to account for individual heterogeneities while modelling the transmission of an infectious disease. Individuals vary in their tendency to interact with others; personal hygiene is a key factor in the propagation of diseases; individuals' community structure and location might be significant in spreading epidemics. The simplest assumption of individual heterogeneity is that individual sociability varies with the age.

The proposed epidemic model can be viewed as a new approach in deriving epidemic models that consider individual heterogeneities and provide insight into underlying dynamics. In the next chapter, we extend the proposed model by taking into account a finite population and the individual heterogeneity in age groups and using a multidimensional Hawkes process for modelling the infection cases.

# Chapter 6

# Age-stratified epidemic model using a marked latent Hawkes process

Modelling the spread of an infectious disease must take into account the mechanism of its transmission, the individual heterogeneities, and the nature and duration of interactions among the population [Isham and Medley, 1996; Wallinga et al., 1999; Farrington et al., 2001].

Cevik and Baral [2021] emphasize the importance of considering individual and network heterogeneity in modelling and setting specific mitigation strategies for respiratory pandemics. The Public Health England group [PHE-Group, 2020], Cevik et al. [2021]; Lemieux et al. [2021] indicate the connection of COVID-19 risk with socioeconomic inequalities, including occupation, crowded housing, job insecurity and poverty. Long-term-care facilities such as nursing homes and homeless shelters saw an increased infection rate [Cevik et al., 2021; Lemieux et al., 2021]. People working in public-facing occupations classified as social workers, people working in hospitality and people facing the greatest socioeconomic deprivation had a higher risk of getting infected [Cevik et al., 2021; PHE-Group, 2020].

There are several proposed epidemic approaches built on basic compartmental models that take into account population heterogeneity usually focusing on a population characteristic (e.g., age) and either dividing the population into several constant-size groups (see, e.g., Gart [1968]) or considering the population as having a continuous distribution (see, e.g., Novozhilov [2008]). Moreno et al. [2002] considered the susceptible-infected-recovered (SIR) model on complex heterogeneous networks with predefined connectivity distribution to understand the epidemics in complex networks; They showed that the large connectivity fluctuations strengthen the incidence of epidemic outbreaks. Novozhilov [2008] formulated a SIR model with an explicit form of a nonlinear transmission function modelling the heterogeneity. Gomes et al. [2022], considering the basic susceptible-exposed-infected-recovered (SEIR) model and continuous distribution of individual susceptibility or exposure to infection, showed that “individual variation in susceptibility and exposure can affect basic metrics such as herd immunity threshold (HIT)”. Tkachenko et al. [2021] adopted a generalized version of the heterogeneous well-mixed theory in the spirit of Moreno et al. [2002], grouping individuals into subpopulations with similar social and biological heterogeneity in their susceptibilities to infection, and suggested the significant role of heterogeneity in finding basic epidemic metrics, such as HIT, and modelling the epidemic dynamics using data for the COVID-19.

Diseases are often spread via social contacts. This induces that the rate at which a disease is spread is dependent on the number of contacts between infectious and susceptibles. Empirical studies (e.g. Mossong et al. [2008]; Leung et al. [2017]; de Waroux et al. [2018]; Béraud et al. [2015]) quantified via matrices, known as contact matrices, the contact patterns relevant for infections transmitted by the respiratory or close-contact route in several countries.

Individuals vary in their tendency to interact with others; personal hygiene is a key factor in the propagation of diseases; individuals' community structure and location might be significant in spreading epidemics. The simplest assumption of individual heterogeneity is to consider that contact rates vary with only one characteristic of an individual, such as age. Age as behavioural and physiological factor is highly correlated to the risk of infection in many diseases like influenza-like diseases [Eames et al., 2012; Apolloni et al., 2013; Luca et al., 2018; Worby et al., 2015], pertussis [Rohani et al., 2010], tuberculosis [Arregui et al., 2018; Guzzetta et al., 2011], varicella [Marangi et al., 2017] and COVID-19 [Davies et al., 2020]. Understanding the impact of age on the transmission of disease is critical for determining and implementing social-distancing and interventions, especially closing schools [Cauchemez et al., 2008; Hens et al., 2009; Davies et al., 2020].

Wallinga et al. [2006] showed that school-aged children and young adults are more likely to get infected and contribute most to the spread of infection due to their high number of contacts. The 65 study participants of Edmunds et al. [1997] showed that the mean age of contacts increased with the age of participants, and older participants ($\geq$ 40 years) had more contacts with older adults and a larger variability in the age of their contacts than younger participants ($<$ 40 years).

Balabdaoui and Mohr [2020] have proposed a compartmental model to capture the dynamic of highly age-sensitive epidemics and evaluate the effect of social contact patterns on a load of hospitals and their intensive care units. Stocks et al. [2020] have introduced a model selection process using transmission models, subdividing the population into age classes. Pellis et al. [2020] have also suggested a mathematical approach to select between age and household structure in designing a model for an initial, rapid assessment of potential epidemic severity.

The epidemic model introduced in Chapter 5 can be viewed as an alternative approach in deriving epidemic models that consider individual heterogeneities and provide insight into underlying dynamics. Here, we extend that model by considering a finite population and the individual heterogeneity in age groups by using a multidimensional Hawkes process for modelling the infections. The aggregated reported cases per age group, in turn, have a probability distribution with a mean driven by the underlying Hawkes process. We apply a KDPF for inferring the latent infections and the instantaneous reproduction number for each age group and for forecasting the epidemic's trajectory in the near future. We demonstrate the performance of the proposed model on COVID-19 data in several London boroughs published by the government in the UK using the empirical contact matrix derived from Jarvis et al. [2020] within the framework of the latent multidimensional Hawkes process.

The outline of the chapter is as follows. Section 6.1 introduces the novel age-stratified epidemic model. Section 6.2 discusses the proposed inference algorithm. Sections 6.3 and 6.4 are applications of the proposed algorithm on synthetic and COVID-19 data.

## 6.1 Model

We introduce a novel age-stratified epidemic model by extending the epidemic model 5.2-5.6 considering a finite population stratified by a finite set of age bands, denoted by $\mathcal{A}$.

The epidemic process is seen as a marked counting process $N(t)$ with a set of jump times $\mathcal{T}^N = \{t_0 < t_1 < t_2 < \ldots\}$ and a set of associated age groups, $A^N = \{a_i\}$ where $a_i \in \mathcal{A}$ is the age group of infection at time $t_i$. The intensity of latent infections in age group $a \in \mathcal{A}$ at time $t$ is given by

$$\lambda^N(t,a) = \frac{S_{t,a}}{N_a}\gamma(t,a) \sum_{t_i \in h_t^0} h(t - t_i)\mathrm{m_{aa_i}}$$

for $t > T_0$ with $S_{t,a}$ being the number of susceptibles in age group $a$ at time $t$, $\gamma(t,a)$ being a process dependent on age group $a$, $N_a$ being the size of the population of age group $a$ and $\mathrm{m_{aa_i}}$ being the average number of contacts per unit time of a person in age group $a$ with people in age group $a_i$. The process $\gamma(t,a)$ conveys information about the biological susceptibility of the age group $a$ and offsets the potential erroneous average number of contacts at time $t$.

Similar to Chapter 5, we model the observed cases of age $a$ falling in $\mathcal{T}_n$ denoted by $Y_{na}$ via a probability distribution $G$ having mean $\mu_{na}$ equal to the expected observed cases of age $a$ in $\mathcal{T}_n$ given by

$$\mu_{na} = \beta \sum_{t_w \in [T_0, T_n), a_w \equiv a} \int_{\max(t_w, T_{n-1})}^{T_n} g(s - t_w) ds.$$

We model $\gamma(t,a)$ as a stepwise function having as many weights as the number of subintervals, that is

$$\gamma(t,a) = \prod_{n=1}^{k} \gamma_{na}^{\mathbb{1}(t \in \mathcal{T}_n)},$$

where $\{\gamma_{na}\}_{n=1}^k$ is assumed to be a Markov process. The average number of secondary cases of age $a' \in \mathcal{A}$ each infected individual of age $a$ would infect is given by

$$R_{aa'}(t) = \frac{S_{t,a'}}{N_{a'}}\gamma(t,a')\mathrm{m_{a'a}}$$

if the conditions remained as they were at $t$. The instantaneous reproduction number of age group $a$ is

$$R_a(t) = \sum_{a'} R_{aa'}(t),$$

which is the number of newly infected people that each infected individual aged $a$ would infect if the conditions remained as they were at $t$. We derive the instantaneous reproduction number $R(t)$ as the infected population-weighted average of $R_a(t)$ at time $t$.

The age-stratified model is described by the equations:

$$\lambda^N(t,a) = \frac{S_{t,a}}{N_a}\gamma(t,a)\sum_{t_i \in h_t^0} h(t-t_i)\mathrm{m}_{\mathrm{aa_i}},\ \mathrm{t} \in [\mathrm{T}_0, \mathrm{T}),\ \mathrm{a} \in \mathcal{A} \tag{6.1}$$

$$Y_{na} \sim \text{G with mean } E(Y_{na}) = \mu_{na},\ n = 1,..,k,\ a \in \mathcal{A} \tag{6.2}$$

$$\gamma(t,a) = \prod_{n=1}^{k} \gamma_{na}^{\mathbb{1}\{t \in \mathcal{T}_n\}},\ a \in \mathcal{A} \tag{6.3}$$

$$\{\gamma_{na}\}_{n=1}^{k} \text{is a Markov process},\ a \in \mathcal{A} \tag{6.4}$$

$$\mu_{na} = \beta \sum_{t_w \in [T_0, T_n), a_w \equiv a} \int_{\max(t_w, T_{n-1})}^{T_n} g(s-t_w)ds\ ,\ n = 1, \ldots, k,\ a \in \mathcal{A}. \tag{6.5}$$

## 6.2 Inference algorithm

Given a set of observed infections with their associated age groups, we seek to infer the marked counting process $N(t)$ and the processes $\{\gamma(t,a)\}_a$.

Following the inference approach of Section 5.4, the proposed epidemic model described by equations 6.1-6.5 is seen as a state-space model with a latent state process $\{X_n : 1 \leq n \leq k\}$ and an observed process $\{Y_n = (Y_{na})_a : 1 \leq n \leq k\}$. Each hidden state $X_n$ consists of the weights $\{\gamma_{na}\}_a$ associated to $\mathcal{T}_n$ and the set of latent cases $S_n^N$ falling into $\mathcal{T}_n$ along with their associated age groups $A_n^N$. The time-constant parameters are the parameters associated with the distribution $G$ and the prior imposed on the weights $\{\{\gamma_{na}\}_{n=1}^{k}\}_a$. We apply the KDPF (Algorithm 16) for inferring the marked counting process $N(t)$, the weights $\{\{\gamma_{na}\}_{n=1}^{k}\}_a$, and the time-constant parameters.

Similar to Chapter 5, we focus on illustrating the performance of the model on COVID-19. We model the observed cases $Y_{na}$ via a negative binomial distribution (NB) with mean $\mu_{na}$ and dispersion $v > 0$. The transition kernels are defined in Section 5.3. Before we proceed with the simulation analysis, we define the prior on weights $\{\{\gamma_{na}\}_{n=1}^{k}\}_a$, an algorithm for sampling the latent cases, the complexity of the Algorithm 16 and a crude initialization of $\mathcal{H}_0$.

**Set of infectious at the beginning of the process, $\mathcal{H}_0$** We adopt a heuristic approach to initialize $\mathcal{H}_0$. The transition kernel of latent cases illustrated in Figure 5.1 shows that a latent case at $t_w$ can influence the latent intensity at $t$ if $t_w$ has occurred at most 21 days before $t$. Otherwise, the influence of $t_w$ is negligible. Therefore, as the history of the process, we consider the latent cases of 21 days/3 weeks before the beginning of the process. The transition kernel of observed cases shown in Figure 5.1 demonstrates that an event is most likely to be observed seven days after the actual infection time. Considering the observed cases are daily, we initialize the history of latent cases in age group $a$, $\mathcal{H}_{0a}$ by uniformly spreading on the day $-i$ the number of cases of age $a$ occurred on the day $(-i+7)$ times $1/\beta$. The times of their infections at the beginning of the process, $\mathcal{H}_0$ is given by the union of the sets $\mathcal{H}_{0a}$, that is, $\mathcal{H}_0 = \cup_a \mathcal{H}_{0a}$. We denote by $A_0^N$ the age groups associated with the time infections in $\mathcal{H}_0$. In simulation analysis, we propose initialization of $\mathcal{H}_0$ when we deal with weekly reported cases.

**Imposed prior on weights $\{\{\boldsymbol{\gamma_{na}}\}_{\mathbf{n=1}}^{\mathbf{k}}\}_a$** A geometric random walk (RW) is imposed as prior on the weights $\{\{\gamma_{na}\}_{n=1}^{k}\}_a$ :

$$\begin{aligned}\log \gamma_{na} &= \log \gamma_{n-1,a} + \log \epsilon_{na}, \quad \epsilon_{na} \sim \text{Gamma}(d,d),\ n = 2,..,k,\ a \in \mathcal{A}\\ \gamma_{1a} &\sim \text{Uniform}(\alpha, b),\ a \in \mathcal{A}.\end{aligned}$$

We assume that the noise of RW $\epsilon_{na}$ has a gamma distribution with equal shape and rate at $d$. This induces that the time-varying number $\gamma_{na}$ is gamma distributed with a mean equal to $\gamma_{n-1,a}$ and standard deviation $\gamma_{n-1,a}/\sqrt{d}$. The stronger fluctuations in the observed data, the more flexible modelling we need. Smaller values of $d$ have higher standard deviation and lead to a wider range of possible values of $\gamma_{n-1,a}$ increasing the flexibility of the model.

**Sampling the latent cases and associated age groups** We sample the latent cases $S_n^N$ along with their associated age groups $A_n^N$ falling into the subinterval $\mathcal{T}_n$ by applying Algorithm 15, which is a simulation procedure based on the branching structure of the Hawkes process [Laub et al., 2015]. The proposed algorithm is a superposition of Poisson processes in the interval $\mathcal{T}_n$; the descendants of each latent event at $t_i$ form an inhomogeneous Poisson process with intensity

$$\lambda_i(t) = h(t - t_i) \sum_a \gamma_{na} \frac{S_{t,a}}{N_a} \mathrm{m}_{\mathrm{aa_i}}$$

for $t > t_i$ and $t \in [T_{n-1}, T_n)$. This induces that:

- The number of events $n_i$ triggered by an event at $t_i$ in the interval $\mathcal{T}_n$ is Poisson distributed with parameter

$$\lambda = \left( \sum_a \frac{\gamma_{na} S_{\max(T_{n-1}, t_i), a}}{N_a} \mathrm{m}_{\mathrm{aa_i}} \right) \int\limits_{\max(t_i, T_{n-1})}^{T_n} h(s - t_i) ds.$$

- The arrival times of the $n_i$ descendants are $t_i + E_i$ with $E_i$ being iid random variables with pdf the truncated distribution $h(t)$ in $[\max(t_i, T_{n-1}), T_n)$.

- Sample the associated age groups w.p. $P(a) = \frac{\frac{S_{\max(T_{n-1},t_i),a}}{N_a} \mathrm{m}_{\mathrm{aa_i}}}{\sum\limits_a \frac{S_{\max(T_{n-1},t_i),a}}{N_a} \mathrm{m}_{\mathrm{aa_i}}}$, $\forall a \in \mathcal{A}$.

**Algorithm 15** Sample $S_n^N$, $A_n^N | S_{1:(n-1)}^N$, $\mathcal{H}_0$, $A_0^N$, $\{\gamma_{na}\}_a$

```
1: Input: $S_{1:(n-1)}^N$, $A_{1:(n-1)}^N$, $\mathcal{H}_0$, $A_0^N$ , $\{\gamma_{na}\}_a$
2: Initialize two empty queues: $Q_t$ and $Q_a$.
3: $Q_t = \{S_v^N\}_{v=n-\eta}^{n-1}$ and $Q_a = \{A_v^N\}_{v=n-\eta}^{n-1}$ with $\eta$ being the number of former subintervals we consider (the value of $\eta$ is determined by the transition kernel of latent cases).
4: while $Q_t$ is not empty do
5:     Remove the first element $t_i$ from $Q_t$ and the first element $a_i$ from $Q_a$.
6:     Draw the number of events $n_i$ triggered by an event $i$ from a Poisson distribution with parameter $\lambda = \left( \sum_a \frac{\gamma_{na} S_{\max(T_{n-1},t_i),a}}{N_a} \mathrm{m}_{\mathrm{aa_i}} \right) \int_{\max(t_i,T_{n-1})}^{T_n} h(s-t_i)ds$ that is the average number of offsprings generated by an event at $t_i$ in $\mathcal{T}_n$.
7:     Generate $n_i$ events from the truncated distribution $h(t)$ in $[\max(t_i, T_{n-1}), T_n)$, and add the new elements to $S_n^N$ and the back of queue $Q_t$.
8:     Sample the associated age groups w.p. $P(a) = \frac{\frac{S_{\max(T_{n-1},t_i),a}}{N_a} \mathrm{m}_{\mathrm{aa_i}}}{\sum_a \frac{S_{\max(T_{n-1},t_i),a}}{N_a} \mathrm{m}_{\mathrm{aa_i}}}$, and add the asociated age groups to $A_n^N$ and the back of queue $Q_a$.
9:     For each age group $a \in \mathcal{A}$, decrease the estimated number of susceptibles aged $a$ falling into interval $\mathcal{T}_n$, $S_{\max(T_{n-1},t_i),a}$, by the number of new infections aged $a$ generated by the infection event at $t_i$ in $\mathcal{T}_n$.
10: end while
11: Return $S_n^N$ and $A_n^N$.
```

The computational cost of Algorithm 15 is $O\left( |\mathcal{A}| \sum_{v=n-\eta}^{n} |S_v^N| \right)$ with $\eta$ being the number of former subintervals that influence the latent cases falling into $\mathcal{T}_n$ determined by the transition kernel of latent cases. Considering a finite population stratified by age bands $\mathcal{A}$, the cost of Algorithm 15 for sampling the latent cases falling into the subinterval $\mathcal{T}_n$ increases compared to the cost of Algorithm 8, that is equal to $O\left( \sum_{v=n-\eta}^{n} |S_v^N| \right)$, due to additional steps 8 and 9.

Accounting for an infinite population and each person's age band, step 9 is not required and the probabilities $\{P(a)\}_{a \in \mathcal{A}}$ are calculated once over all states leading to the computational cost $O\left( \sum_{v=n-\eta}^{n} |S_v^N| \right)$, which is also the cost of sampling latent cases in an unstructured homogeneously mixing epidemic model (Algorithm 8).

**Who infected whom** The proposed model can capture the process' branching pattern by saving the parent of each latent infection at step 7 of Algorithm 15. Alternatively, the parent of each infection $j$ aged $a$ falling into $\mathcal{T}_z$ is assumed to be sampled from a multinomial distribution parameterized by $\pi_{ja}$, where $\pi_{ja} =$

$\{\pi_{jia}\}_{i\in h_j}$ with

$$\pi_{jia} = \frac{h(t_j - t_i)\mathrm{m}_{\mathrm{aa_i}}}{\sum_{t_w \in h_j^P} h(t_j - t_w)\mathrm{m}_{\mathrm{aa_w}}}$$

being the probability of secondary infection $j$ aged $a$ having been caused by primary infection $i$, where $h_j^P = \{t_i | t_i \in \cup_{v=z-\eta}^{z} \mathcal{T}_v,\ t_i < t_j\}$ and $h_j = \{i :\ t_i \in h_j^P\}$.

**Computational complexity** The computational cost of each propagation step (steps 7 and 12 of Algorithm 16) at state (interval) $n$ is equal to the cost of Algorithm 15 times the number of particles ($N$), that is $O\left(N|\mathcal{A}| \sum\limits_{v=n-\eta}^{n} |S_v^N|\right)$. The cost of finding weights (steps 8 and 13 of Algorithm 16) at state (interval) $n$ is $O\left(N \sum\limits_{v=n-\eta}^{n} |S_v^N|\right)$. Hence, the computational cost of Algorithm 16 over all states (intervals) is $O\left(N|\mathcal{A}|\eta|\mathcal{T}^N|\right)$. $S_n^N$ is the set of latent cases falling into subinterval $\mathcal{T}_n$ and $\mathcal{T}^N = \cup_n S_n^N$. The algorithm is easily parallelized over $N$.

Accounting for an infinite population and each person's age band, the computational cost is $O\left(N\eta|\mathcal{T}^N|\right)$, which is also the cost of Algorithm 9 for an unstructured homogeneously mixing epidemic model.

**Model complexity** The set of parameters for inference includes the two time-constant parameters, $d$ governing the variability of the noise in the processes $\gamma(t,a)$, $\forall a \in \mathcal{A}$ and $v$ the dispersion parameter of the observed counts, the latent infections, $\{S_n^N\}_{n=1}^k$, and their associated age groups, $\{A_n^N\}_{n=1}^k$ and the steps of processes $\gamma(t,a)$, $\{\{\gamma_{na}\}_{n=1}^k\}_{a\in\mathcal{A}}$. There is a set of model parameters, including the ascertainment rate $\beta$, the transition kernels of latent and observed cases, which we consider as given. The set of infectious at the beginning of the process, $\mathcal{H}_0$, is described applying the heuristic approach described above. We rely on the Bayesian paradigm for regularizing the parameters for inference.

## 6.3 Simulation Analysis

We carried out a simulation study to illustrate the performance of the KDPF (Algorithm 16) for inferring the weights $\{\{\gamma_{na}\}_{n=1}^k\}_a$, the intensity of latent infections $\lambda^N(t,a)$ and the weekly hidden cases per age group $a$ over various numbers of groups.

The performance of particle filter methods degrades as the model dimension rises [Beskos et al., 2017; Li et al., 2005; Park and Ionides, 2020]. Particle filters can have an exponential cost in the dimensionality of the hidden state to be stable [Beskos et al., 2017]. Increasing the number of age groups raises the dimensionality of the hidden state and, by extension, the computational resources required to keep the uniform-in-time convergence results dependent on recent states [Kantas et al., 2015]. A simple approach to mitigate the dimensionality problem would be to consider the process $\gamma(t,a) = c_a q(t)$, where $c_a$ is a given constant associated with age group $a$ and $q(t)$ a random process independent of age groups. However, that approach requires

**Algorithm 16** Kernel density particle filter

```
1:  Initialize the parameters $\{\theta_{j1}\}_{j=1}^{N}$, $\theta_{j1} = (d_{j1}, v_{j1})$ with $d_{min} \leq d \leq d_{max}$ and $v_{min} \leq v \leq v_{max}$:
      for $j$ in $1:N$ do
        $\log d_{j1} = \mathcal{N}(\mu_d, \sigma_d^2)$ with $\mu_d = \frac{\log d_{max} + \log d_{min}}{2}$ and $\sigma_d = \frac{\log d_{max} - \log d_{min}}{8}$
        $\log v_{j1} = \mathcal{N}(\mu_v, \sigma_v^2)$ with $\mu_v = \frac{\log v_{max} + \log v_{min}}{2}$ and $\sigma_v = \frac{\log v_{max} - \log v_{min}}{8}$
      end for
2:  Sample $N$ particles $\{X_{j1}\}_{j=1}^{N}$, $X_{j1} = \left(\{\gamma_{j1a}\}_a,\ S_{j1}^N,\ A_{j1}^N\right)$:
      for $j$ in $1:N$ do
        $\gamma_{j1a} \sim \text{Uniform}(\alpha, \beta)\ \forall$ age group $a$
        $\left(S_{j1}^N,\ A_{j1}^N\right) \sim P\left(S_1^N, A_1^N | \{\gamma_{j1a}\}_a, \mathcal{H}_0, A_0^N\right)$ using Algorithm 15
      end for
3:  Find the weights, $\tilde{w}_1 = \{\tilde{w}_{j1}\}_{j=1}^{N}$:
      for $j$ in $1:N$ do
        $\tilde{w}_{j1} = P\left(Y_1 | S_{j1}^N, A_{j1}^N, \beta, \mathcal{H}_0, A_0^N, v_{j1}\right) = \prod_a P\left(Y_{1a} | S_{j1a}^N, \beta, \mathcal{H}_0, A_0^N, v_{j1}\right)$
      end for
    where $S_{j1a}^N$ is the set of latent cases of age $a$ in $\mathcal{T}_1^N$ associated to particle $j$.
4:  Normalize the weights, $w_1 = \{w_{j1}\}_{j=1}^{N}$:
      for $j$ in $1:N$ do
        $w_{j1} = \frac{\tilde{w}_{j1}}{\sum_{j=1}^{N} \tilde{w}_{j1}}$
      end for
5:  for $n = 1, .., k$ do
6:        for $j$ in $1:N$ do
            $m_{j,dn}^{(L)} = a \log d_{jn} + (1-a)\bar{d}_{Ln}, \qquad \bar{d}_{Ln} = \sum_{j=1}^{N} w_{jn} \log d_{jn}$
            $m_{j,vn}^{(L)} = a \log v_{jn} + (1-a)\bar{v}_{Ln}, \qquad \bar{v}_{Ln} = \sum_{j=1}^{N} w_{jn} \log v_{jn}$
            $m_{j,dn} = a d_{jn} + (1-a)\bar{d}_n, \qquad \bar{d}_n = \sum_{j=1}^{N} w_{jn} d_{jn}$
            $m_{j,vn} = a v_{jn} + (1-a)\bar{v}_n, \qquad \bar{v}_n = \sum_{j=1}^{N} w_{jn} v_{jn}$
          end for
7:        For each particle $j$, we calculate an estimate of $X_{j,n+1}$ called $\tilde{X}_{j,n+1}$ by drawing a sample from $P(X_{n+1}|X_n, \mathcal{H}_0)$:
          for $j$ in $1:N$ do
            $\tilde{\gamma}_{j,n+1,a} \sim P\left(\gamma_{n+1,a} | \gamma_{jna}, m_{j,dn}\right)\ \forall$ age group $a$
            $\left(\tilde{S}_{j,n+1}^N, \tilde{A}_{j,n+1}^N\right) \sim P\left(S_{n+1}^N, A_{n+1}^N | S_{j,1:n}^N, A_{j,1:n}^N, \tilde{\gamma}_{j,n+1,a}, \mathcal{H}_0, A_0^N\right)$ using Algorithm 15
          end for
8:        Find the auxiliary weights, $\tilde{g}_{n+1} = \{\tilde{g}_{j,n+1}\}_{j=1}^{N}$:
          for $j$ in $1:N$ do
            $\tilde{g}_{j,n+1} = g_{jn} w_{jn} P\left(Y_{n+1} | S_{j,1:n}^N, A_{j,1:n}^N, \tilde{S}_{j,n+1}^N, \tilde{A}_{j,n+1}^N, \beta, \mathcal{H}_0, A_0^N, m_{j,vn}\right)$
                     $= g_{jn} w_{jn} \prod_a P\left(Y_{n+1,a} | \{S_{jka}\}_{k=1}^{n}, \beta, \mathcal{H}_0, A_0^N, m_{j,vn}\right)$
          end for
        where $S_{jka}^N$ is the set of latent cases of age $a$ in $\mathcal{T}_k$ associated to particle $j$.
9:        Normalize the auxiliary weights, $g_{n+1} = \{g_{j,n+1}\}_{j=1}^{N}$:
          for $j$ in $1:N$ do
            $g_{j,n+1} = \frac{\tilde{g}_{j,n+1}}{\sum_{j=1}^{N} \tilde{g}_{j,n+1}}$
          end for
10:       if$\left(ESS(g_{n+1}) = 1/\sum_{j=1}^{N} g_{j,n+1} < 0.8N\right)$ then resample and form $N$ equally weighted particles, $\bar{X}_{1:n} = \{\bar{X}_{1:n}^i\}_{i=1}^{N}$:
          for $j$ in $1:N$ do
            (i) sample index $i_j$ from a multinomial distribution with probabilities $g_{n+1}$
            (ii) $\bar{X}_{1:n}^j = X_{i_j,1:n}$
            (iii) $g_{j,n+1} = 1$
          end for
        end if
11:       Regenerate the fixed parameters:
          for $j$ in $1:N$ do
            $\log v_{j,n+1} \sim \mathcal{N}(m_{i_j,vn}^{(L)}, h^2 V_{nv}^{(L)})$
            $\log d_{j,n+1} \sim \mathcal{N}(m_{i_j,dn}^{(L)}, h^2 V_{nd}^{(L)})$
          end for
        where $V_{nv}^{(L)}$ is the weighted variance of $\{\log v_{jn}\}_{n=1}^{N}$ and $V_{nd}^{(L)}$ the weighted variance of $\{\log d_{jn}\}_{n=1}^{N}$.
12:       Using $\bar{X}_{1:n}$ propagate:
          for $j$ in $1:N$ do
            $\gamma_{j,n+1,a} \sim P(\gamma_{n+1,a} | \gamma_{jna}, d_{j,n+1})$
            $\left(S_{j,n+1}^N, A_{j,n+1}^N\right) \sim P\left(S_{n+1}^N, A_{n+1}^N | S_{j,1:n}^N, A_{j,1:n}^N, \{\gamma_{jna}\}_a, \mathcal{H}_0, A_0^N\right)$ using Algorithm 15
            Set $X_{j,1:n+1} = (\bar{X}_{1:n}^j, \left(\{\gamma_{j,n+1,a}\}_a, S_{j,n+1}^N)\right)$
          end for
13:       Find the weights, $\tilde{w}_{j,n+1} = \{\tilde{w}_{j,n+1}\}_{j=1}^{N}$:
          for $j$ in $1:N$ do
            $\tilde{w}_{j,n+1} = \frac{P\left(Y_{n+1} | S_{j,1:n+1}^N, A_{j,1:n+1}^N, \beta, \mathcal{H}_0, A_0^N, v_{j,n+1}\right)}{P\left(Y_{n+1} | S_{j,1:n}^N, A_{j,1:n}^N, \tilde{S}_{i_j,n+1}^N, \tilde{A}_{i_j,n+1}^N \beta, \mathcal{H}_0, A_0^N, m_{i_j,vn}\right)}$
          end for
14:       Normalize the weights, $w_{n+1} = \{w_{j,n+1}\}_{j=1}^{N}$:
          for $j$ in $1:N$ do
            $w_{j,n+1} = \frac{\tilde{w}_{j,n+1}}{\sum_{j=1}^{N} \tilde{w}_{j,n+1}}$
          end for
15:       To draw a sample from $P\left(X_{1:n+1}, d_{n+1}, v_{n+1} | Y_{1:n+1}\right)$. We do resampling with weights $\{w_{j,n+1}\}_{j=1}^{N}\}$ if resampling was
        performed at step 10. Otherwise, we do resampling with weights $k_{j,n+a} \propto \tilde{w}_{j,n+1} g_{j,n+1}$.
16: end for
```

methods for estimating the constants $\{c_a\}_a$ determining the biological susceptibilities of age groups. An alternative method would be to see $\{c_a\}$ as time-constant parameters and infer them via the KDPF. Appendix C includes simulation concepts considering six and nine age groups with similar biological susceptibilities and independent processes $\{\gamma(t,a)\}_a$ on age groups, that is, $\gamma(t,a) = q(t)$, $\forall$ age group $a \in \mathcal{A}$.

To confirm the convergence of posterior estimates of weights and weekly hidden cases per age group concerning the number of particles, we find the associated Monte Carlo Standard Errors (MCSEs) that give a sense of the average variability of particle mean per state. The MCSEs of the average of posterior means of weights $\gamma_a = \{\gamma_{ia}\}_{i=1}^{k}$, denoted $\overline{\gamma_a}$, and the average of weekly latent cases of age $a$, $Y_a = \{Y_{ia}\}_{i=1}^{k}$, denoted $\overline{Y_a}$ are given by

$$\text{MCSE}(\overline{\gamma_a}) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{\text{var}(\gamma_{ia})}{N}\right)^{1/2}$$

and

$$\text{MCSE}(\overline{Y_a}) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{\text{var}(Y_{ia})}{N}\right)^{1/2}$$

where var$(z)$ is the variance of $z$ and $Y_{ia}$ the aggregate latent cases of age $a$ in $i_{th}$ week.

In the simulation concepts, we adopt the demographic features in Leicester published by the Office for National Statistics (ONS) [ONS, 2021]. We deal with 16 hidden states $\{X_n\}_{n=1}^{16}$ and 16 subintervals $\{\mathcal{T}_n\}_{n=1}^{16}$; each subinterval corresponds to the duration of one week. We infer the latent intensity $\lambda^N(t,a)$, the weights $\{\gamma_{na}\}_{n=1}^{16}$ and the weekly latent infections per age group $a$ via the particle sample derived by drawing samples from the smoothing density with lag equal to 4.

We illustrate a simulation study for 2 age groups. Appendix C.1 also includes a simulation study for 4 age groups. The simulation analysis showed that the KDPF (Algorithm 16) approaches well the ground truth. The reported infections carry information about the progress of the epidemic with a maximum delay of 21 days between the reported and actual infection time for the COVID-19 pandemic. For this reason, the uncertainty of estimates increases in the last days. The MCSEs verified the convergence of posterior estimates concerning the number of particles.

### 6.3.1 Two age groups

ONS shows that 83.1% of the population is aged under 60 years (hereafter 0-59) and 16.9% aged 60 years and over (60+). We coarse the age groups of the contact matrix for reopening schools [Jarvis et al., 2020] and get the matrix:

$$\text{m} = \begin{bmatrix} 6.81 & 0.66 \\ 2.14 & 1.27 \end{bmatrix}.$$

The process is triggered by 4963 infectious uniformly allocated in 21 days before the beginning of the process ($[0,21)$) with a day being the time unit.

We generate weekly latent infections and observed cases according to the model equations (6.1)-(6.5) for weeks -2 $-$ 17 ($[0, 161)$) given $v = 0.004$, $d = 15.22$, $\beta = 0.5$, $\gamma_{-2,0-59} = 0.2$ and $\gamma_{-2,60+} = 0.17$. We consider that about 70% of the population is susceptible at the beginning of week -2 ; 249073 susceptibles ($0-59:$ 206969, $60+:$ 42104).

We are interested in inferring the latent infections in weeks $1-16$. That induces, that $\mathcal{H}_0$ is the set of times of latent infections in weeks -2 $-$ 0. Using the generated observed cases in weeks -1 $-$ 1 as described above, we estimate the latent infections with their associated age groups in weeks -2 $-$ 0 as follows: The latent infections of age group $a_v$ on the week $i$ is equal to the number of events in age $a_v$ reported on the week $(i+1)$ times $1/\beta$, and are spread uniformly in $[(i+2)*7+21, (i+3)*7+21)$ for $-2 \leq i \leq 0$. We assume $\alpha = 0$, $b = 0.5$, $d_{min} = 10$, $d_{max} = 20$, $v_{min} = 0.0001$ and $v_{max} = 0.5$.

The ground truth is characterized by $\mathcal{H}_0$ consisting of 6644 infections ($0-59:$ 4974, $60+:$ 1670), and, 242429 susceptibles ($0-59:$ 201995, $60+:$ 40434) at the beginning of week 1. The estimated $\mathcal{H}_0$ consists of 6624 infectious' times ($0-59:$ 5018, $60+:$ 1606), and the estimated susceptibles are 242449 ($0-59:$ 201951, $60+:$ 40498), respectively. The observed cases in weeks $1-17$ are, 13560 ($0-59:$ 11650, $60+:$ 1910) (Figure 6.1).

Figure 6.2 shows the estimated latent intensities and weekly latent infections for 30000 particles. Figure 6.3 illustrates the estimated weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ and the instantaneous reproduction number per age group $a$. The Credible Intervals (CIs) widen in the last weeks, as we infer the associated random quantities without being aware of the reported cases three weeks ahead. Figure 6.4 illustrates the ESS. The MCSEs show that increasing the number of particles does not have an essential impact on the average variability of the particle mean per state (see Table 6.1). The 95% CIs of the time-constant parameters include the actual values of the parameters.

Most of the time, the estimated weekly latent infections and latent infection intensities are close to the ground truth. The 95% CIs do not cover the ground truth of the weekly latent infections and intensity of latent infections aged 60+ corresponding to the 16th week. This might be observed due to the shortage of information for inferring the latent infections aged 60+ during the last week. We note that the 95% CIs cover the actual values of intensity of latent infections and weekly latent infections concerning all age groups over all states. The simulation analysis demonstrates that the algorithm achieves more accurate estimations concerning all age groups. The analysis also illustrates that the estimated reproduction number per age group $a$ and weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ approaches well the ground truth the majority of the time.

## 6.4 Real Data

We apply the KDPF (Algorithm 16) to real cases among individuals aged 0 to 29 ($0-29$), 30 to 49 ($30-49$), 50 to 69 ($50-69$) and 70+ years in the local authorities: Leicester (10/9/2021 -24/12/2021) [GOV.UK, 2022d], Kinston upon the Thames (11/12/2021 -1/4/2022) [GOV.UK, 2022c] and Ashford (19/12/2021 -

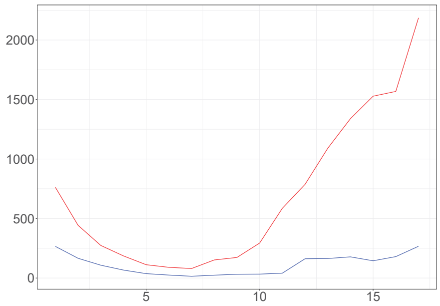
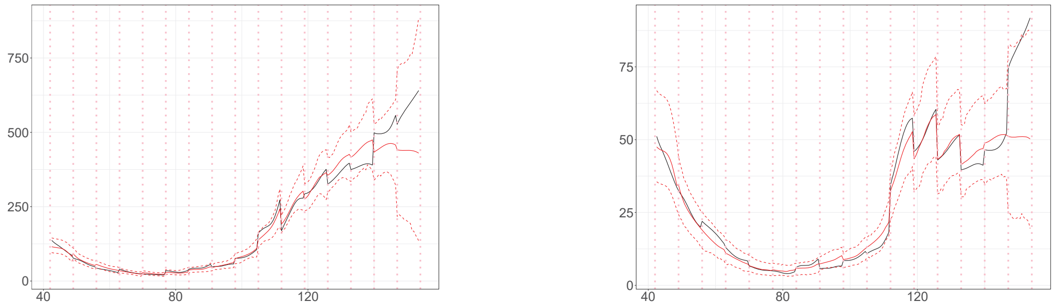
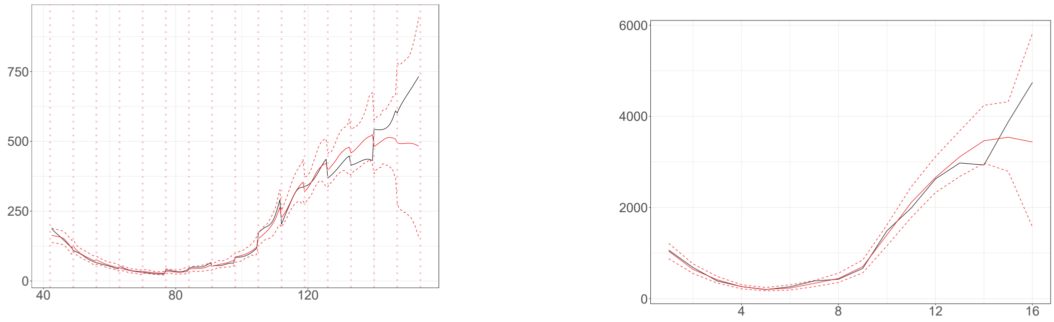


Figure 6.1: The weekly observed cases aged 0-59 (red line) and 60+ (blue line).



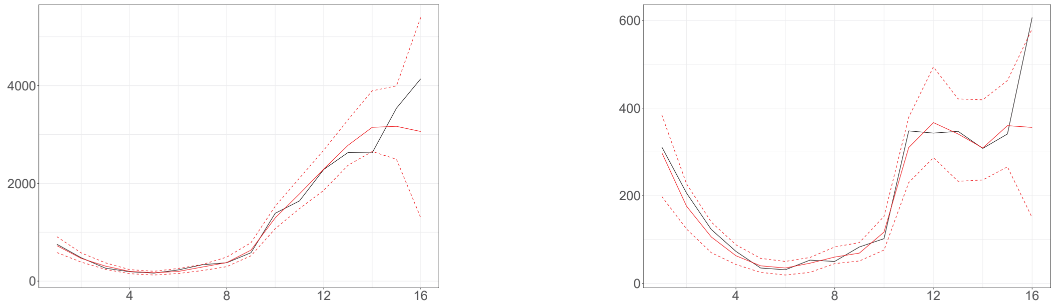
(a) The estimated and true intensity of latent cases aged 0-59.

(b) The estimated and true intensity of latent cases aged 60+.



(c) The aggregate estimated and true intensity of latent cases.

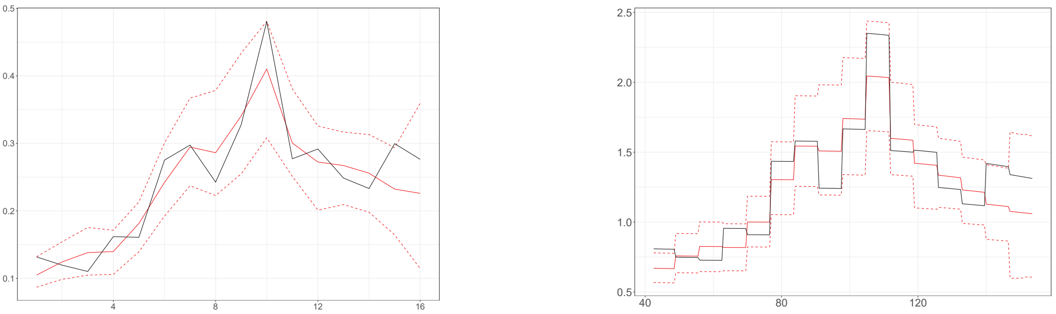
(d) The aggregate estimated weekly hidden cases.



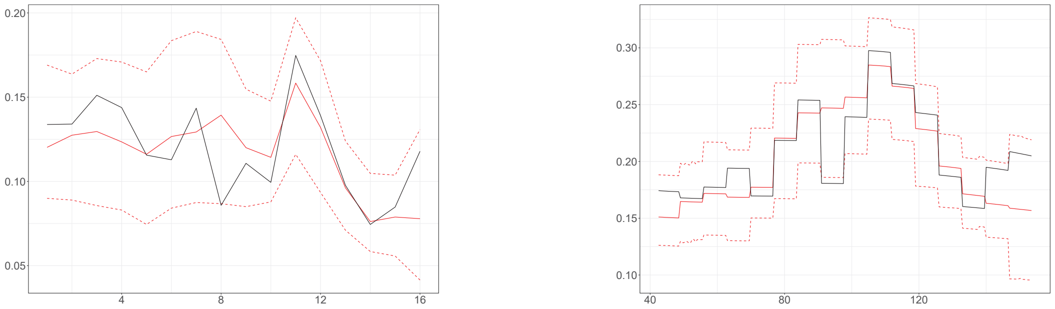
(e) The estimated weekly infected cases aged 0-59.

(f) The estimated weekly infected cases aged 60+.

Figure 6.2: The ground truth (black line), and the estimated intensities of latent infections and weekly latent infections for the synthetic scenario using 2 age groups (posterior median (red line) ; 95% CI (red dashed line)). The vertical dotted lines show the beginning of each week in the period we examine.



(a) $\{\gamma_{i,0-59}\}_{i=1}^{16}$ (b) $R_{0-59}$



(c) $\{\gamma_{i,60+}\}_{i=1}^{16}$ (d) $R_{60+}$

Figure 6.3: The estimated weights $\{\gamma_{na}\}_a$ and instantaneous reproduction numbers $\{R_a\}_a$ (posterior median (red line); 95% CI (red dashed line)) and the true values (black line) for the synthetic scenario using 2 groups plotted against subintervals and time, respectively.

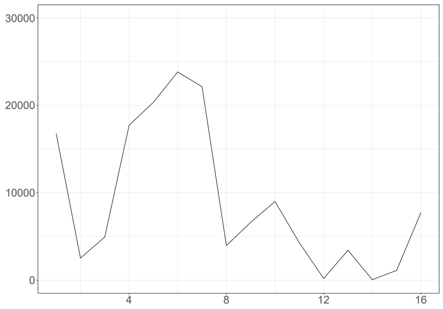


Figure 6.4: The ESS for the synthetic scenario using 2 age groups.

| Convergence of the posterior estimates | | | |
|---|---|---|---|
| $MCSE$ | $N = 10000$ | $N = 20000$ | $N = 30000$ |
| $\overline{\gamma_1}$ | 0.000173 | 0.000189 | 0.000182 |
| $\overline{\gamma_2}$ | 0.000123 | 0.000119 | 0.000118 |
| $\overline{Y_1}$ | 1.05305 | 1.114323 | 1.077618 |
| $\overline{Y_2}$ | 0.193806 | 0.191016 | 0.193658 |

Table 6.1: MCSEs of the average of weights and weekly hidden infections, estimated by the posterior means for the synthetic scenario using 2 age groups.

9/4/2022) [GOV.UK, 2022a] (see Figure 6.5). Estimates of the local authorities' population are available from ONS [ONS, 2021]. We deal with 16 hidden states $\{X_n\}_{n=1}^{16}$ and 16 subintervals $\{\mathcal{T}_n\}_{n=1}^{16}$; each subinterval corresponds to the duration of one week. We infer the latent intensity $\lambda^N(t,a)$ and the instantaneous reproduction number, $R_a(t)$ per age group $a$ and the weights, $\{\gamma_{na}\}_a$ as well as the weekly and daily latent cases via the particle sample derived by drawing samples from the smoothing density with lag equal to 4. We also infer the instantaneous reproduction number, $R(t)$ as the infected population-weighted average of $R_a(t)$. Appendix C.3 includes the simulation study for the local authorities: Leicester and Ashford.

We assume that the weights $\{\gamma_{1a}\}_a$ are uniformly distributed over the interval from $2\left(\sum_{a_i}\sum_a \frac{S_{t,a}}{N_a}\mathrm{m}_{\mathrm{aa_i}}\right)^{-1}$ to $8\left(\sum_{a_i}\sum_a \frac{S_{t,a}}{N_a}\mathrm{m}_{\mathrm{aa_i}}\right)^{-1}$ by assuming that the average number of secondary cases an individual would infect at the beginning of the first week is uniformly distributed over $[0.5, 2]$. Our initialization is based on the published reproduction number in England from the government in the UK (see section 5.5). We also assume that 50% of the infections are reported ($\beta = 0.5$), $d_{min} = 1$, $d_{max} = 10$, $v_{min} = 0.0001$ and $v_{max} = 0.5$. We coarse the age groups of the contact matrix for reopening schools [Jarvis et al., 2020] to find the matrix m.

We use the percentage of the population aged $a$ with levels of antibodies against SARS-CoV-2 at or above a threshold of 179 nanograms per millilitre (ng/ml), denoted by $p_a$, estimated by the lower 95% credible interval available from ONS [2022], to initialize the number of susceptibles aged $a$ at the beginning of the first week. The percentage of people 0-15 having antibodies at or above the threshold is not available. We assume the percentage is the same as that of 16-24. Following this methodology, $(1-p_a)N_a$ gives the susceptible population aged $a$ at the start of the first week. However, the estimated susceptible population is less than the total number of reported infections in Ashford and Kingston for some age bands. The antibodies might not fully protect against infection in December 2021; this could be due to declining immunity or immune escape (new variants being different from old variants and thus the previous infection being less protective against a new one). For this reason, we assume that a smaller percentage of the population aged $a$, given by $p_a - z$ with $0 < z < \min_a p_a$, has enough antibodies against a new infection in December 2021. We choose the minimum value of $z$ so that the susceptible population is at least twice as large as the reported cases for each age group to maintain consistency with our assumption that we see 50% of the infections. Summarizing the number of susceptibles aged $a$ at the beginning of the first week in Leicester (4/9/2021), Kingston upon Thames (11/12/2021) and Ashford (19/12/2021) are given by $(1-p_a)N_a$, $(1-p_a+0.2)N_a$ and $(1-p_a+0.2)N_a$, respectively. Appendix C.4 includes an alternative method to initialize the size of susceptible population per age group at the beginning of the process.

Figures 6.6-6.9, C.18- C.32 show the estimated intensity, the estimated weekly and daily hidden cases, the estimated susceptibles, the estimated instantaneous reproduction number and the estimated weights $\{\gamma_{na}\}_{n=1}^{16}$ per age group $a$, the 95% CIs of time-constant parameters, and the ESS using 40000 particles. We observe that the estimated latent intensity and the estimated weekly latent cases of age $a$ are consistent with the weekly observed cases of age $a$. The analysis demonstrates that the instantaneous reproduction numbers $\{R_a(t)\}_a$ and $R(t)$ reflect the progress of the pandemic and capture the changes. As expected, the estimates of last days are more uncertain.

**Comparing the age-stratified model (model A) with the unstructured homogeneously mixing model (model U) introduced in Chapter 5.** Model A includes model U. Following Pellis et al. [2020], we assume model A reflects better the reality and measures the other model's deviation from it.

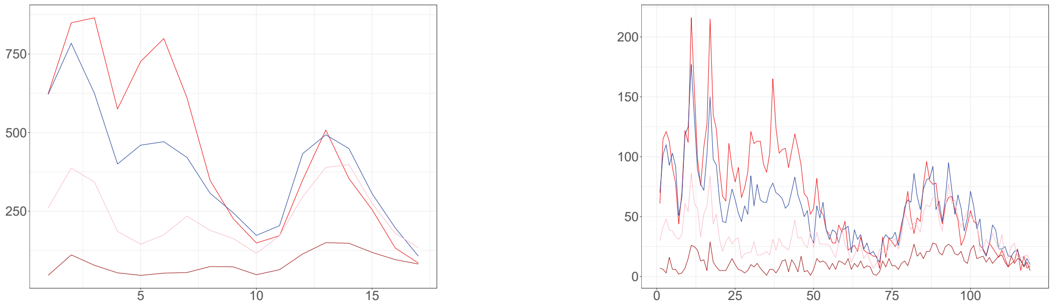
We run model U (Algorithm 9) considering all reported infections and separately for each age band to estimate the aggregated and per age group latent intensity, the aggregated and per age group weekly and daily hidden cases, and the instantaneous reproduction number. We assume that $d_{min} = 1$, $d_{max} = 10$, $(\alpha, b) = (0.5, 3)$ for Ashford, $(\alpha, b) = (0.1, 3.5)$ for Kingston upon Thames and $(\alpha, b) = (0.1, 3)$ for Leicester. Inspired by Pei et al. [2020], we evaluate the goodness of fit of model U using the weighted mean absolute percentage error (wMAPE) of the forecast of model U in relation to that of model A

$$\text{wMAPE} = \frac{\sum_{i=1}^{k} |A_i - U_i|}{\sum_{i=1}^{k} A_i},$$

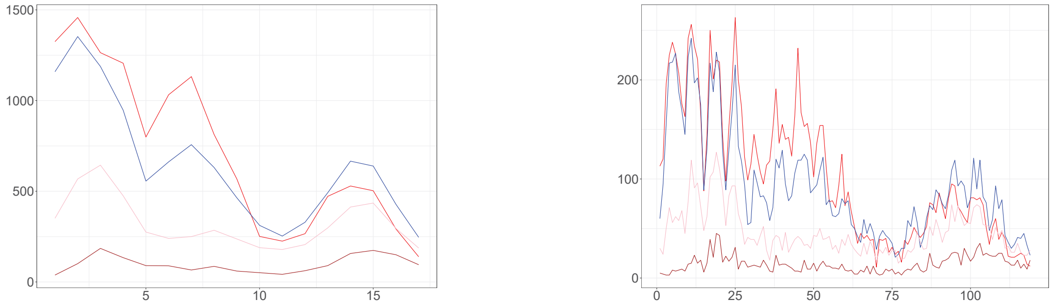
where $A_i$ and $U_i$ are the estimations of models A and U via posterior median associated with the interval $\mathcal{T}_i$. Table 6.2 illustrates the goodness fit of the model using the metric wMAPE.

The instantaneous reproduction number's posterior medians of both models follow the same pattern in general lines (see Figures C.26, 6.13, C.36). The estimated aggregated and per age group latent intensity, weekly and daily hidden cases via posterior median given by model A are similar to the ones of model U (see Figures C.23-C.25, 6.10-6.12, C.33-C.35). Model U derives similar estimates to those of model A, as the latent intensity and cases per age band are strongly dependent on the infections of the associated age group and independent of the other age groups' infections. The simulation analysis shows that considering the individual inhomogeneity in age and finite population, the width of CIs decreases.

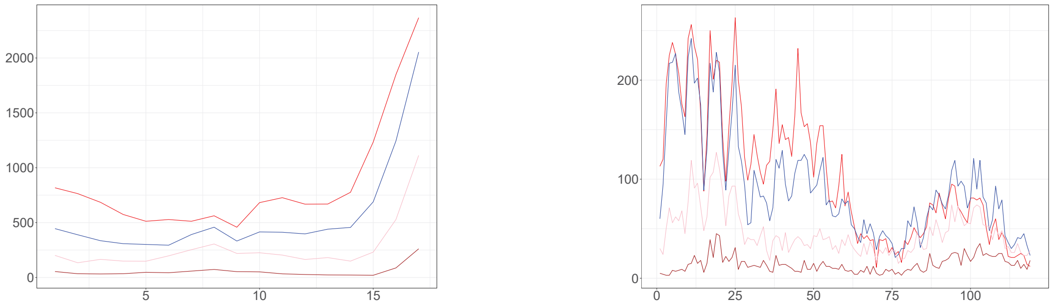
The comparison between models A and U demonstrates that model U provides estimates close to the reality for the latent intensity, weekly and daily hidden cases, and a rough approximation of the instantaneous reproduction number. The possible noticeable differences in both models' estimates during the first three weeks are due to different initializations. Table 6.2 illustrates that the metric is higher than 0.1 when models' medians present differences in the first weeks. We suggest running both models considering the former $\eta$ days of the horizon we are interested in, where $\eta$ is the delay in which the transition kernel of latent cases is negligible; $\eta = 21$ days or 3 weeks for COVID-19 data. The analysis also shows that model A decreases the uncertainty of estimates and infers the reproduction number per age group. Model U cannot be applied to infer the reproduction number per age group number and investigate the age groups' behaviours in an epidemic indicating the importance of model A. The instantaneous reproduction numbers per age group provide a real-time measurement of interventions and behavioural changes.



(a) The weekly observed cases in Ashford (b) The daily observed cases in Ashford

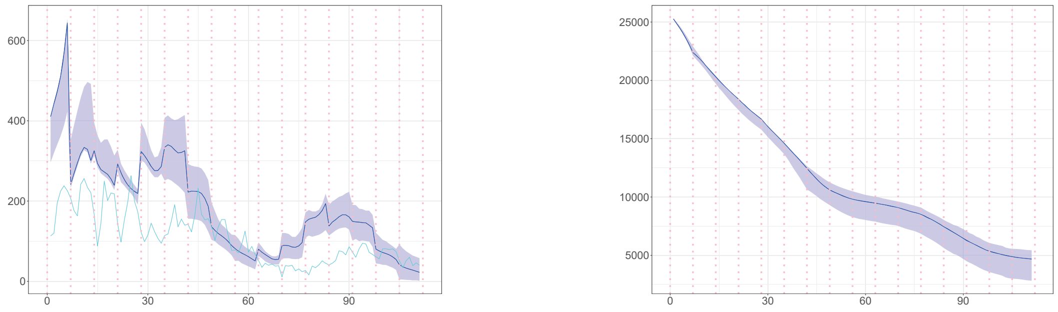


(c) The weekly observed cases in Kingston (d) The daily observed cases in Kingston



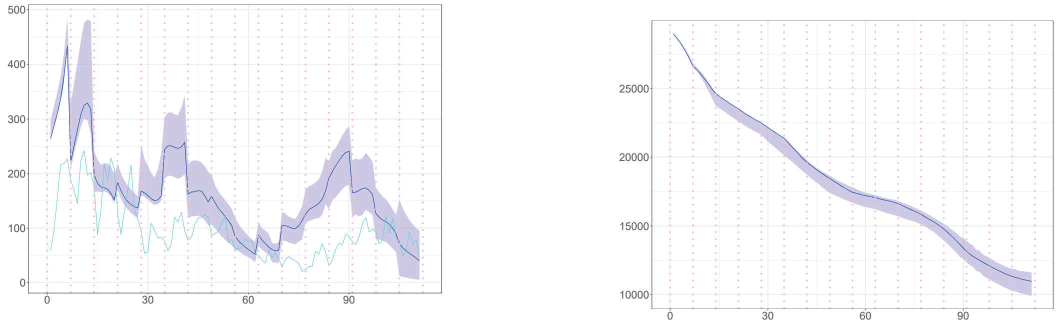
(e) The weekly observed cases in Leicester (f) The daily observed cases in Leicester

Figure 6.5: The weekly and observed cases in age groups 0-29 (red line), 30-49 (blue line), 50-69 (pink line) and 70+ (brown line) in the local authorities plotted against time.



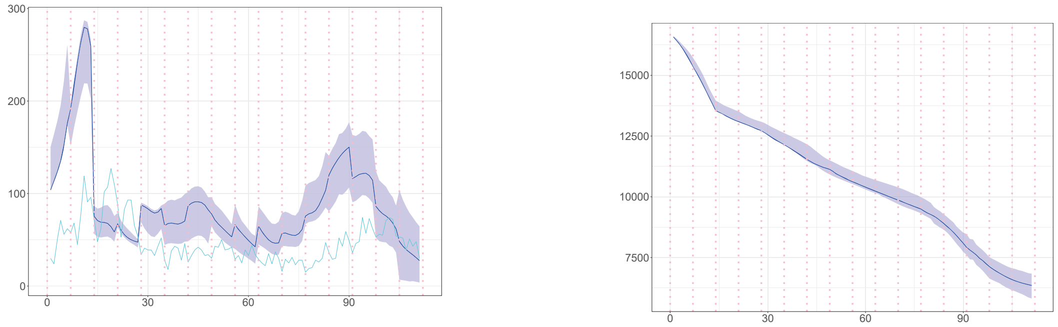
(a) The estimated intensity of latent cases aged 0-29.

(b) The estimated susceptibles aged 0-29.



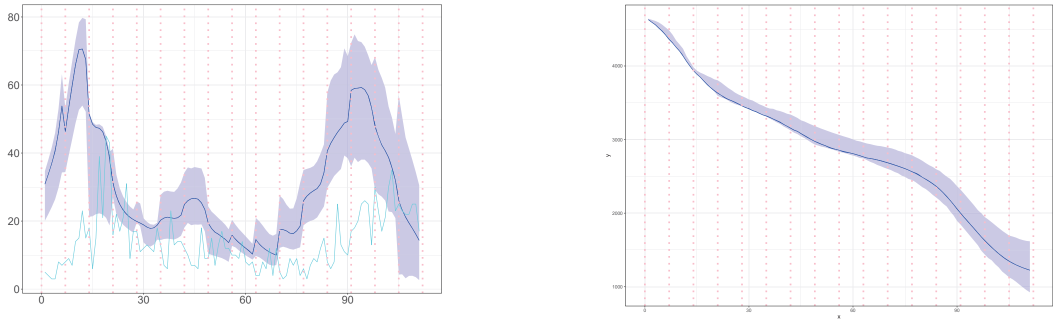
(c) The estimated intensity of latent cases aged 30-49.

(d) The estimated susceptibles aged 30-49.



(e) The estimated intensity of latent cases aged 50-69.

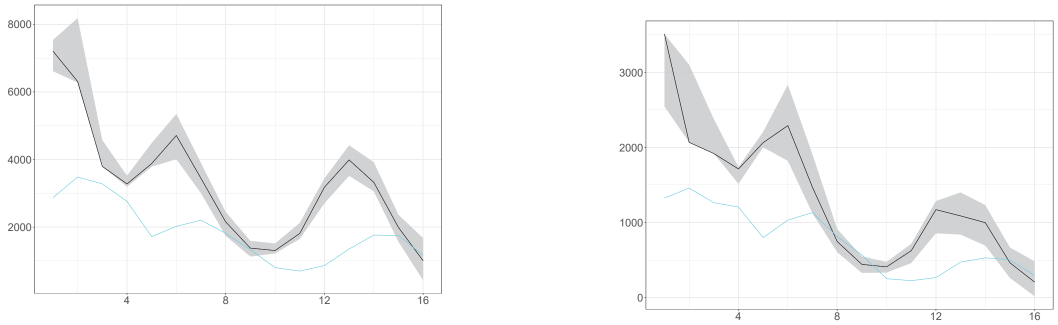
(f) The estimated susceptibles aged 50-69.



(g) The estimated intensity of latent cases aged 70+.

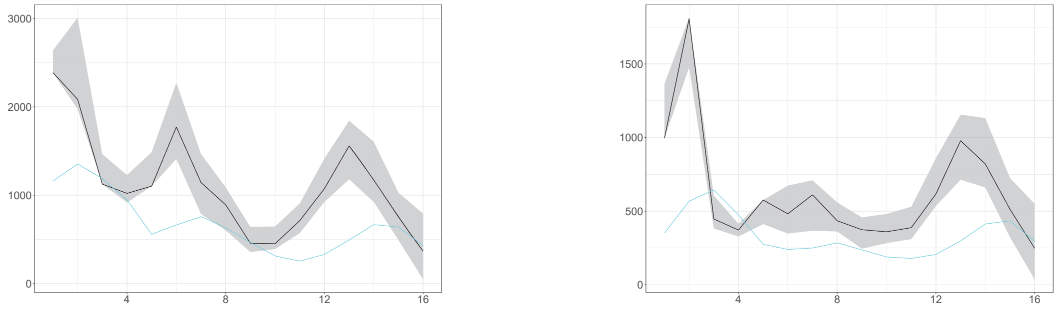
(h) The estimated susceptibles aged 70+.

Figure 6.6: The estimated latent intensity and susceptibles (posterior median (blue line) ; 95% CI (ribbon)) and the daily observed cases (cyan line) in Kingston plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



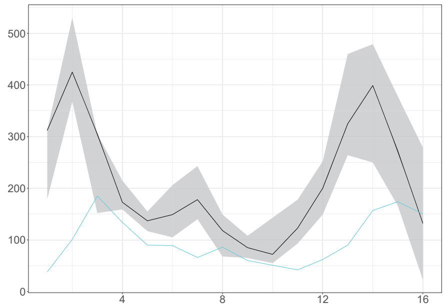
(a) The estimated aggregated weekly hidden cases.

(b) The estimated weekly hidden cases aged 0-29.



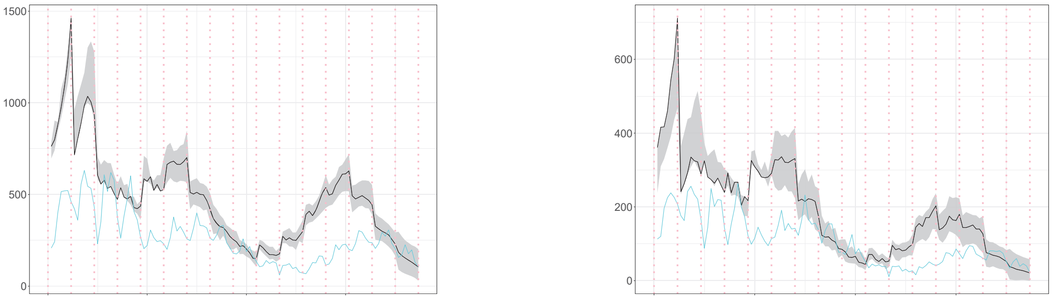
(c) The estimated weekly hidden cases aged 30-49.

(d) The estimated weekly hidden cases aged 50-69.

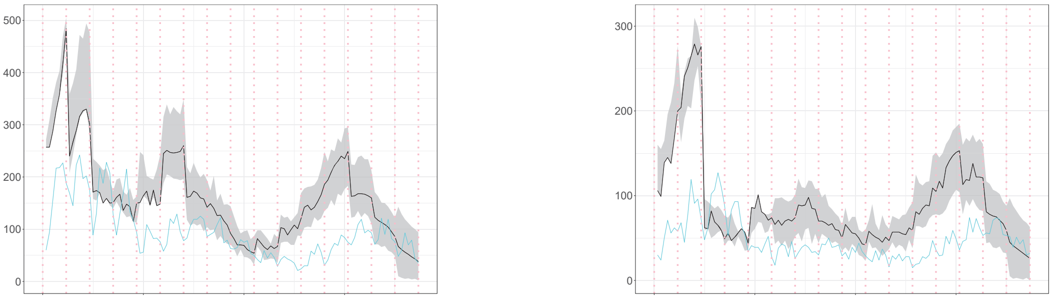


(e) The estimated weekly hidden cases aged 70+.

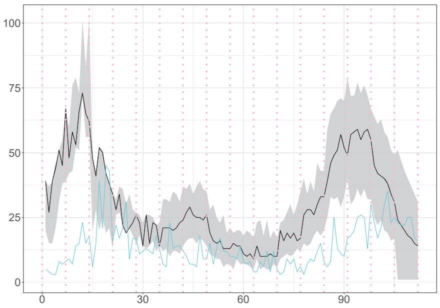
Figure 6.7: The estimated weekly latent cases (black line; 95% CI (ribbon)) and the weekly observed cases (cyan line) in Kingston plotted against time.



(a) The estimated aggregated daily hidden cases. (b) The estimated daily hidden cases aged 0-29.

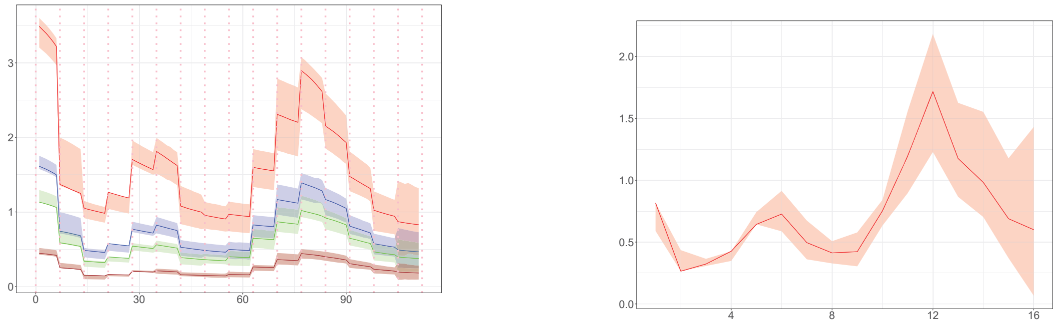


(c) The estimated daily hidden cases aged 30-49. (d) The estimated daily hidden cases aged 50-69.



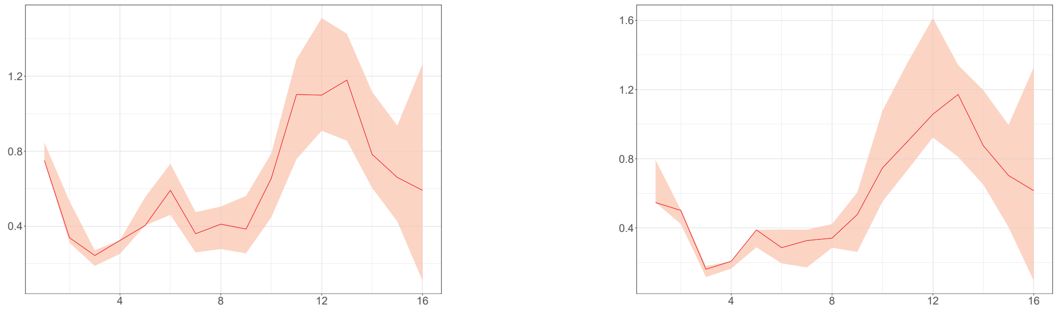
(e) The estimated daily hidden cases aged 70+.

Figure 6.8: The estimated daily latent cases (posterior median (black line); 95% CI (ribbon)) and the daily observed cases (cyan line) in Kingston plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



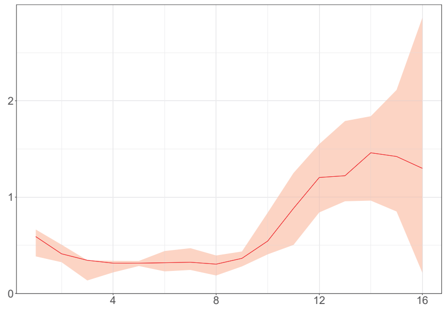
(a) The per age group instantaneous reproduction numbers.

(b) The estimated weights $\{\gamma_{i,0-29}\}_{i=1}^{16}$.



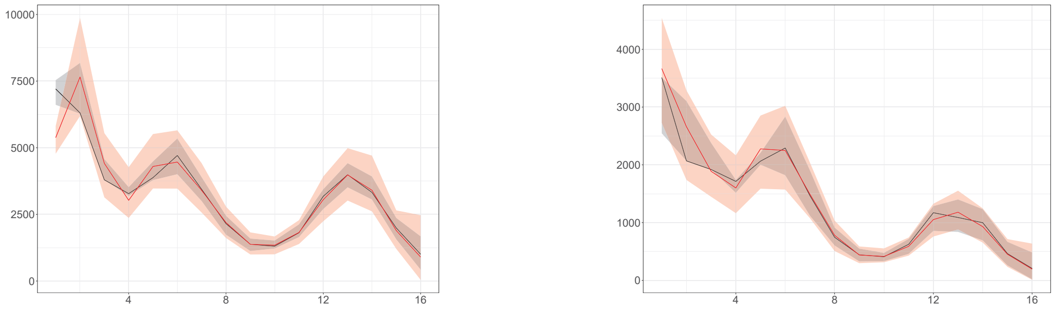
(c) The estimated weights $\{\gamma_{i,30-49}\}_{i=1}^{16}$.

(d) The estimated weights $\{\gamma_{i,50-69}\}_{i=1}^{16}$.

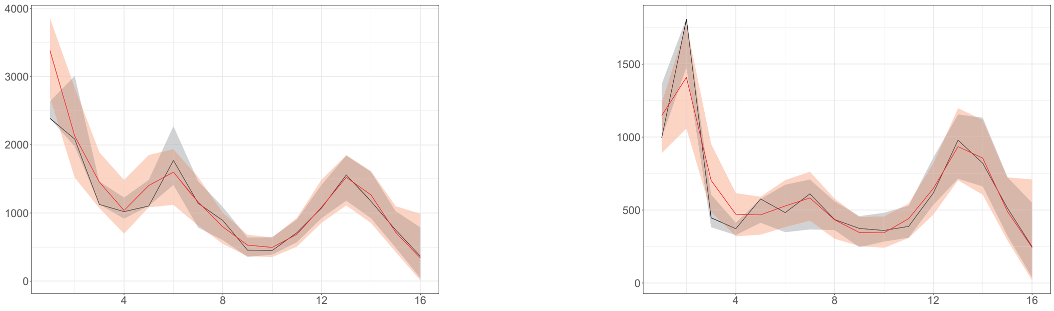


(e) The estimated weights $\{\gamma_{i,70+}\}_{i=1}^{16}$.

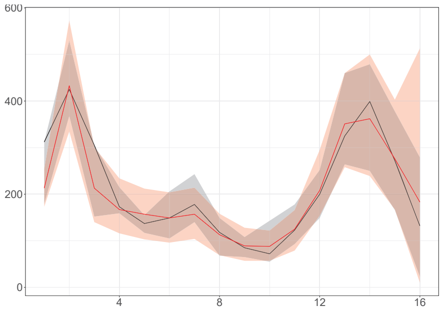
Figure 6.9: The posterior median estimate of instantaneous reproduction number per age group (0-29 (red line), 30-49 (blue line), 50-69 (green line) and 70+ (brown line)), the posterior median estimate of weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ (red line) and the 95% CIs (ribbon) for Kingston plotted against time.



(a) Aggregated weekly hidden cases. (b) Aged 0-29.

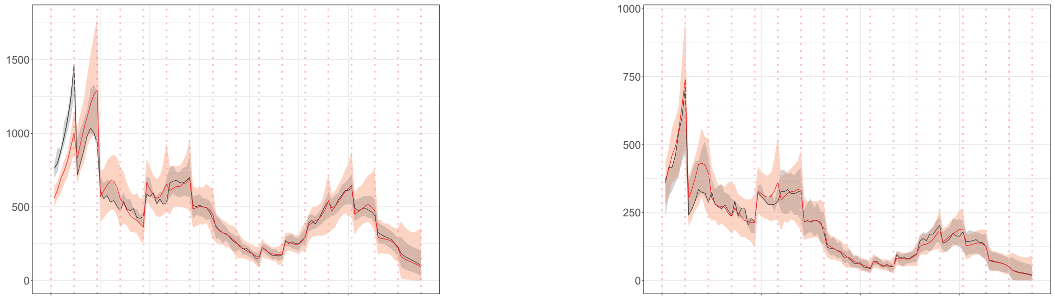


(c) Aged 30-49. (d) Aged 50-69.



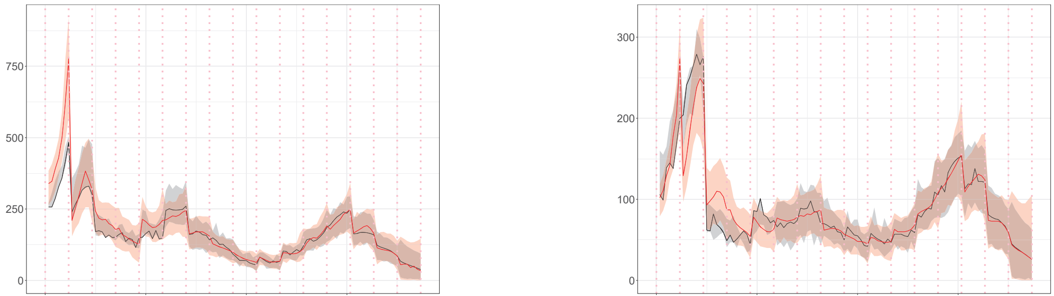
(e) Aged 70+.

Figure 6.10: The posterior median estimate of weekly hidden cases of model A (black line) and model U (red line), and the 95% CIs (ribbon) in Kingston plotted against time.



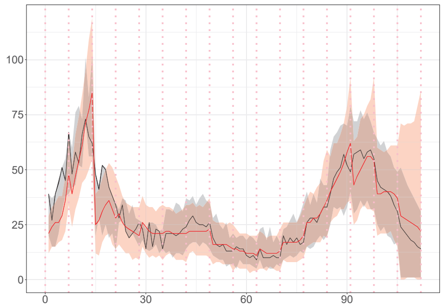
(a) Aggregated daily hidden cases.

(b) Aged 0-29.



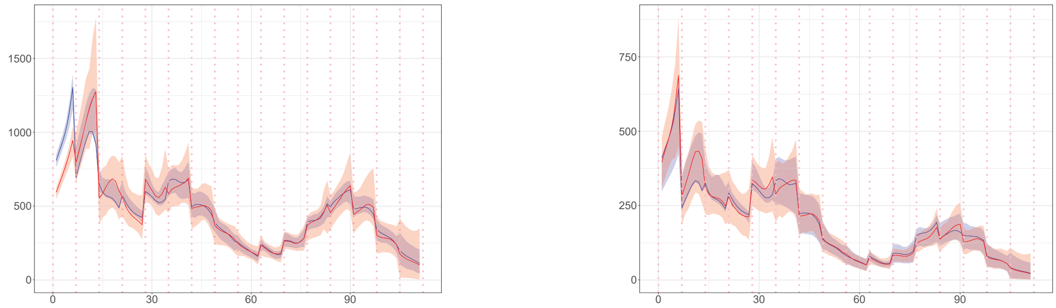
(c) Aged 30-49.

(d) Aged 50-69.

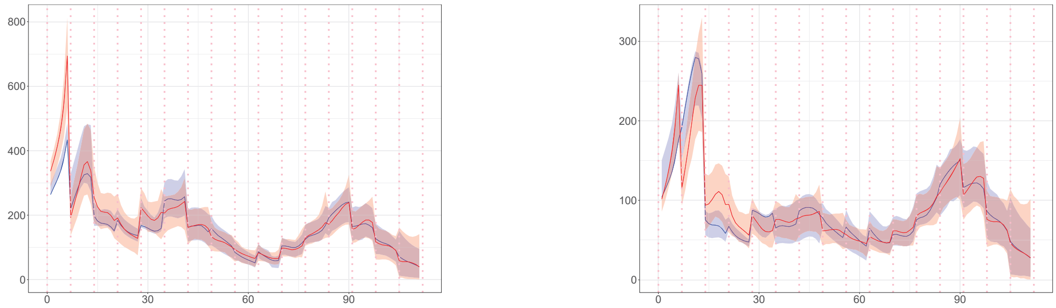


(e) Aged 70+.

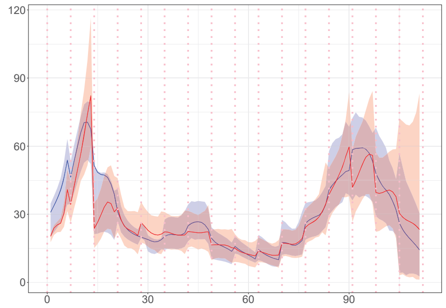
Figure 6.11: The posterior median estimate of daily hidden cases of model A (black line) and model U (red line), and the 95% CIs (ribbon) in Kingston plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



(a) Aggregated latent intensity. (b) Aged 0-29.



(c) Aged 30-49. (d) Aged 50-69.



(e) Aged 70+.

Figure 6.12: The posterior median estimate of latent intensity of model A (blue line) and model U (red line), and the 95% CIs (ribbon) in Kingston plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.

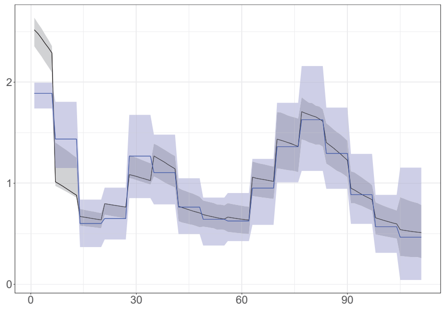


Figure 6.13: The posterior median estimate of instantaneous reproduction number of model A (black line) and model U (blue line), and the 95% CIs (ribbon) in Kingston plotted against time.

| Goodness of fit of Model U | | | |
|---|---|---|---|
| wMAPE | Kingston | Ashford | Leicester |
| R | 0.12 | 0.06 | 0.12 |
| WHC aged 0-29 | 0.07 | 0.06 | 0.11 |
| WHC aged 30-49 | 0.13 | 0.05 | 0.14 |
| WHC aged 50-69 | 0.13 | 0.05 | 0.18 |
| WHC aged 70+ | 0.12 | 0.17 | 0.33 |
| aggregated WHC | 0.1 | 0.03 | 0.11 |
| DHC aged 0-29 | 0.08 | 0.07 | 0.12 |
| DHC aged 30-49 | 0.14 | 0.07 | 0.14 |
| DHC aged 50-69 | 0.14 | 0.08 | 0.19 |
| DHC aged 70+ | 0.16 | 0.2 | 0.36 |
| aggregated DHC | 0.1 | 0.05 | 0.11 |
| LI aged 0-29 | 0.08 | 0.09 | 0.11 |
| LI aged 30-49 | 0.13 | 0.07 | 0.13 |
| LI aged 50-69 | 0.14 | 0.07 | 0.16 |
| LI aged 70+ | 0.15 | 0.2 | 0.3 |
| aggregated LI | 0.1 | 0.06 | 0.1 |

Table 6.2: The goodness of fit of model U for instantaneous reproduction number (R), weekly hidden cases (WHC), daily hidden cases (DHC) and latent intensity (LI) using the metric wMAPE.

**Future prediction** Applying the steps outlined in Algorithm 17, it is also possible to forecast with relative accuracy how many new infections would be reported during the next week, $\mathcal{T}_{17}$ (see Tables 6.3,C.4, C.5). Our forecast is subject to similar levels of uncertainty as last week's estimates. However, the algorithm can be applied to forecast the epidemic's future trajectory-whether it will be upward, downward or stable-using the posterior mean and median.

| Proposed Method | | | | | |
|---|---|---|---|---|---|
| Reported infections | Posterior Mean | Posterior Median | 95% CIs | True Number ($\mathcal{T}_{17}$) | True Number ($\mathcal{T}_{16}$) |
| aggregated | 664 | 649 | (427, 921) | 667 | |
| aged 0-29 | 154 | 145 | (64, 258) | 139 | 292 |
| aged 30-49 | 250 | 238 | (112, 404) | 246 | 428 |
| aged 50-69 | 172 | 164 | (77, 279) | 187 | 295 |
| aged 70+ | 89 | 85 | (39, 142) | 95 | 150 |

Table 6.3: The true number of reported infections in $\mathcal{T}_{16}$ and $\mathcal{T}_{17}$, and the posterior median, the posterior mean and the 95% CIs of the estimated infections in $\mathcal{T}_{17}$ in Kingston.

**Algorithm 17** Predicting the new aggregated and per age group observed cases in near future

```
Let $\hat{v} = \frac{1}{N}\sum_{j=1}^{N} v_{j,16}$ and $\hat{d} = \frac{1}{N}\sum_{j=1}^{N} d_{j,16}$.
for $j = 1, .., N$ do
    $\gamma_{j,17,a} \sim P(\gamma_{17,a}|\gamma_{j,16,a}, \hat{d})$, $\forall$ age group $a$
    $\left(S_{j,17}^{N}, A_{j,17}^{N}\right) \sim P\left(S_{17}^{N}, A_{17}^{N}|S_{j,1:16}^{N}, A_{j,1:16}^{N}, \{\gamma_{j,17,a}\}_a, \mathcal{H}_0, A_0^N\right)$
    Calculate the mean of observed cases of age $a$ in the interval denoted by $\mu_{j,17,a}$, $\forall$ age group $a$.
    $Y_{j,17,a} \sim \mathrm{NB}(\mu_{j,17,a}, \hat{v})$, $\forall$ age group $a$.
end for
Use the sample $\{Y_{j,17,a}\}_{j=1}^{N}$ to find the posterior mean, the posterior median and the 95% CI of the estimated observed cases of
    age a in $\mathcal{T}_{17}$.
```

## 6.5 Epidemic Dynamics

The KDPF (Algorithm 16) can also infer the number of directed links between age groups $a \in \mathcal{A}$ and $a' \in \mathcal{A}$ (hereafter $a \to a'$) by capturing the process' branching structure. The underlying dynamics are revealed either by saving the parent of each latent infection (Method A) or by sampling its parent from a multinomial distribution (Method B) as described in Section 6.2. We employed 30 randomly selected particles in Method B due to computational constraints.

Similar to Section 6.3, we use the ONS' released demographic data for Leicester in the simulation concepts. We work with 16 hidden states $\{X_n\}_{n=1}^{16}$ and 16 subintervals $\{\mathcal{T}_n\}_{n=1}^{16}$, where each subinterval has a week-long length. We coarse the age groups of the contact matrix for reopening schools [Jarvis et al., 2020] to find $\{\mathrm{m_{aa'}}\}_{\mathrm{a,a'}\in\mathcal{A}}$. We infer the number of directed links between age groups in weeks 1-16.

We illustrate the simulation study for 2 and 4 age groups:

- **2 age groups:** We generated weekly latent and observed cases according to the model equations 6.1-6.5 for weeks -2 $-$ 17 ($[21, 161)$) given that the process is triggered by 1630 infectious, 56% of the population is susceptible at the beginning of week -2 ($0-59:\ 126629,\ 60+:\ 25777$), $v = 0.003$, $d = 25.4$, $\beta = 0.5$, $\gamma_{-2,0-59} = 0.35$ and $\gamma_{-2,60+} = 0.35$. The observed cases in weeks $1-17$ are 9890 ($0-59:\ 8157,\ 60+:\ 1733$) (Figure 6.14). We ran the KDPF by assuming $\alpha = 0$, $b = 0.5$, $d_{min} = 20$, $d_{max} = 30$, $v_{min} = 0.0001$, $v_{max} = 0.5$ and 30000 particles.

- **4 age groups:** We generated weekly latent and observed cases according to the model equations 6.1-6.5 for weeks -2 $-$ 17 ($[21, 161)$) given that the process is triggered by 564 infectious, 56% of the population is susceptible at the beginning of week -2 ($0-29:\ 72426,\ 30-49:\ 39018,\ 50-69:\ 29034,\ 70+:\ 12994$), $v = 0.005$, $d = 23.7$, $\beta = 0.5$, $\gamma_{-2,0-29} = 0.4$, $\gamma_{-2,30-49} = 0.47$, $\gamma_{-2,50-69} = 0.36$ and $\gamma_{-2,70+} = 0.18$. The observed cases in weeks $1-17$ are 8769 ($0-29:\ 4861,\ 30-49:\ 1849,\ 50-69:\ 1869, 70+:\ 190$) (Figure 6.14). We ran the KDPF by assuming $\alpha = 0$, $b = 1$, $d_{min} = 20$, $d_{max} = 30$, $v_{min} = 0.0001$, $v_{max} = 0.5$ and 40000 particles.

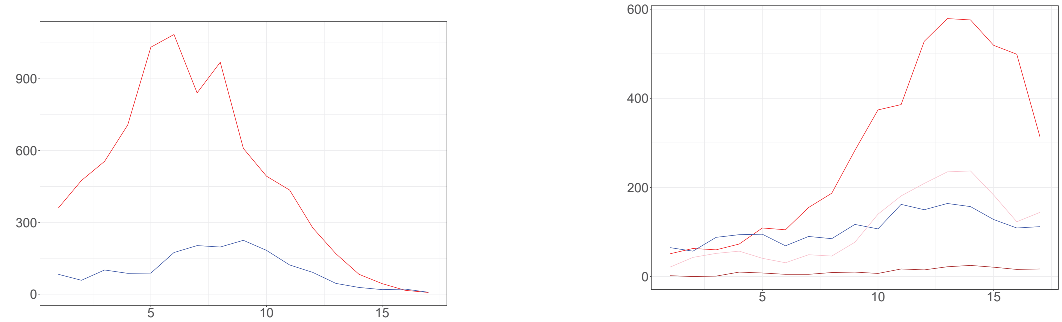
Both scenarios are consistent with Edmunds et al. [1997], i.e., older participants have more contacts with older people, and Wallinga et al. [2006], i.e., younger adults are more likely to catch an infection and contribute more to its transmission due to their high number of contacts. Tables 6.4-6.7 demonstrate that the proposed algorithm approaches well the ground truth.

| Estimating the number of directed links $a_i \to a_j$ using Method A | | | |
|---|---|---|---|
| $a_i \to a_j$ | Posterior Median | 99% CI | True Number |
| $a_1 \to a_1$ | 8527 | (7635, 9580) | 8774 |
| $a_1 \to a_2$ | 2543 | (2271, 2831) | 2654 |
| $a_1 \to a_3$ | 2289 | (2039, 2543) | 2515 |
| $a_1 \to a_4$ | 194 | (146, 243) | 217 |
| $a_2 \to a_1$ | 1071 | (946, 1207) | 1093 |
| $a_2 \to a_2$ | 686 | (573, 805) | 710 |
| $a_2 \to a_3$ | 639 | (544, 735) | 662 |
| $a_2 \to a_4$ | 90 | (64, 119) | 106 |
| $a_3 \to a_1$ | 563 | (481, 647) | 616 |
| $a_3 \to a_2$ | 357 | (296, 424) | 392 |
| $a_3 \to a_3$ | 729 | (606, 854) | 789 |
| $a_3 \to a_4$ | 95 | (67, 123) | 109 |
| $a_4 \to a_1$ | 13 | (5, 23) | 16 |
| $a_4 \to a_2$ | 13 | (5, 23) | 12 |
| $a_4 \to a_3$ | 24 | (11, 37) | 27 |
| $a_4 \to a_4$ | 11 | (4, 22) | 14 |

Table 6.4: The true number, the posterior median and the 99% CI of the number of directed links between age groups ($a_1 : 0 - 29$, $a_2 : 30 - 49$, $a_3 : 50 - 69$, $a_4 : 70+$).

| Estimating the number of directed links $a_i \to a_j$ using Method B | | | |
|---|---|---|---|
| $a_i \to a_j$ | Posterior Median | 99% CI | True Number |
| $a_1 \to a_1$ | 8438 | (7934, 9100) | 8774 |
| $a_1 \to a_2$ | 2539 | (2339, 2681) | 2654 |
| $a_1 \to a_3$ | 2285 | (2122, 2445) | 2515 |
| $a_1 \to a_4$ | 189 | (156, 252) | 217 |
| $a_2 \to a_1$ | 1077 | (950, 1232) | 1093 |
| $a_2 \to a_2$ | 623 | (604, 787) | 710 |
| $a_2 \to a_3$ | 632 | (552, 721) | 662 |
| $a_2 \to a_4$ | 91 | (69, 116) | 106 |
| $a_3 \to a_1$ | 572 | (514, 628) | 616 |
| $a_3 \to a_2$ | 361 | (323, 397) | 392 |
| $a_3 \to a_3$ | 736 | (593, 875) | 789 |
| $a_3 \to a_4$ | 94 | (77, 111) | 109 |
| $a_4 \to a_1$ | 14 | (3, 20) | 16 |
| $a_4 \to a_2$ | 12 | (5, 24) | 12 |
| $a_4 \to a_3$ | 25 | (14, 43) | 27 |
| $a_4 \to a_4$ | 12 | (4, 19) | 14 |

Table 6.5: The true number, the posterior median and the 99% CI of the number of directed links between age groups ($a_1 : 0 - 29$, $a_2 : 30 - 49$, $a_3 : 50 - 69$, $a_4 : 70+$).



(a) The weekly observed cases aged 0-59 (red line) and 60+ (blue line).

(b) The weekly observed cases aged 0-29 (red line), 30-49 (blue line), 50-69 (brown line) and 70+ (pink line).

Figure 6.14: The weekly observed cases.

| Estimating the number of directed links $a_i \to a_j$ using Method A | | | |
|---|---|---|---|
| $a_i \to a_j$ | Posterior Median | 99% CI | True Number |
| $a_1 \to a_1$ | 15219 | (14083, 16453)) | 15214 |
| $a_1 \to a_2$ | 2936 | (2638, 3205) | 3025 |
| $a_2 \to a_1$ | 297 | (243, 350) | 300 |
| $a_2 \to a_2$ | 389 | (317, 461) | 381 |

Table 6.6: The true number, the posterior median and the 99% CI of the number of directed links between age groups ($a_1 : 0-59$, $a_2 : 60+$).

| Estimating the number of directed links $a_i \to a_j$ using Method B | | | |
|---|---|---|---|
| $a_i \to a_j$ | Posterior Median | 99% CI | True Number |
| $a_1 \to a_1$ | 15214 | (14478, 16294)) | 15214 |
| $a_1 \to a_2$ | 2946 | (2697, 3076) | 3025 |
| $a_2 \to a_1$ | 298 | (250, 339) | 300 |
| $a_2 \to a_2$ | 391 | (330, 472) | 381 |

Table 6.7: The true number, the posterior median and the 99% CI of the number of directed links between age groups ($a_1 : 0-59$, $a_2 : 60+$).

## 6.6 Discussion

Isham and Medley [1996]; Wallinga et al. [1999]; Farrington et al. [2001] contend that it is necessary to account for individual heterogeneities while modelling the transmission of an infectious disease. The unstructured homogeneously mixing epidemic model (Model U) introduced in Chapter 5 is a step toward developing epidemic models considering individual heterogeneities and revealing underlying dynamics. In this chapter, working in this direction, we suggest a novel age-stratified epidemic model considering a finite population (Model A). We propose a KDPF for inferring the marked counting process and forecasting the epidemic's future trajectory in the near future. We demonstrate the performance of the proposed algorithm on COVID-19.

The simulation analysis of synthetic data shows that the KDPF approaches well the ground truth. We demonstrate that the estimated latent cases and the latent intensity are consistent with the aggregated observed cases for each age group in various local authorities in the UK. The analysis also reveals that each age group's instantaneous reproduction number reflects the pandemic's progression and the group's evolving behavioural characteristics. The uncertainty of estimates increases in the last states of the algorithm, as the reported infections carry information about the progress of the epidemic by transferring the delay between the reported and the actual infection time. We also show how the algorithm can be employed to project the epidemic's course over short time horizons. However, our forecast is subject to similar levels of uncertainty as the last state's estimates.

According to Cori et al. [2013], the size of the time window will impact the estimations of the instantaneous reproduction number. Small sizes lead to faster detection of transmission changes and higher statistical noise, whereas large sizes lead to more smoothing and reductions in statistical noise. In accordance with Cori et al. [2013], who suggest an appropriate way of choosing the time window size, we have selected a weekly time window to analyse the real data.

Model A includes model U, and according to Pellis et al. [2020], model A reflects better the spreading of the epidemic. The estimated aggregated and per age group latent intensity, weekly and daily hidden cases via posterior median given by model A are similar to the ones of model U. The instantaneous reproduction number's posterior medians of both models follow the same pattern in general lines. The analysis shows that model A derives narrower CIs, indicating that considering the individual inhomogeneity in age and finite population decreases the uncertainty of estimates; that conclusion is under investigation by increasing the number of particles, proposing alternative inference methods and priors. The instantaneous reproduction numbers for each age group offer a real-time gauge for interventions and behavioural changes. The inapplicability of model U to determine the reproduction number per age group and give insight into each group's behaviour also demonstrates the necessity of model A.

When modelling the epidemic using a marked Hawkes process, it is straightforward to consider more individual inhomogeneities, such as community structure (e.g. household) and location, immunization status against the disease, and medical conditions. This type of modelling reveals the dynamics of the epidemic, including who infected whom, which is challenging because of invisible transmission pathways [Yang and Zha, 2013; Kim et al., 2019].

# Chapter 7

# Conclusions and Further Work

In Chapter 3, we investigated the application of the Bayesian Additive Regression Trees (BART) model to estimate the intensity of Poisson processes. The BART framework offers a flexible non-parametric method for identifying additive and non-linear effects in the intensity's underlying functional form. With ensembles of fewer than 10 trees, our numerical results demonstrate that our technique yields accurate intensity approximations. This enables our method to capture the dimensions that are most contributing to the intensity. The simulation analysis using synthetic data sets demonstrates that our method is capable of identifying change points and can produce accurate estimations of the intensity using either the posterior mean or posterior median. In both one and two dimensions, the Haar-Fisz algorithm and kernel approaches are competitive with the proposed algorithm. Our method excels in higher dimensions, and we show that it outperforms the kernel approach for multidimensional intensities due to edge effects and is competitive with inference using spatial log-Gaussian Cox process. We also show how the variability of the rate of occurrences in both real and synthetic data is consistent with our estimated intensity.

The axis-alignment nature of the introduced algorithm in Chapter 3 is a drawback; the imposed prior on the trees is a stochastic process that takes axis-aligned cuts into account and results in partitioning the domain into hyper-rectangular regions. The simulation analysis revealed that points close to jumps have less reliable estimates, which is expected due to that drawback. In Chapter 4, we discuss this problem and propose approaches to remedy it. We illustrate using helpful rotations captures the diagonal nature of intensity boundaries, and outperforms kernel smoothing with edge correction in two dimensions. Further work in two dimensions includes the Bayesian inference of the rotation matrix and allowing trees of the ensemble to be associated with different rotation matrices. Future work also includes investigating the approach into higher-dimensional intensities.

In Chapter 5, we introduce a novel epidemic model that models the infections via a Hawkes process and the aggregated reported cases via a probability distribution with a mean driven by the underlying Hawkes process. We apply a KDPF for inferring the latent cases and the instantaneous reproduction number and for predicting the new observed cases in the near future. It is feasible to employ particle filter type algorithms, like the KDPF, because the computational effort is linear to the number of infections. Modelling the infections via a Hawkes process allows us to estimate by whom an infected individual was infected. We demonstrate the performance of the proposed algorithm on COVID-19. We demonstrate that our predicted new cases, and our inference for the latent intensity, the daily and weekly hidden cases are consistent with the observed cases in various local authorities in the UK. The simulation analysis shows that, in comparison to the approach introduced by Koyama et al. [2021], the proposed algorithm offers comparable estimates of the observed case fluctuations. Similar reproduction number estimations are given by the EpiEstim, the method of Koyama et al. [2021] and the proposed model.

Isham and Medley [1996]; Wallinga et al. [1999]; Farrington et al. [2001] contend that it is necessary to account for individual heterogeneities while modelling the transmission of an infectious disease. Individuals vary in their tendency to interact with others; personal hygiene is a key factor in the propagation of diseases; individuals' community structure and location might be significant in spreading epidemics. The proposed epidemic model introduced in Chapter 5 can be seen as a new method for constructing epidemic models that take individual heterogeneities into account and reveal underlying dynamics.

The simplest assumption of individual heterogeneity is that individual sociability varies with the age. In Chapter 6, we extend the unstructured homogeneously mixing epidemic model introduced in Chapter 5 considering a finite population stratified by age bands and modelling the epidemic process as a marked counting process. We propose a KDPF for inferring the marked counting process and forecasting the epidemic's future trajectory in the near future. We demonstrate the performance of the proposed algorithm on COVID-19. We show that the aggregated observed cases for each age group in different local authorities in the UK are consistent with the estimated latent cases and the latent intensity. The analysis also demonstrates that the instantaneous reproduction rate for each age group corresponds to the development of the epidemic and the group's changing behavioural patterns. The instantaneous reproduction numbers for each age group offer a real-time gauge for interventions and behavioural changes. The inapplicability of the model introduced in Chapter 5 to determine the reproduction number per age group and give insight into each group's behaviour also demonstrates the necessity of the age-stratified epidemic model.

It is straightforward to take into account more individual inhomogeneities when modelling the epidemic using a marked Hawkes process, such as community structure (for example, household) and location, immunisation status for the disease, and medical issues. Such modelling also estimates the dynamics of the epidemic, including who infected whom, which is difficult due to hidden transmission pathways [Yang and Zha, 2013; Kim et al., 2019]. Future work considers alternative inference methods and modelling of the processes $\gamma(t, a)$, the inference of ascertainment rate, using various transition kernels for modelling the latent and the reported infection cases, as well as more sophisticated ways for initializing the set of infectious triggering the epidemic process and the number of susceptibles at the beginning of the process for each age group.

# Appendix A

# Appendix of Chapter 3

In this appendix, we present more simulation results on synthetic and real data to illustrate the performance of Algorithm 4 on estimating the intensity of inhomogeneous Poisson processes. We describe the model considering only one tree.

## A.1 The model for the case of one tree

The proposed model for considering only one tree can be written as follows

$$\begin{aligned}\lambda(s_i) = g(s_i;T,\Lambda) = \sum_{k=1}^{b} \lambda_k \, I(s_i \in \Omega_k)\\ T \sim \text{heterogeneous Galton-Watson process for a partition of } S\\ \lambda_k|T \sim \text{Gamma}(\alpha,\beta)\end{aligned}$$

underpinned by a tree-shaped partition $T = \{\Omega_k\}_{k=1}^{b}$ where $b$ is the number of terminal nodes in the tree $T$. Each leaf node k associated to region $\Omega_k$ is linked with a parameter $\lambda_k$. All parameters $\lambda_k$ are collected in the vector $\Lambda = (\lambda_1, \lambda_2, .., \lambda_b)$. The parameters of the model are

1. the regression tree $T$
2. the parameters $\Lambda = (\lambda_1, \lambda_2, .., \lambda_b)$.

We assume that the leaf parameters are independent, i.e., $P(\Lambda|T) = \prod_{k=1}^{b} P(\lambda_k|T)$.

### A.1.1 Poisson process conditional likelihood

The conditional likelihood of a finite realization of an inhomogeneous Poisson process with $n$ points $s_1, \dots, s_n$ is derived by describing $\lambda(s)$ using one tree $(\Lambda, T)$ as: $\lambda(s) = g(s;T,\Lambda)$.

$$P(s_1,\dots,s_n|\Lambda,T) = \prod_{i=1}^{n} \lambda(s_i) \exp\left(-\int_S \lambda(s)ds\right) = \prod_{i=1}^{n} g(s_i;T,\Lambda) \exp\left(-\int_S g(s;T,\Lambda)ds\right). \tag{A.1}$$

The first term of the above equation can be written as follows

$$\prod_{i=1}^{n} g(s_i; T, \Lambda) = \prod_{k=1}^{b} \lambda_k^{n_k}$$

where $n_k$ is the cardinality of the set $\{i : s_i \in \Omega_k\}$.

The exponential term of (A.1) can be expressed as follows

$$\begin{aligned}\exp\left(-\int_S g(s; T, \Lambda) ds\right) &= \exp\left(-\int_S \sum_{k=1}^{b} \lambda_k I(s \in \Omega_k) ds\right)\\ &= \exp\left(-\sum_{k=1}^{b} \lambda_k \int_S I(s \in \Omega_k) ds\right) = \exp\left(-\sum_{k=1}^{b} \lambda_k |\Omega_k|\right)\end{aligned}$$

Hence the conditional likelihood can be written as

$$P(s_1, \ldots, s_n | \Lambda, T) = \prod_{k=1}^{b} \lambda_k^{n_k} e^{-\lambda_k |\Omega_k|}, \tag{A.2}$$

where $|\Omega_k|$ is the volume of the region $\Omega_k$.

### A.1.2 Inference Algorithm

Inference on the model parameters $(\Lambda, T)$ induces sampling from the posterior $P(\Lambda, T | s_1, \ldots, s_n)$. A Metropolis Hastings within Gibbs sampler (Algorithm 18) is proposed for sampling from the posterior $P(\Lambda, T | s_1, \ldots, s_n)$. Noting that,

$$P(\Lambda, T | s_1, \ldots, s_n) = P(\Lambda | T, s_1, \ldots, s_n)\, P(T | s_1, \ldots, s_n)$$

and

$$P(\Lambda | T, s_1, \ldots, s_n) \propto P(s_1, \ldots, s_n | \Lambda, T)\, P(\Lambda | T) \propto \prod_{k=1}^{b} \lambda_k^{n_k + \alpha - 1} e^{-(|\Omega_k| + \beta)\lambda_k},$$

a draw from $(T, \Lambda) | s_1, \ldots, s_n$ can be achieved in $(b+1)$ successive steps:

- sample $T | n, s_1, \ldots, s_n$ through Metropolis-Hastings Algorithm summarized in Algorithm 19
- sample $\lambda_k | T, n, s_1, \ldots, s_n$ from a Gamma distribution with shape equal to $n_k + \alpha$ and rate equal to $|\Omega_k| + \beta$ for $k = 1, .., b$.

Noting that

$$P(T | s_1, \ldots, s_n) \propto P(s_1, \ldots, s_n | T)\, P(T)$$

the integrated likelihood (integrating out the parameters $\Lambda$) is:

$$
\begin{aligned}
P(s_1,\ldots,s_n|T) &= \int P(s_1,\ldots,s_n,\Lambda|T)d\Lambda = \int P(s_1,\ldots,s_n|\Lambda,T)P(\Lambda|T)d\Lambda \\
&= \left(\frac{\beta^\alpha}{\Gamma(a)}\right)^b \prod_{k=1}^{b} \int \lambda_k^{n_k+\alpha-1} e^{-(|\Omega_k|+\beta)\lambda_k} d\lambda_k \\
&= \left(\frac{\beta^\alpha}{\Gamma(a)}\right)^b \prod_{k=1}^{b} \frac{\Gamma(n_k+\alpha)}{(\beta+|\Omega_k|)^{n_k+\alpha}}. \qquad \text{(A.3)}
\end{aligned}
$$

In the tree sampling Algorithm 19, the transition kernel $q$ is chosen from the three proposals: GROW, PRUNE, CHANGE [Chipman et al., 2010; Kapelner and Bleich, 2016], and Eq. (A.3) allows us to compute the Metropolis Hastings ratio to accept or reject the proposal.

**Algorithm 18** Proposed Algorithm: Metropolis Hastings within Gibbs sampler

```
for $t = 1, 2, 3, ..$ do
    Sample $T^{(t+1)}|s_1,\ldots,s_n$
    for $k = 1$ to $b$ do
        Sample $\lambda_k^{(t+1)}|s_1,\ldots,s_n,T^{(t+1)}$
    end for
end for
```

**Algorithm 19** Metropolis Hastings Algorithm for sampling from the posterior $P(T|s_1,\ldots,s_n)$

```
Generate a candidate value $T^*$ with probability $q(T^*|T^{(t)})$.
Set $T^{(t+1)} = T^*$ with probability
    $\alpha(T^{(t)},T^*) = \min\left(1, \frac{q(T^{(t)}|T^*)}{q(T^*|T^{(t)})}\frac{P(s_1,\ldots,s_n|T^*)}{P(s_1,\ldots,s_n|T^{(t)})}\frac{P(T^*)}{P(T^{(t)})}\right)$
Otherwise, set $T^{(t+1)} = T^{(t)}$.
```

## A.2 Simulation results on synthetic data with various number of sampling iterations

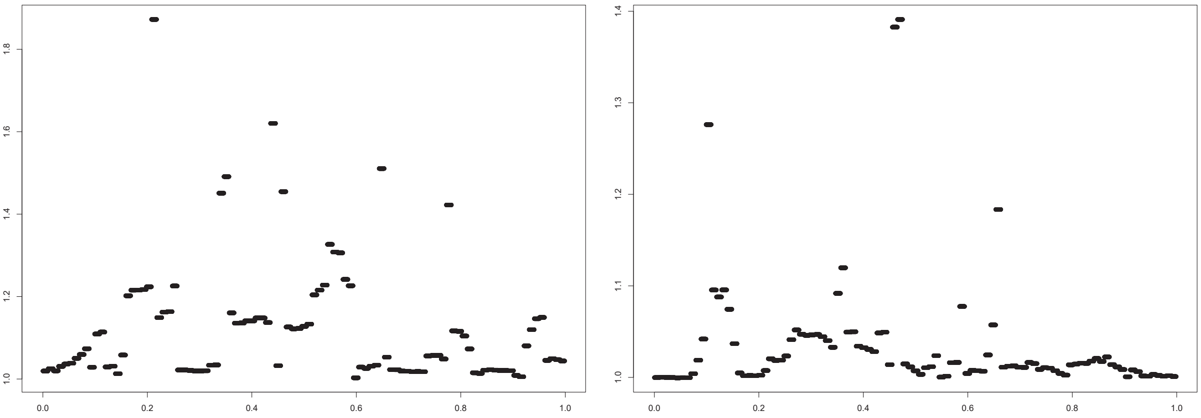
In this appendix we show that our algorithm works equally well for 10000 iterations by running three parallel chains, examining their convergence and assessing the performance of our algorithm via AAE and RMSE of computed estimates over various number of iterations. We also check the convergence of chains using the Gelman-Rubin criterion in all cases.

### A.2.1 One dimensional Poisson process with stepwise intensity

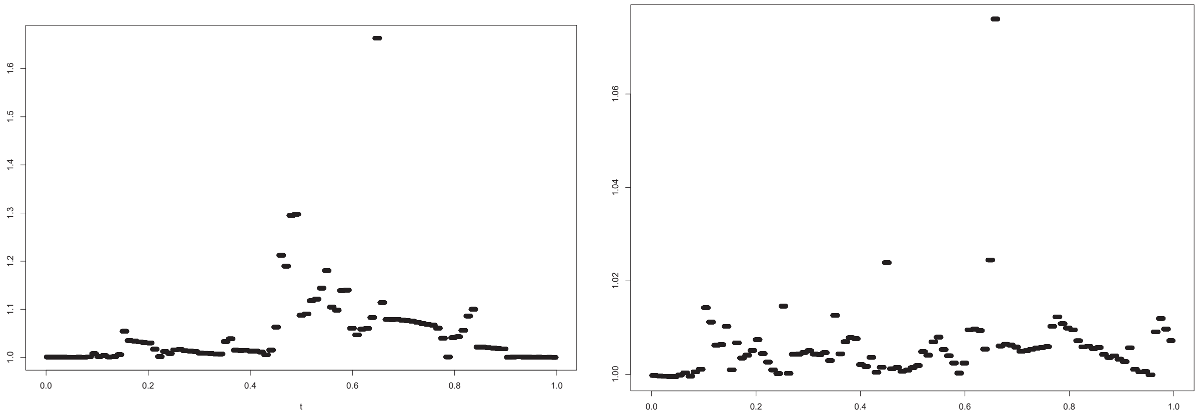
Table A.1 shows that there are no significant difference in errors increasing the number of iterations from 10000 to 200000. Figure A.1 reveals that the chains work less well at points close to jumps for small number of iterations.

| Proposed Algorithm | | | | | |
|---|---|---|---|---|---|
| Number of trees | Number of Iterations | AAE for Posterior Mean | AAE for Posterior Median | RMSE for Posterior Mean | RMSE for Posterior Median |
| 5 | 10000 | 284.61 | 274.3 | 588.88 | 590.5 |
| | 50000 | 289.11 | 284.56 | 575.11 | 579.17 |
| | 200000 | 279.88 | 269.81 | 572.94 | 576.94 |
| 7 | 10000 | 265.22 | 257.49 | 572.33 | 576.58 |
| | 50000 | 276.19 | 267.75 | 580.35 | 584.47 |
| | 200000 | 278.37 | 269.78 | 582.82 | 584.1 |

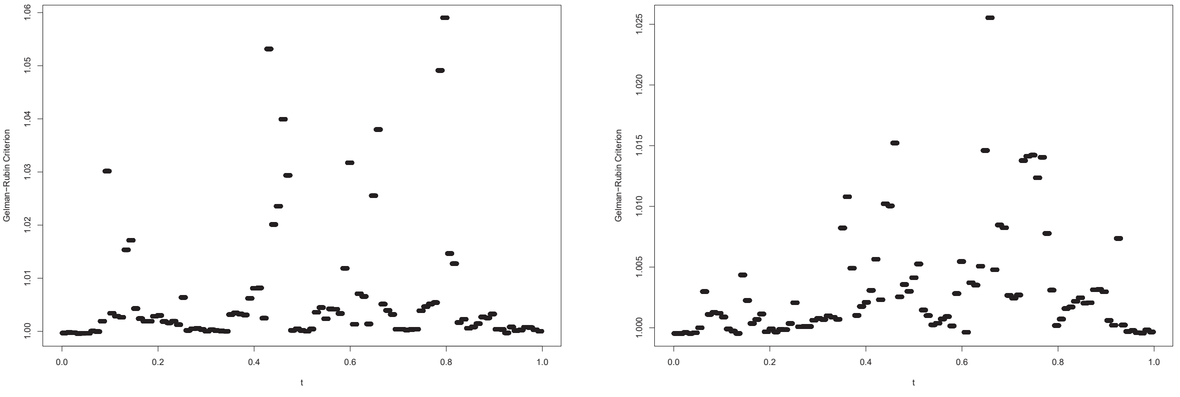
Table A.1: Average Absolute Error and Root Mean Square Error for various number of iterations and trees.



(a) 5 Trees and 10000 iterations (b) 7 Trees and 10000 iterations



(c) 5 Trees and 50000 iterations (d) 7 Trees and 50000 iterations



(e) 5 Trees and 200000 iterations (f) 7 Trees and 200000 iterations

Figure A.1: The Gelman-Rubin Criterion for various number of iterations and trees.

### A.2.2 One dimensional Poisson process with continuously varying intensity

Table A.2 shows that increasing the number of iterations does not change essentially the error for the synthetic data presented in Section A.4.1. The convergence criterion indicates that even for small number of iterations, the chains converge for 10 trees.

| Proposed Algorithm | | | | | |
|---|---|---|---|---|---|
| Number of trees | Number of Iterations | AAE for Posterior Mean | AAE for Posterior Median | RMSE for Posterior Mean | RMSE for Posterior Median |
| 5 | 10000 | 6.27 | 6.71 | 9.83 | 10.62 |
| | 50000 | 6.16 | 6.51 | 9.63 | 10.42 |
| | 100000 | 6.14 | 6.38 | 9.52 | 10.17 |
| 7 | 10000 | 5.99 | 6.03 | 9.54 | 9.95 |
| | 50000 | 6.04 | 6.1 | 9.49 | 9.88 |
| | 100000 | 5.95 | 6.01 | 9.39 | 9.8 |

Table A.2: Average Absolute Error and Root Mean Square Error for various number of iterations and trees.

| Proposed Algorithm | | | | | |
|---|---|---|---|---|---|
| Number of trees | Number of Iterations | AAE for Posterior Mean | AAE for Posterior Median | RMSE for Posterior Mean | RMSE for Posterior Median |
| 4 | 10000 | 241.82 | 240.1 | 464.99 | 489.93 |
| | 50000 | 209.95 | 209.58 | 392.43 | 418.37 |
| | 100000 | 208.74 | 213.04 | 410.19 | 447.86 |

Table A.3: Average Absolute Error and Root Mean Square Error for 4 Trees and various number of iterations.

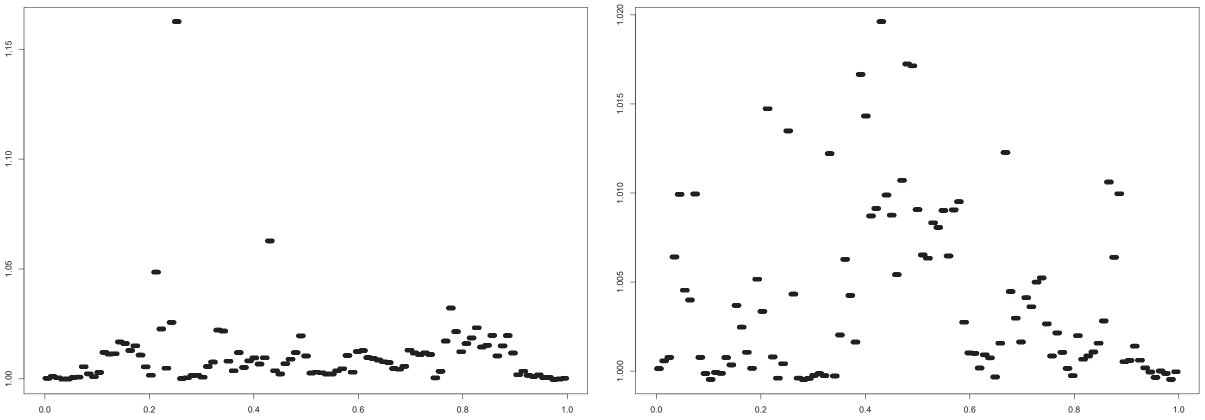
For 5 trees they converge for the majority of the range (Figure A.2).

### A.2.3 Two dimensional Poisson process with stepwise intensity function

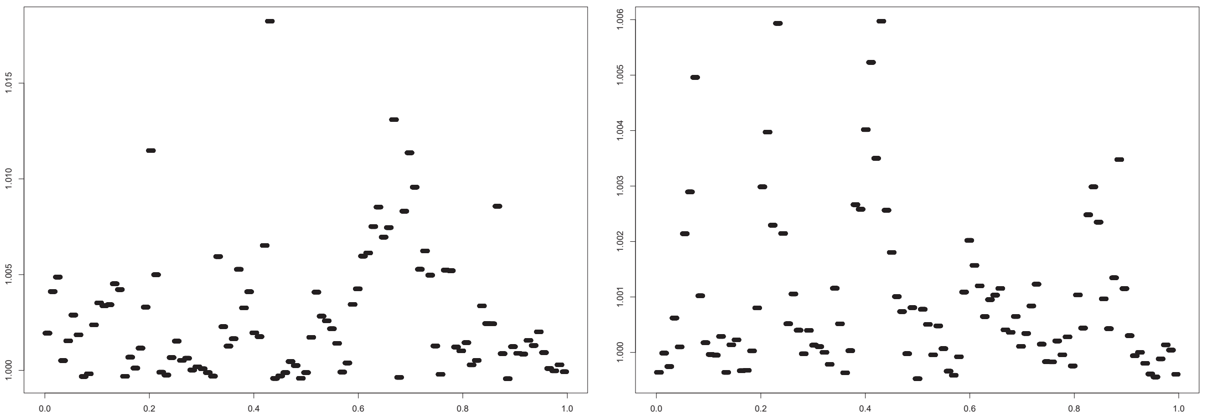
Likewise, we do not observe significant improvement in AAE and RMSE beyond 10000 iterations (see Table A.3). Moreover, increasing the number of iterations does not fix the convergence issues at points close to jumps (see Figure A.3).

### A.2.4 Inhomogeneous two dimensional Poisson process with Gaussian intensity

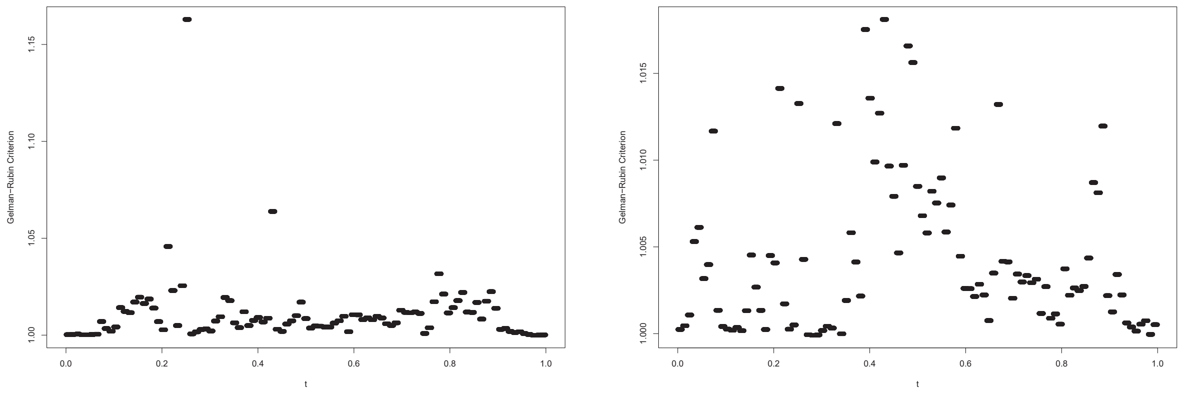
Similarly to all the above scenarios, the error with 10000 iterations are comparable to those obtained with a larger number of iterations (see Table A.4). Figure A.4 shows that the chains converge for 10 Trees even if we consider a relatively small number of iterations. The same holds for the majority of testing points for 8 Trees. The algorithm only provides less accurate estimations for the testing points close to the upper end of the domain for 8 Trees and relatively small number of iterations.



(a) 5 Trees and 10000 iterations (b) 10 Trees and 10000 iterations

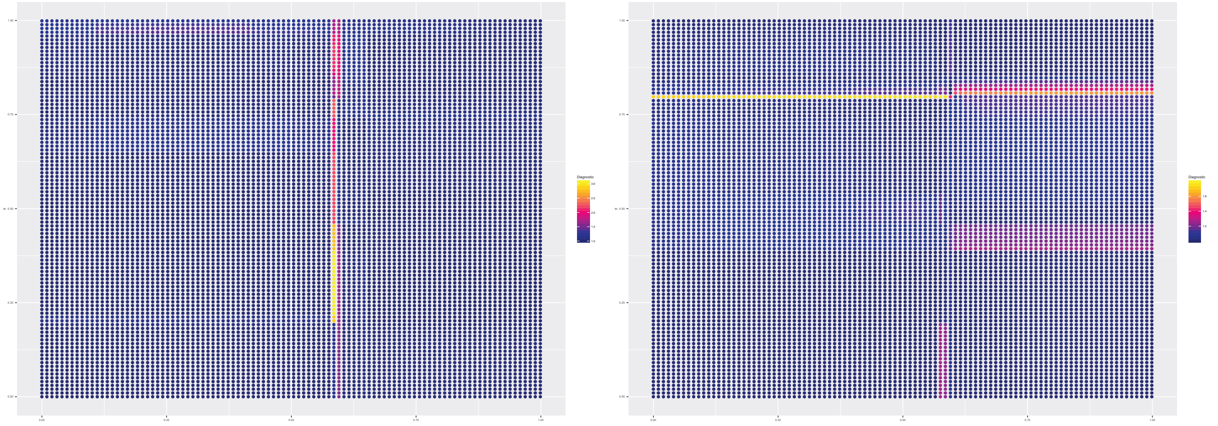


(c) 5 Trees and 50000 iterations (d) 10 Trees and 50000 iterations

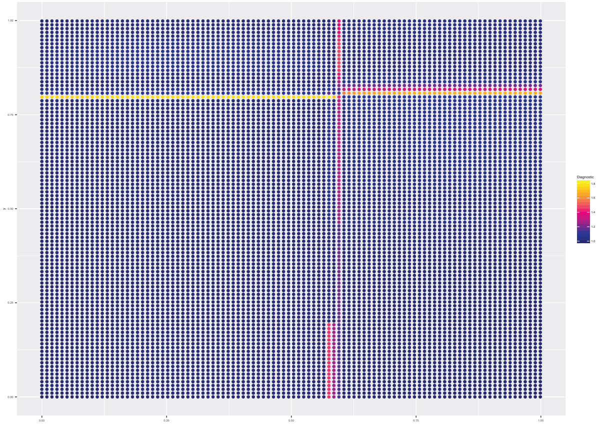


(e) 5 Trees and 100000 iterations (f) 10 Trees and 100000 iterations

Figure A.2: The Gelman-Rubin Criterion for various number of iterations and trees.



(a) 4 Trees and 10000 iterations (b) 4 Trees and 50000 iterations

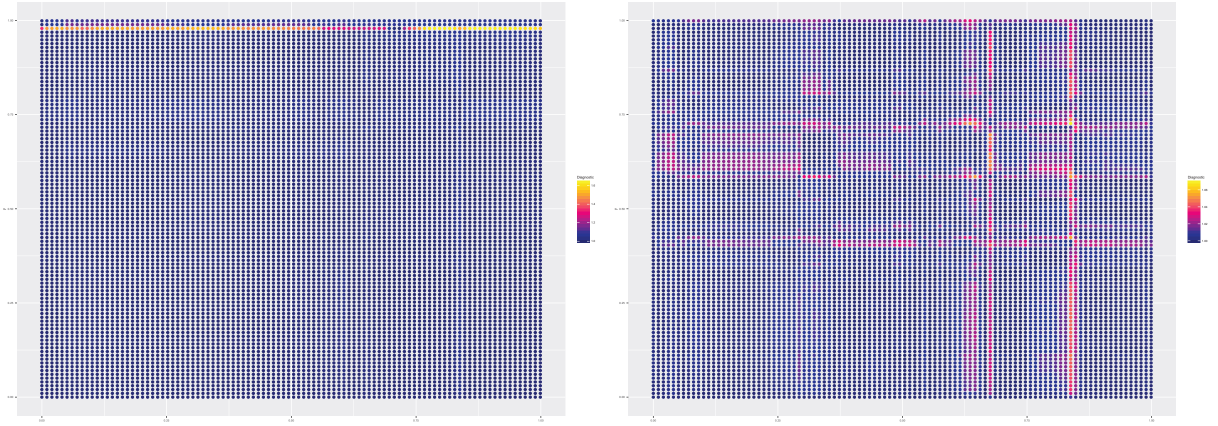


(c) 4 Trees and 100000 iterations

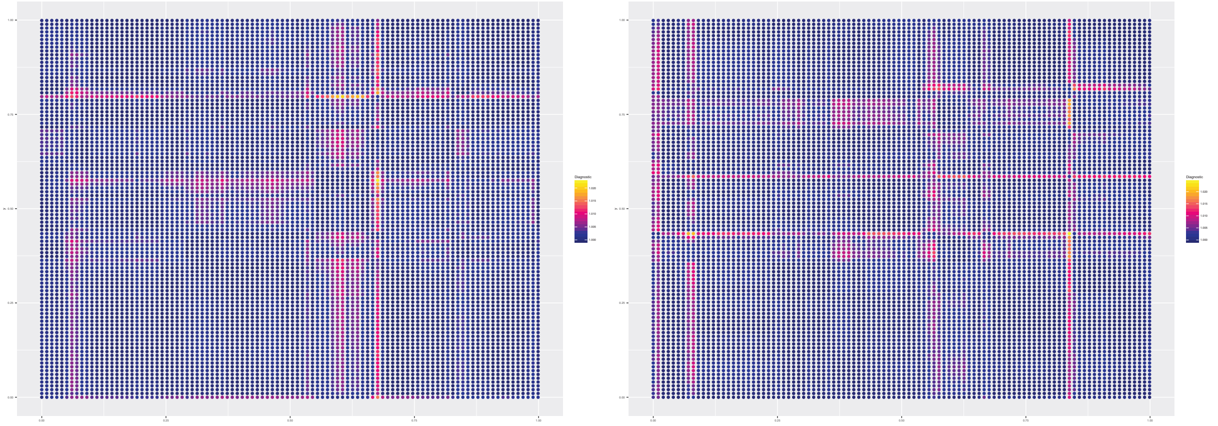
Figure A.3: The Gelman-Rubin Criterion for 4 trees and various number of iterations.

| Proposed Algorithm | | | | | |
|---|---|---|---|---|---|
| Number of trees | Number of Iterations | AAE for Posterior Mean | AAE for Posterior Median | RMSE for Posterior Mean | RMSE for Posterior Median |
| 8 | 10000 | 173.02 | 175.61 | 247.5 | 255.81 |
| | 50000 | 169.54 | 170.5 | 242.03 | 250.74 |
| | 200000 | 177.44 | 175.62 | 255.23 | 258.88 |
| 10 | 10000 | 168.91 | 168.78 | 242.62 | 249.38 |
| | 50000 | 177.72 | 173.93 | 254.67 | 256.32 |
| | 200000 | 176.52 | 174.02 | 253.14 | 255.92 |

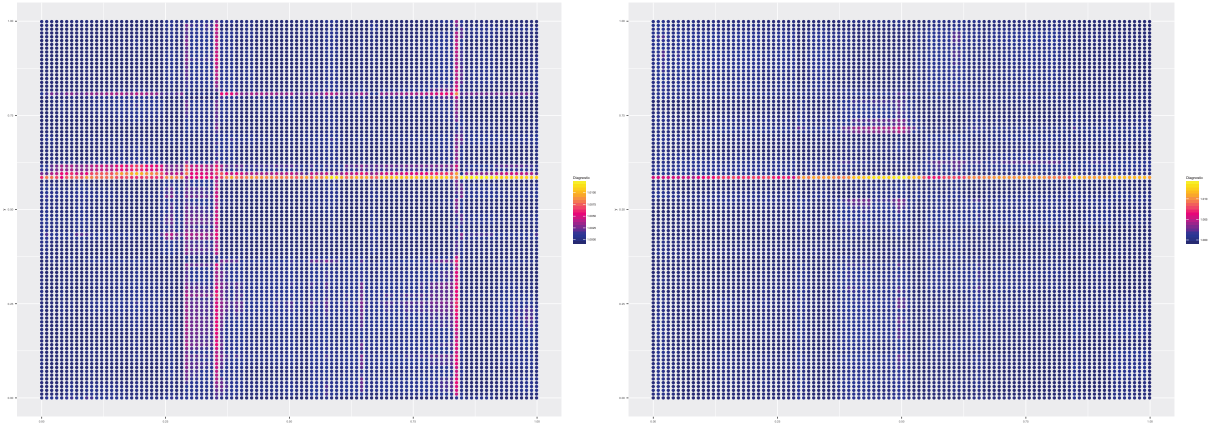
Table A.4: Average Absolute Error and Root Mean Square Error for various number of iterations and trees.



(a) 8 Trees and 10000 iterations (b) 10 Trees and 10000 iterations



(c) 8 Trees and 50000 iterations (d) 10 Trees and 50000 iterations



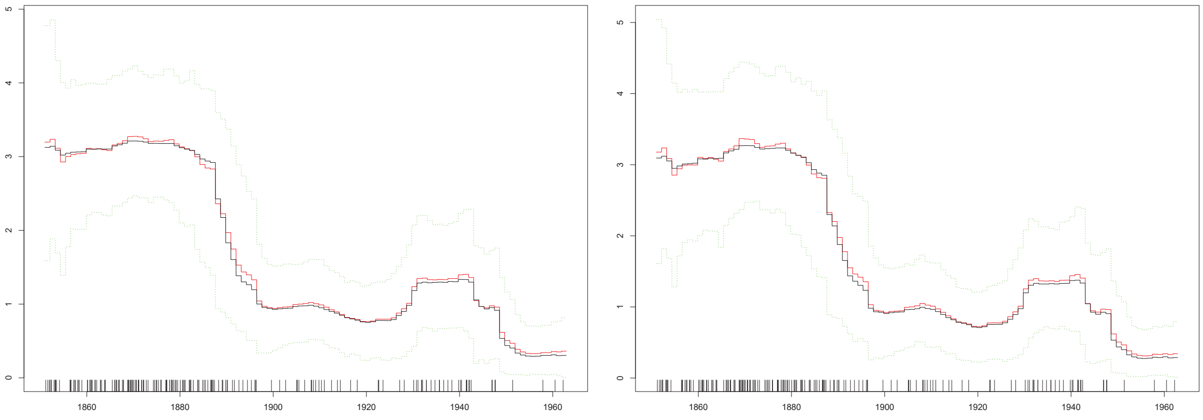
(e) 8 Trees and 200000 iterations (f) 10 Trees and 200000 iterations

Figure A.4: The Gelman-Rubin Criterion for various number of iterations and trees.

## A.3 Intensity estimation for real data

### A.3.1 Coal Data

The first real data set under consideration is composed of the dates of 191 explosions which caused at least 10 occurrences of death from March 22, 1962 until March 15, 1981. The data set is available in the **R** package **boot** [Canty and Ripley, 2019] as **coal**. Figure A.5 illustrates the Posterior Mean and the Posterior Median for 8 and 10 Trees. We observe that our algorithm captures the fluctuations of the rate of accidents in the period under consideration. The diagnostic criteria included in Appendix A.6 indicate that the considered chains have converged. See Adams et al. [2009], Gugushvili et al. [2018] and Lloyd et al. [2015] for alternative analyses.



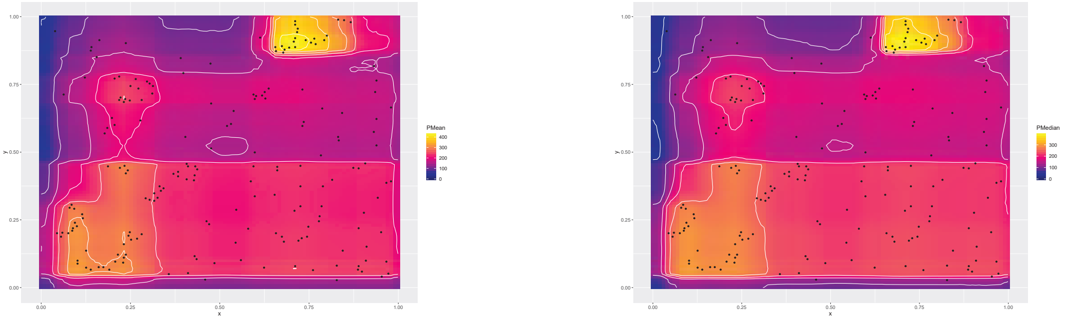
(a) 8 Trees

(b) 10 Trees

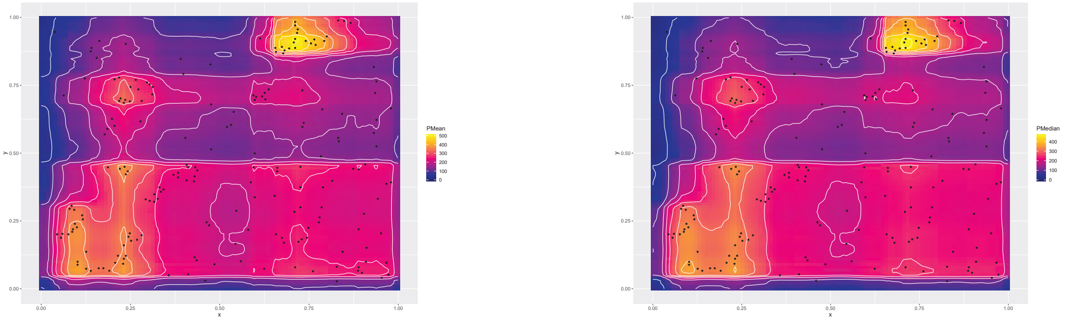
Figure A.5: Coal Data: The posterior mean (red curve), the posterior median (black curve), the 95% hdi interval of the estimated intensity illustrated by the dotted green lines. The rug plot on the bottom displays the event times.

### A.3.2 Redwoodfull Data

Finally, we use a data set available in the **R** package **spatstat** describing the locations of 195 trees in a square sampling region shown with dots in the figures below. Adams et al. [2009] analyzed the **redwoodfull** data using their recommended algorithm. We present the posterior mean and the posterior median obtained with our algorithm for different number of trees and the result of kernel estimators. Intensity inference via posterior mean (Figure A.6c) or posterior median (Figure A.6d) for 10 Trees is similar to the fixed-bandwidth kernel estimator with edge correction and bandwidth selected using likelihood cross-validation (Figure A.7a), and the inference from Adams et al. [2009].

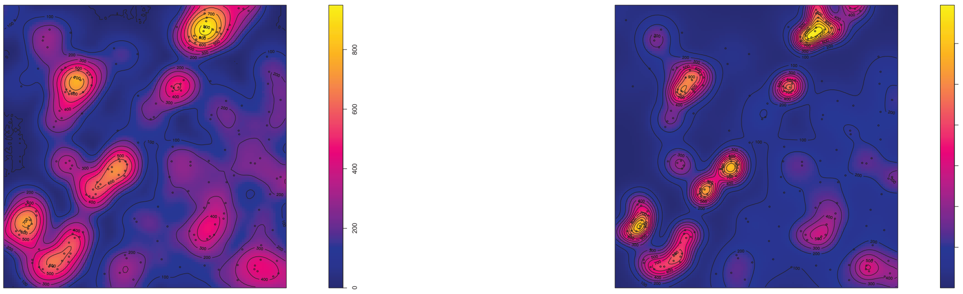


(a) Posterior Mean for 5 Trees (b) Posterior Median for 5 Trees



(c) Posterior Mean for 10 Trees (d) Posterior Median for 10 Trees

Figure A.6: Posterior Mean and Posterior Median for 3, 5 and 10 Trees



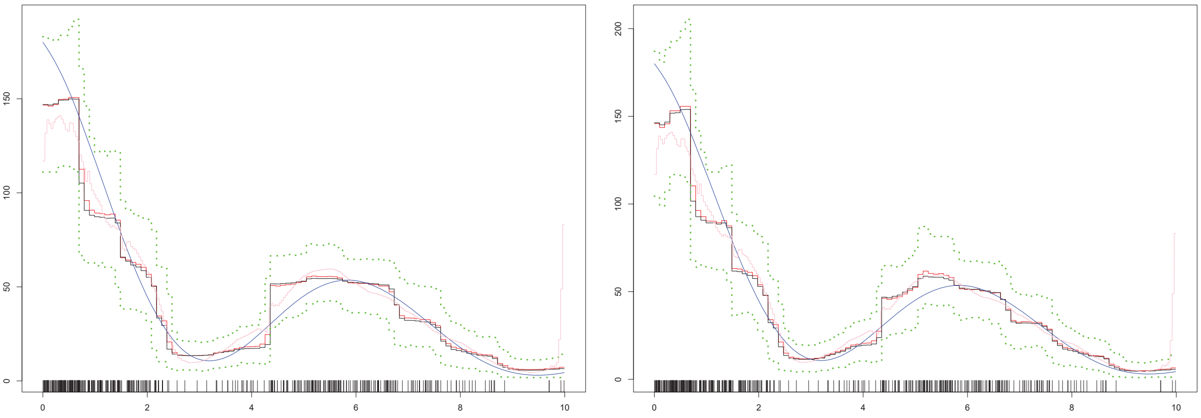
(a) Fixed-bandwidth estimator (b) Adaptive-bandwidth estimator

Figure A.7: Fixed-bandwidth chosen using likelihood cross-validation and adaptive-bandwidth kernel estimators.

# A.4 Simulation study on synthetic data

## A.4.1 One dimensional Poisson process with continuously varying intensity

We have applied our algorithm to samples of a one dimensional Poisson process with intensity $\lambda(x) = 20e^{-x/5}(5 + 4\cos(x))$ for $x \in [0, 10]$. Figure A.8 and Tables A.5-A.6 show that the algorithm works well on a smoothy varying intensity with fewer sample points and outperforms the Haar-Fisz Estimator for the majority the range. The convergence criteria indicate convergence of the simulated chains for 10 Trees and for the most testing points for 5 Trees (see supplementary material).



(a) 5 Trees (b) 10 Trees

Figure A.8: Scenario 2: The original intensity (blue curve), the posterior mean (red curve), the posterior median (black curve), the 95% hdi interval of the estimated intensity illustrated by the dotted green lines and the Haar-Fisz estimator (pink curve). The rug plot on the bottom displays the 440 event times.

| Proposed Algorithm | | | | |
|---|---|---|---|---|
| Number of trees | AAE for Posterior Mean | AAE for Posterior Median | RISE for Posterior Mean | RISE for Posterior Median |
| 5 | 6.14 | 6.38 | 9.52 | 10.17 |
| 10 | 5.95 | 6.01 | 9.39 | 9.8 |

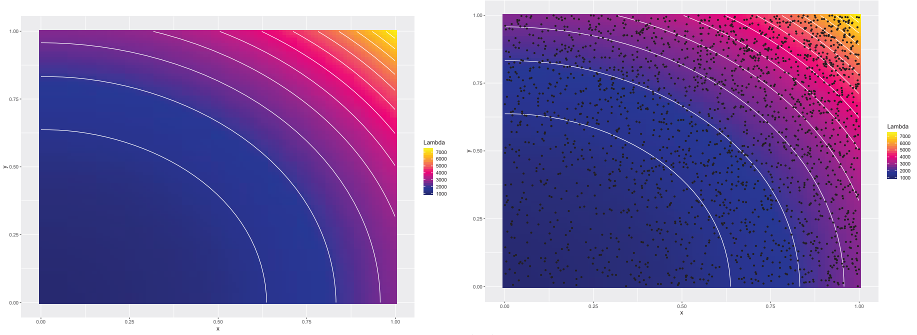
Table A.5: Average Absolute Error and Root Integrated Square Error for various number of trees for the data in Fig. A.8.

| Haar-Fisz Algorithm | |
|---|---|
| AAE | RISE |
| 7.16 | 11.67 |

Table A.6: Average Absolute Error and Root Integrated Square Error for Haar-Fisz estimator for the data in Fig. A.8

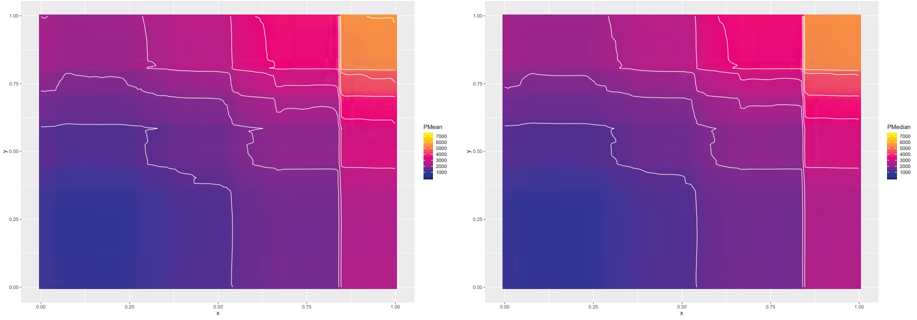
### A.4.2 Inhomogeneous two-dimensional Poisson process with Gaussian intensity

We also considered a two dimensional Poisson process with intensity $\lambda(x, y) = 1000\, e^{x^2+y^2}$ for $x, y \in [0, 1)$. The outcomes of the algorithm, log-Gaussian Cox processes (LGCP) and kernel smoothing are illustrated in Figures A.9-A.10 and Tables A.7-A.9. The results demonstrate that the proposed algorithm performs well in this setting, is competitive with the kernel method, and spatial log-Gaussian Cox processes. In this scenario, the hyperparameter $\beta$ has been set equal to 1. The convergence criteria indicate convergence of the simulated chains (see Appendix A.5).



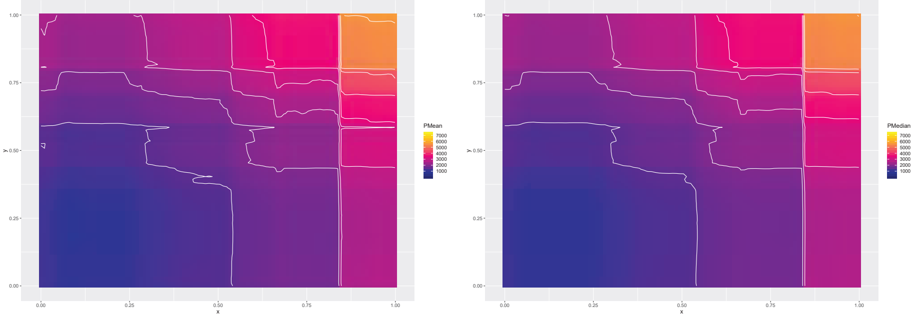
(a) Original Intensity

(b) Realization of Process consisting of 2176 points



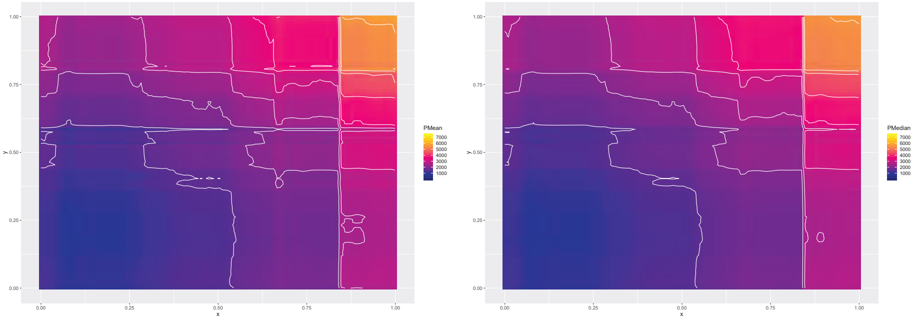
(c) Posterior Mean for 8 Trees

(d) Posterior Median for 8 Trees



(e) Posterior Mean for 10 Trees

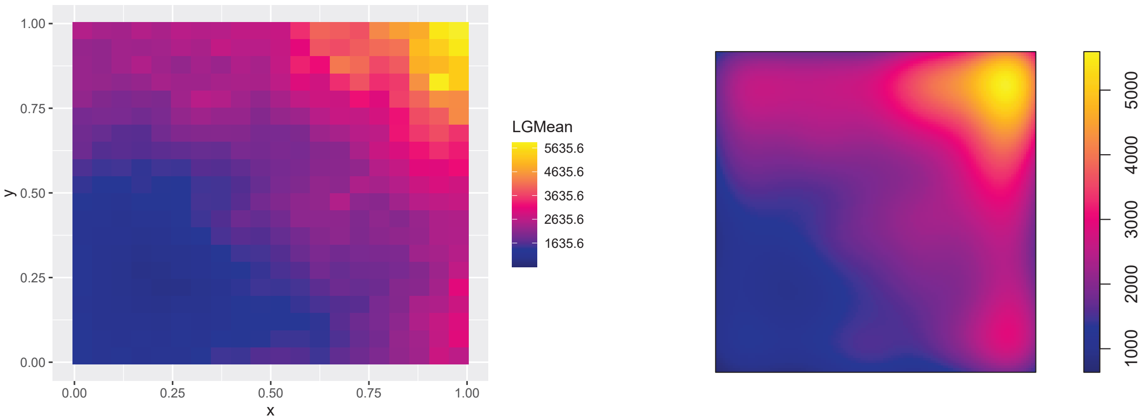
(f) Posterior Median for 10 Trees



(g) Posterior Mean for 15 Trees

(h) Posterior Median for 15 Trees

Figure A.9: Posterior Mean and Posterior Median for 8, 10 and 15 Trees



(a) LGCP

(b) Kernel Smoothing with $h = 0.087$.

Figure A.10: Kernel estimator and inference with spatial log-Gaussian Cox processes.

| Inference with spatial log-Gaussian Cox processes | | |
|---|---|---|
| grid | AAE | RISE |
| $10 \times 10$ | 195 | 263 |
| $20 \times 20$ | 182 | 245 |

Table A.7: Average Absolute Error and Root Integrated Square Error with LGCP for the data in Figure A.9.

| Proposed Algorithm | | | | |
|---|---|---|---|---|
| Number of trees | AAE for Posterior Mean | AAE for Posterior Median | RISE for Posterior Mean | RISE for Posterior Median |
| 8 | 177.44 | 175.62 | 255.23 | 258.88 |
| 10 | 176.52 | 174.02 | 253.14 | 255.92 |
| 15 | 177.48 | 172.62 | 254.22 | 251.96 |

Table A.8: Average Absolute Error and Root Integrated Square Error for various number of trees for data in Fig. A.9.

| Kernel Smoothing | | |
|---|---|---|
| Bandwidth (sigma) | AAE | RISE |
| 0.03 | 360.11 | 463.1 |
| 0.04 | 277.89 | 353.82 |
| 0.087 (LCV) | 167.74 | 227.85 |
| 0.095 | 166.51 | 230.27 |

Table A.9: Average Absolute Error and Root Integrated Square Error for fixed bandwidth estimators for data in Fig. A.9.

### A.4.3 Inhomogeneous five dimensional Poisson process with Gaussian intensity

Our next example is a five dimensional Poisson process with intensity $\lambda(x) = 50e^{x^T x}$ for $x \in [0,1)^5$ and the generated process via thinning consists of 343 points. The statistics are presented in Tables A.10 and A.11 for our algorithm and kernel smoothing, respectively. We have checked that the Gelman-Rubin criterion indicates convergence of chains.

| Proposed Algorithm | | | | |
|---|---|---|---|---|
| Number of trees | AAE for Mean | AAE for Median | RISE for Mean | RISE for Median |
| 8 | 65.99 | 66.68 | 104.24 | 106.19 |
| 10 | 66.94 | 67 | 106.09 | 106.36 |

Table A.10: Average Absolute Error and Root Integrated Square Error with different number of trees.

| Kernel Smoothing | | |
|---|---|---|
| $h$ | AAE | RISE |
| 0.03 | 631.1 | 5060.9 |
| 0.06 | 409.6 | 825.1 |
| 0.08 | 287.8 | 419.8 |
| 0.1 | 213.5 | 295.6 |
| 0.15 (LCV) | 181.2 | 278.5 |
| 0.3 | 258.4 | 363 |
| 0.5 | 311.2 | 409.7 |

Table A.11: Average Absolute Error and Root Integrated Square Error for various isotropic variance matrices.

## A.5 Additional simulation results on synthetic data

We provide further simulation results including trace and autocorrelation plots, the Gelman-Rubin Criterion in a variety of scenarios with established ground truth to demonstrate the performance of Algorithm 4.

## A.5.1 One dimensional Poisson process with stepwise intensity

| Convergence Criterion ($R_{hat}$) | |
|---|---|
| Number of Trees | $R_{hat} < 1.1$ |
| 3 | 87% |
| 4 | 98% |
| 5 | 100% |
| 7 | 100% |
| 8 | 100% |
| 9 | 100% |
| 10 | 100% |
| 20 | 100% |

Table A.12: The percentage of testing points having convergence criterion smaller than 1.1.

### A.5.1.1 5 Trees

We run 3 parallel chains each for 200000 iterations keeping every 100th sample.

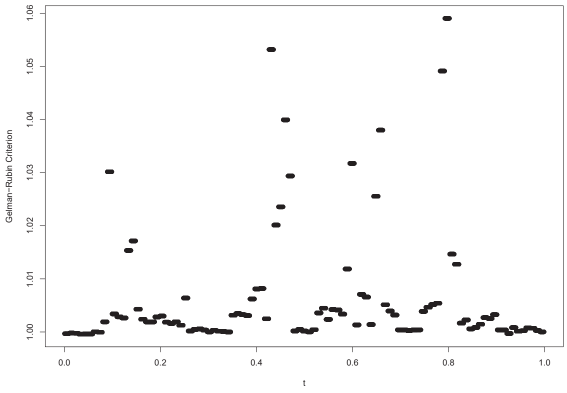


Figure A.11: The Gelman-Rubin Criterion for 5 Trees

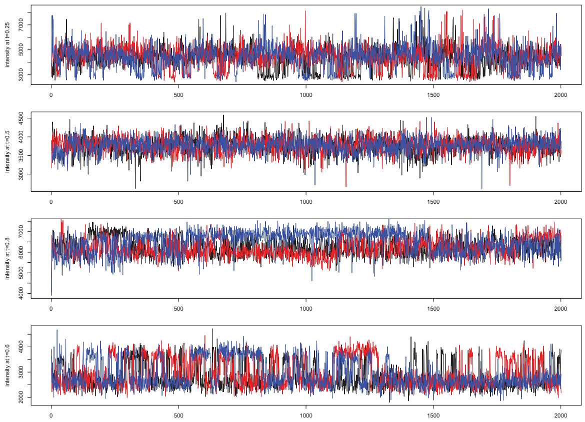


Figure A.12: Trace plots for 5 Trees

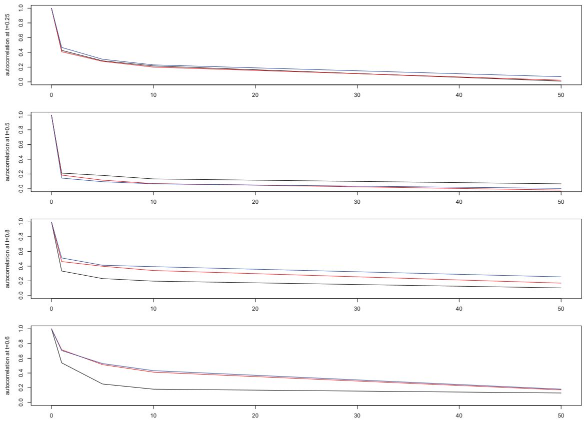


Figure A.13: Autocorrelation plots for 5 Trees

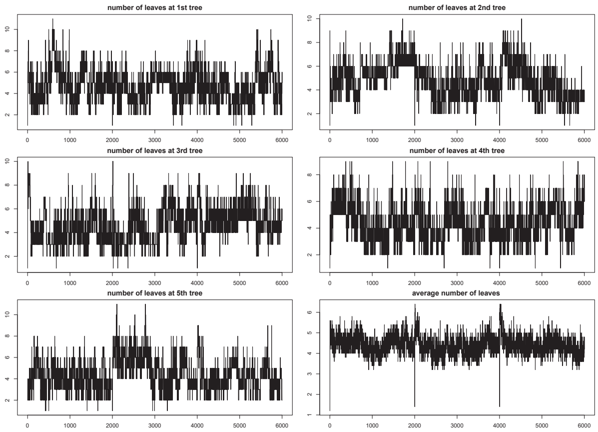


Figure A.14: Average number of leaves at trees

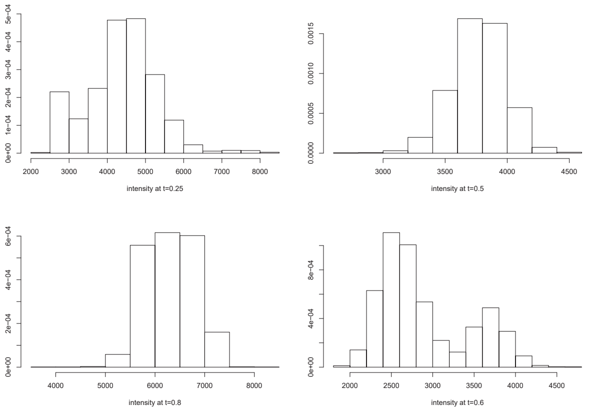


Figure A.15: Density of the estimated intensity for 5 Trees

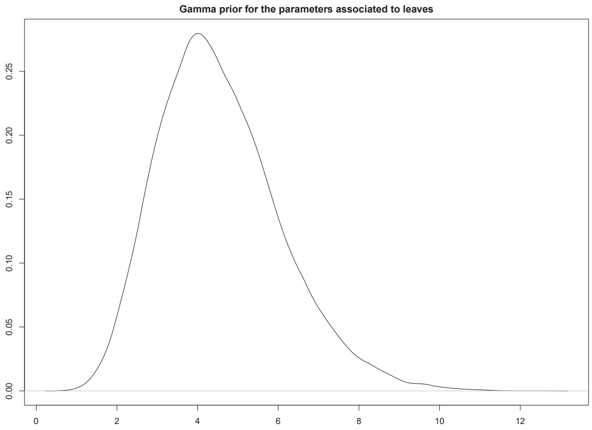


Figure A.16: Prior for 5 Trees

#### A.5.1.2 7 Trees

We run 3 parallel chains each for 200000 iterations keeping every 100th sample.

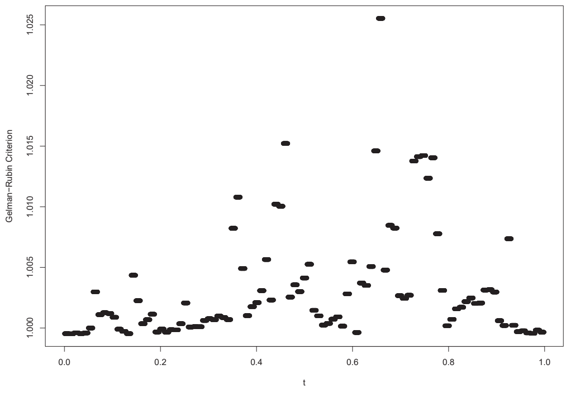


Figure A.17: The Gelman-Rubin Criterion for 7 Trees

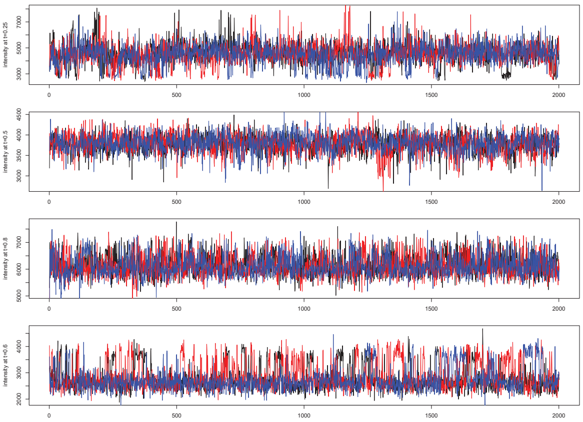


Figure A.18: Trace plots for 7 Trees

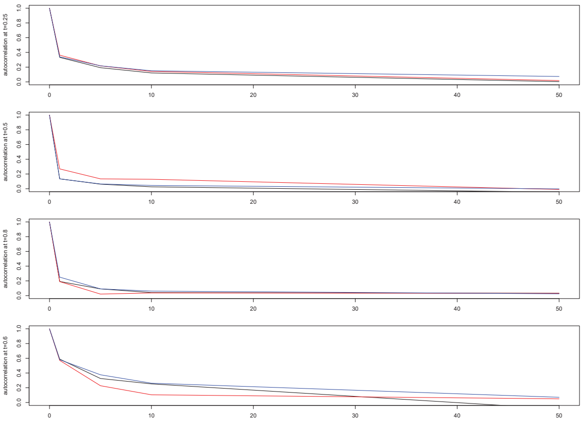


Figure A.19: Autocorrelation plots for 7 Trees

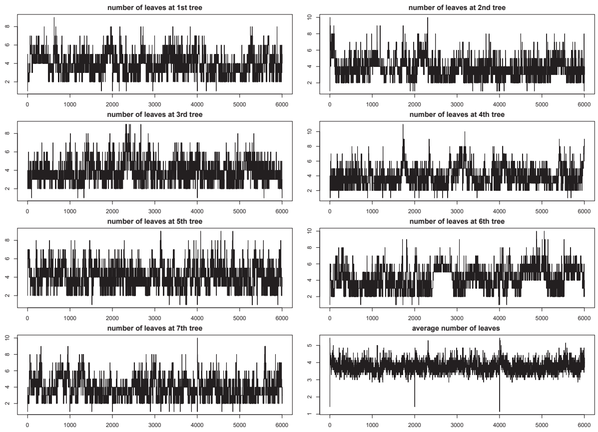


Figure A.20: Average number of leaves at trees

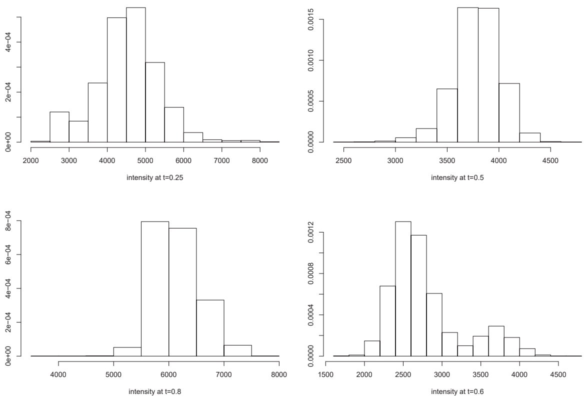


Figure A.21: Density of the estimated intensity for 7 Trees

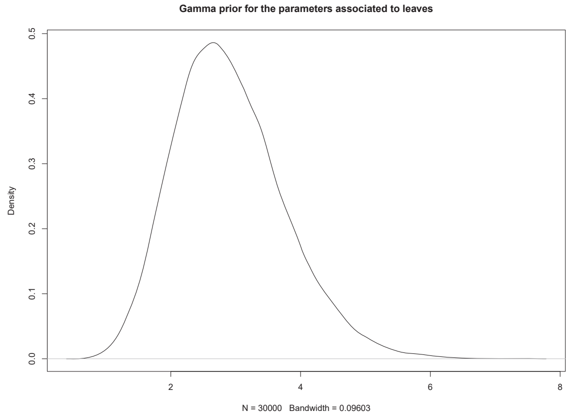


Figure A.22: Prior for 7 Trees

## A.5.2 One dimensional Poisson process with with continuously varying intensity

### A.5.2.1 5 Trees

We run 3 parallel chains each for 100000 iterations keeping every 50th sample.

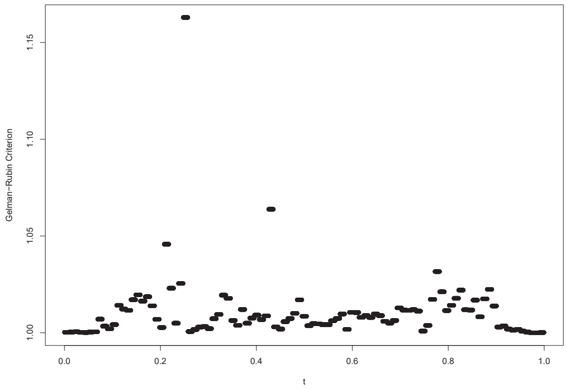


Figure A.23: The Gelman-Rubin Criterion for 5 Trees

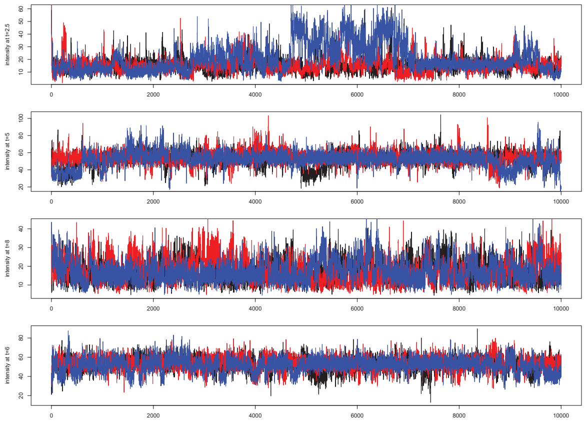


Figure A.24: Trace plots for 5 Trees

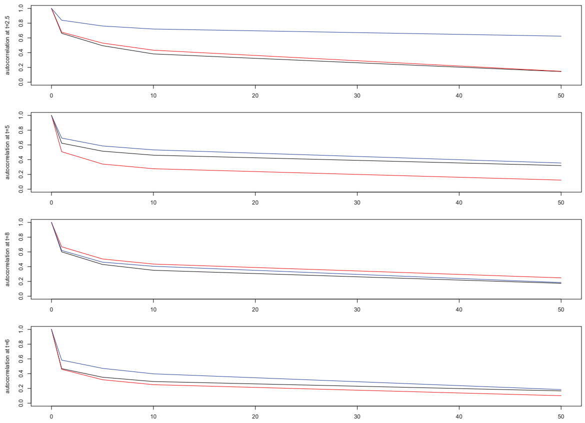


Figure A.25: Autocorrelation plots for 5 Trees

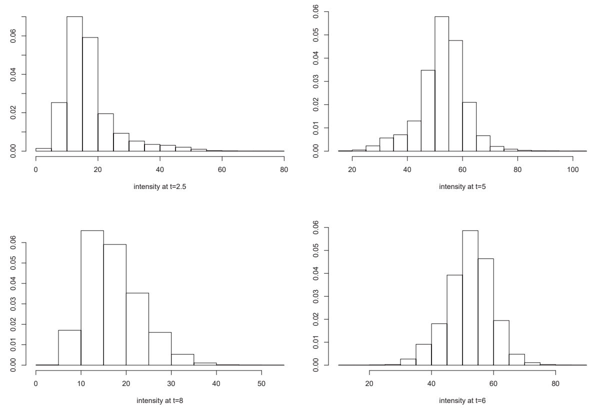


Figure A.26: Density of the estimated intensity for 5 Trees

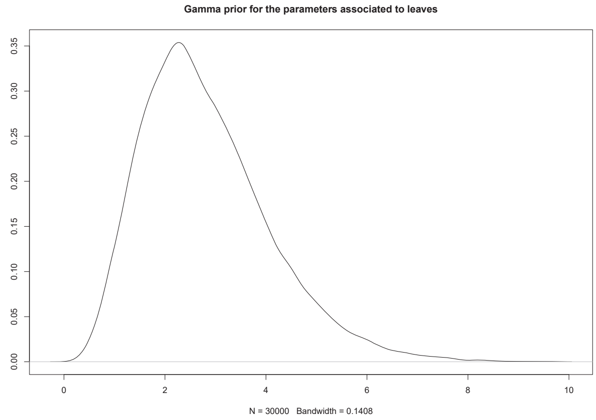


Figure A.27: Prior for 5 Trees

### A.5.2.2 10 Trees

We run 3 parallel chains each for 100000 iterations keeping every 50th sample.

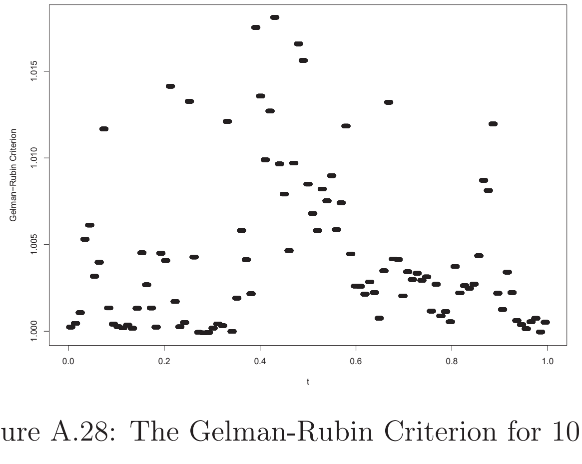


Figure A.28: The Gelman-Rubin Criterion for 10 Trees

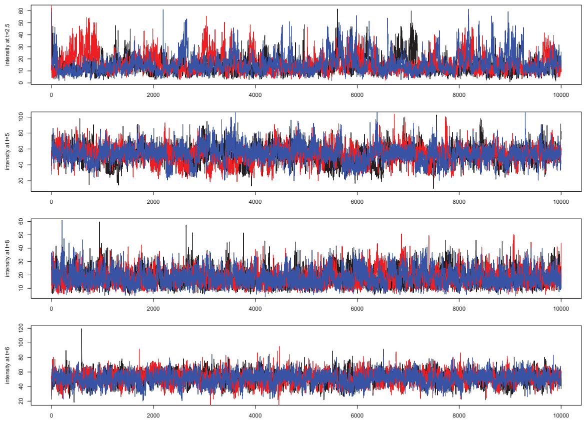


Figure A.29: Trace plots for 10Trees

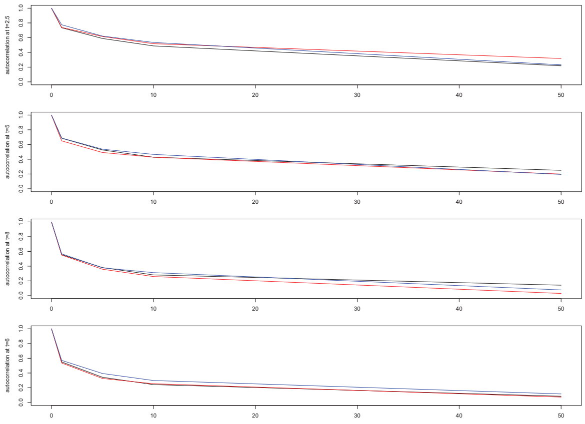


Figure A.30: Autocorrelation plots for 10 Trees

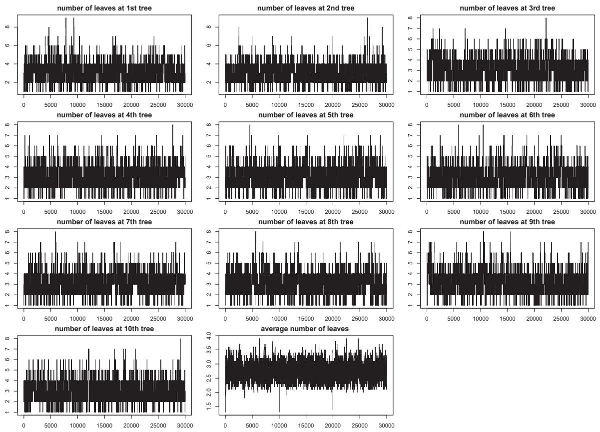


Figure A.31: Average number of leaves at trees

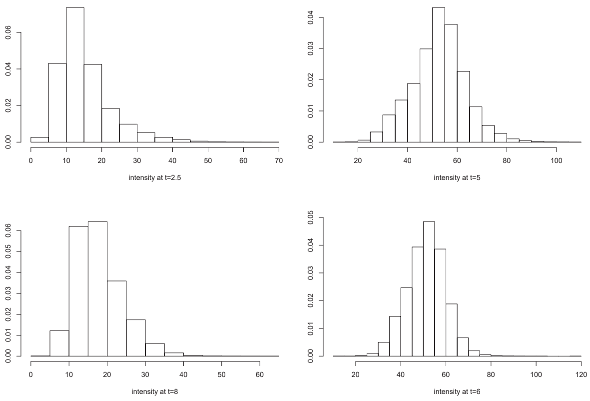


Figure A.32: Density of the estimated intensity for 10 Trees

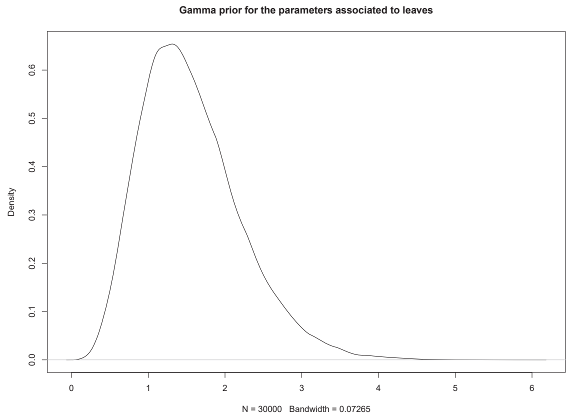


Figure A.33: Prior for 10 Trees

## A.5.3 Inhomogeneous two dimensional Poisson process with Gaussian intensity

### A.5.3.1 8 Trees

We run 3 parallel chains each for 200000 iterations keeping every 100th sample.

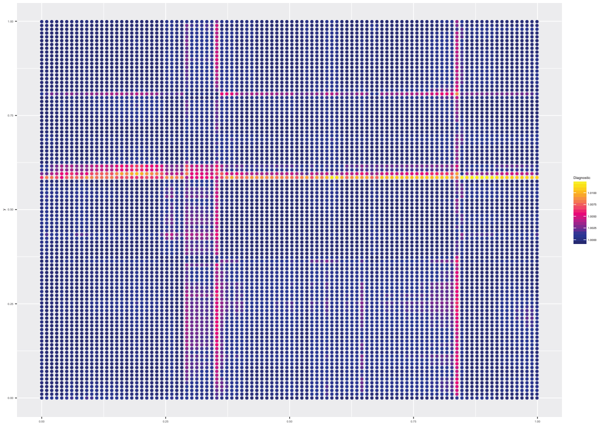


Figure A.34: The Gelman-Rubin Criterion for 8 Trees

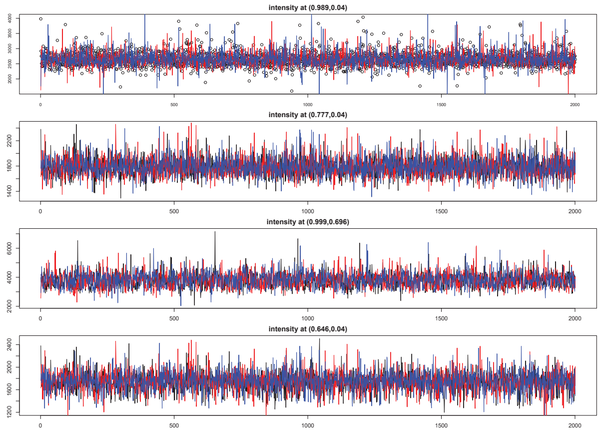


Figure A.35: Trace plots for 8 Trees

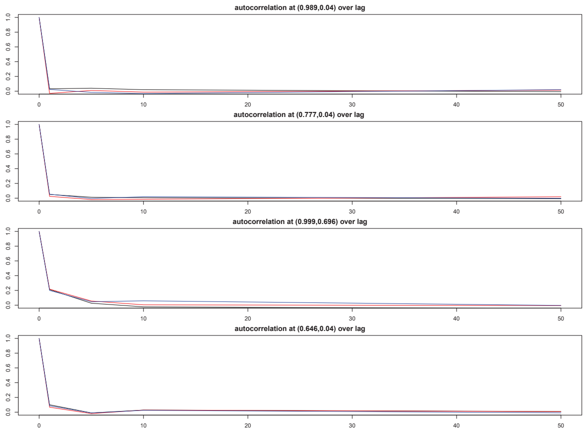


Figure A.36: Autocorrelation plots for 8 Trees

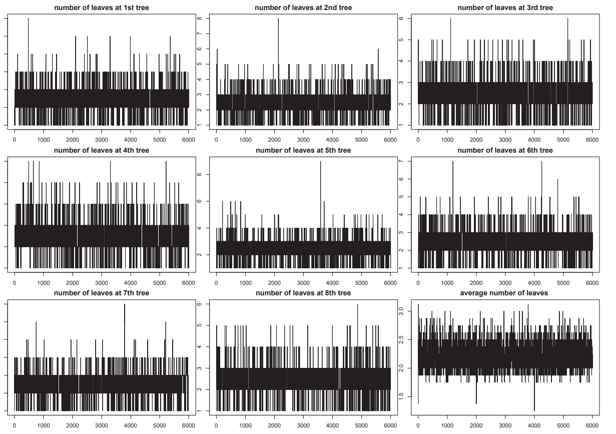


Figure A.37: Average number of leaves at trees

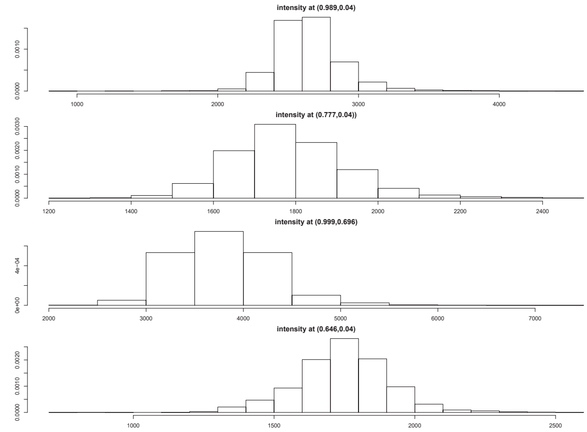


Figure A.38: Density of the estimated intensity for 8 Trees

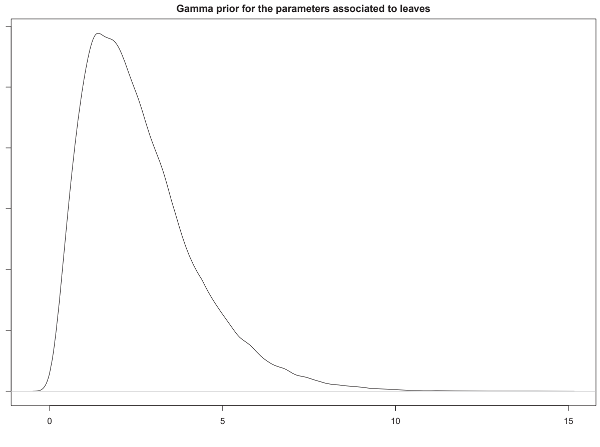


Figure A.39: Prior for 8 Trees

#### A.5.3.2 10 Trees

We run 3 parallel chains each for 200000 iterations keeping every 100th sample.

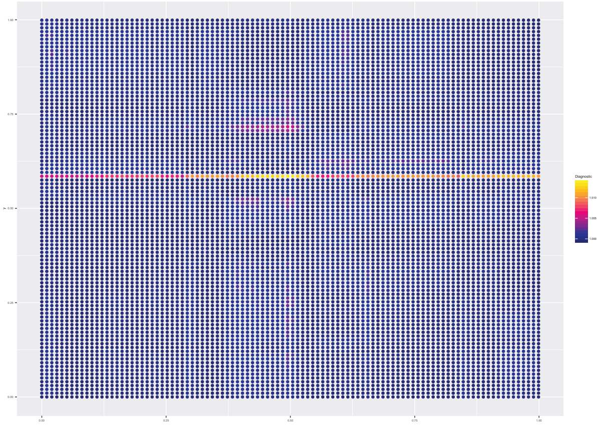


Figure A.40: The Gelman-Rubin Criterion for 10 Trees

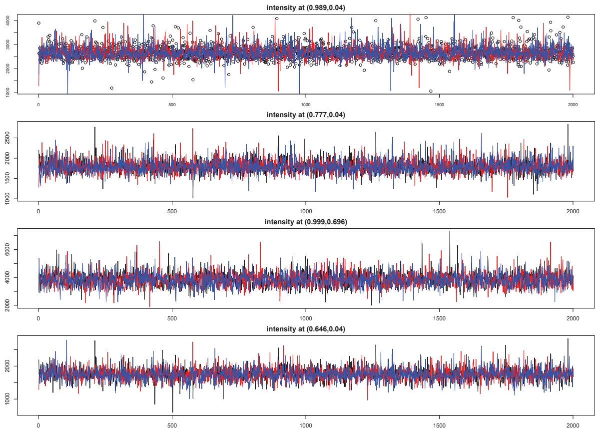


Figure A.41: Trace plots for 10 Trees

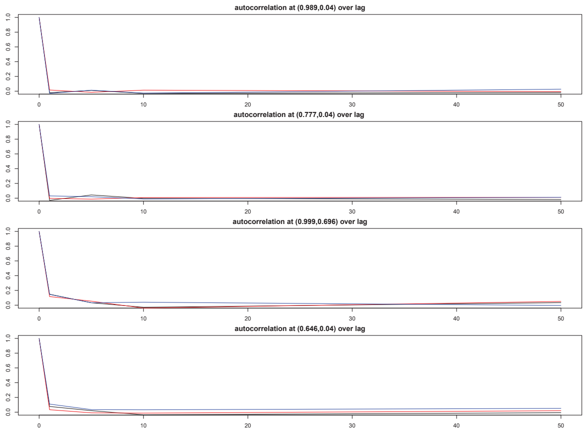


Figure A.42: Autocorrelation plots for 10 Trees

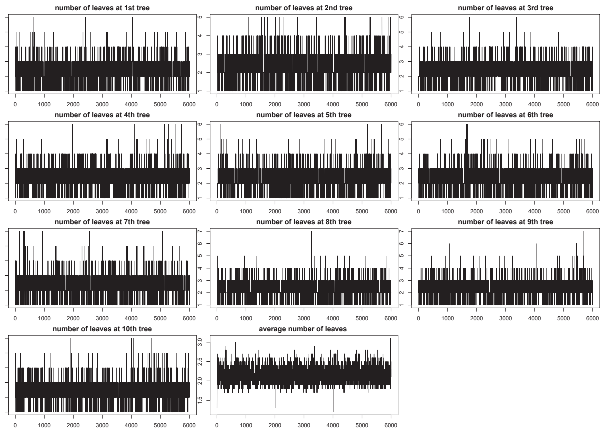


Figure A.43: Average number of leaves at trees

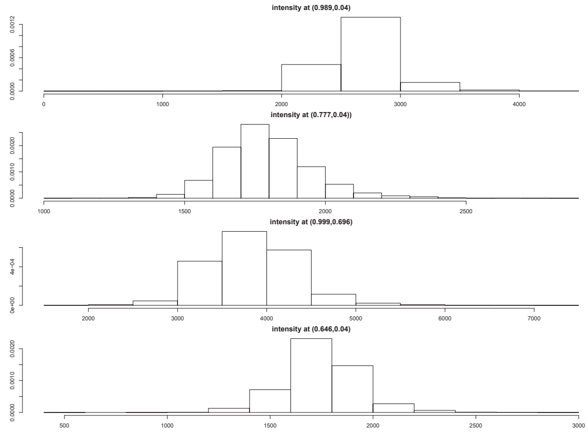


Figure A.44: Density of the estimated intensity for 10 Trees

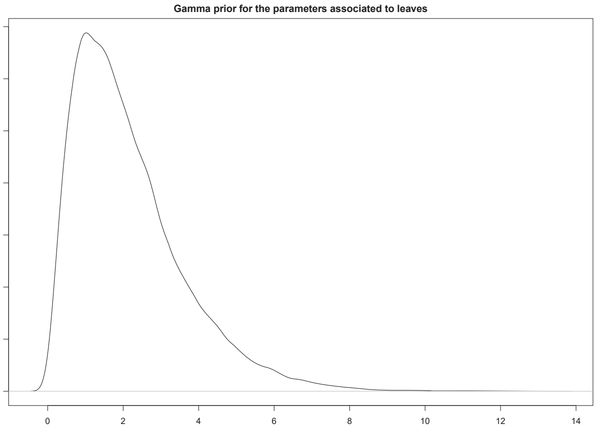


Figure A.45: Prior for 10 Trees

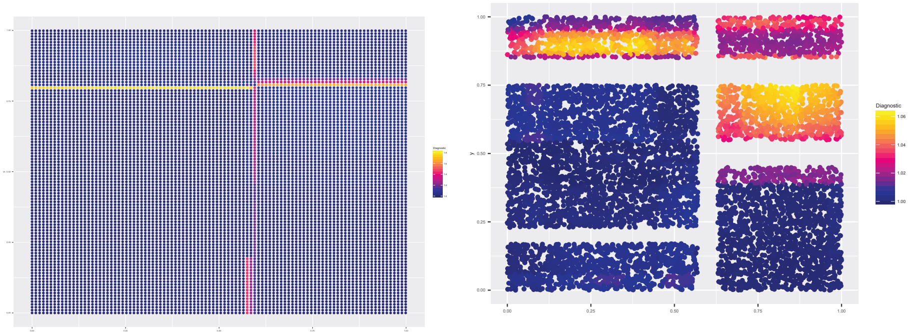
### A.5.4 Two dimensional Poisson process with stepwise intensity function

| Convergence Criterion ($R_{hat}$) | |
|---|---|
| Number of Trees | $R_{hat} < 1.1$ |
| 3 | 90% |
| 4 | 99% |
| 5 | 100% |
| 6 | 100% |

Table A.13: The percentage of testing points having convergence criterion smaller than 1.1.

#### A.5.4.1 4 Trees

We run 3 parallel chains each for 100000 iterations keeping every 50th sample.



(a) Gelman-Rubin Criterion for 4 Trees

(b) Gelman-Rubin Criterion for 4 Trees with removal of points close to jumps

Figure A.46: Gelman-Rubin Criterion

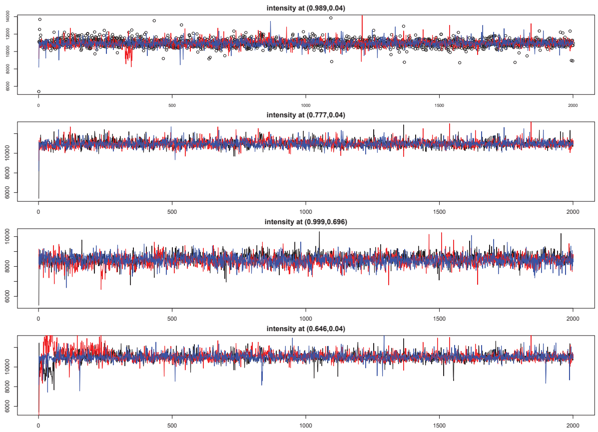


Figure A.47: Trace plots for 4 Trees

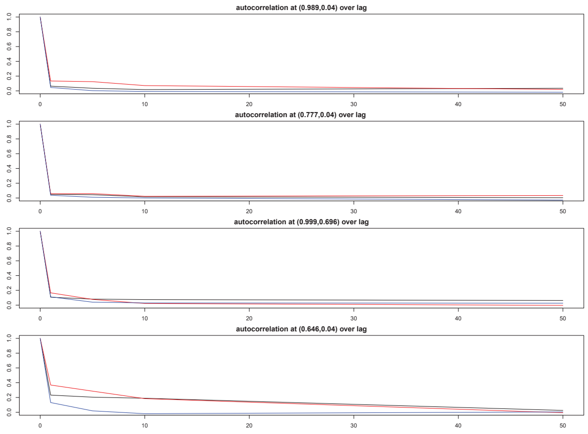


Figure A.48: Autocorrelation plots for 4 Trees

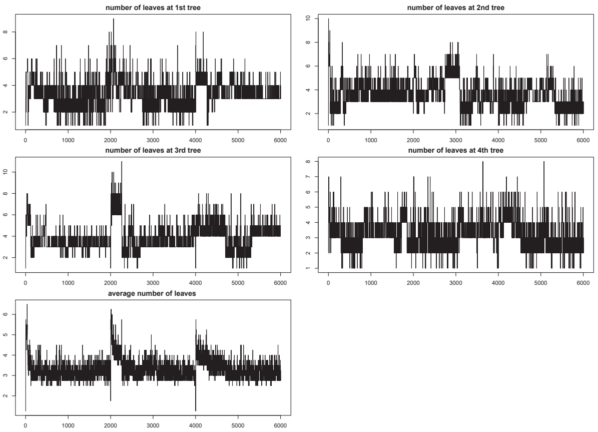


Figure A.49: Average number of leaves at trees

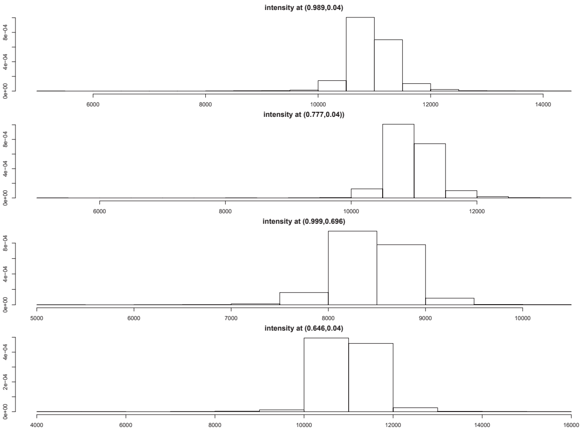


Figure A.50: Density of the estimated intensity for 4 Trees

## A.6 Additional simulation results on real data

To further illustrate the performance of Algorithm 4, we present additional simulation results in real data sets, such as trace and autocorrelation plots, as well as the Gelman-Rubin Criterion.

### A.6.1 Coal Data

#### A.6.1.1 8 Trees

We run 3 parallel chains each for 200000 iterations keeping every 100th sample.

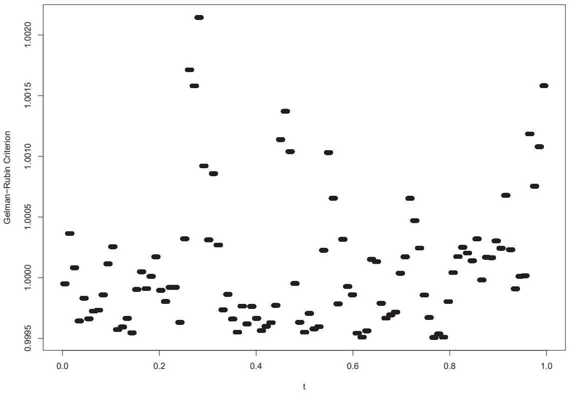


Figure A.51: The Gelman-Rubin Criterion for 8 Trees

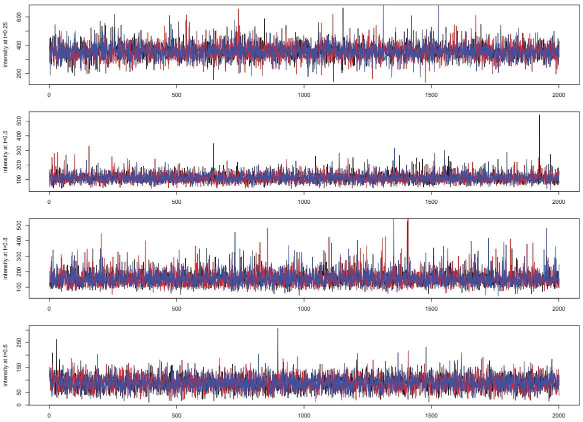


Figure A.52: Trace plots for 8 Trees

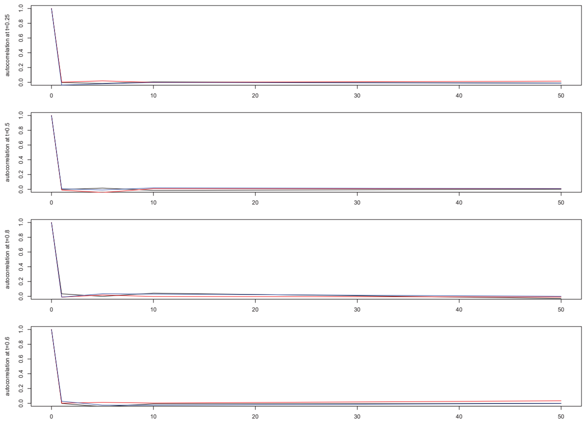


Figure A.53: Autocorrelation plots for 8 Trees

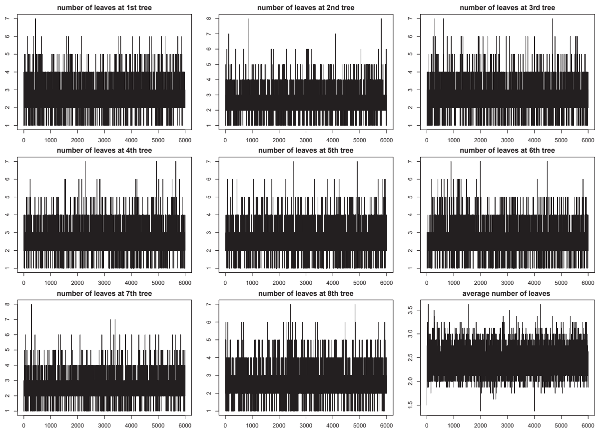


Figure A.54: Average number of leaves at trees

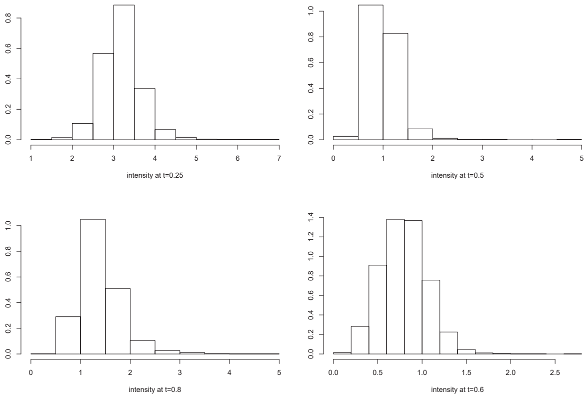


Figure A.55: Density of the estimated intensity for 8 Trees

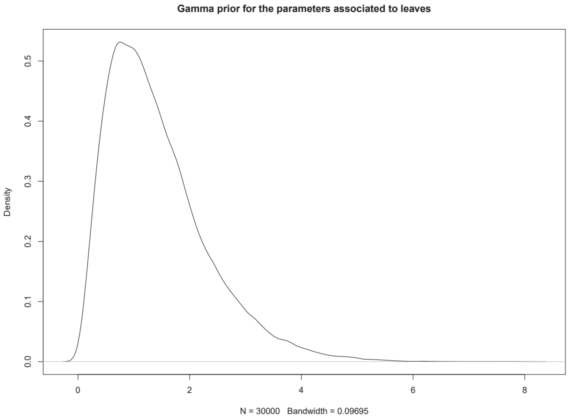


Figure A.56: Prior for 8 Trees

#### A.6.1.2 10 Trees

We run 3 parallel chains each for 200000 iterations keeping every 100th sample.

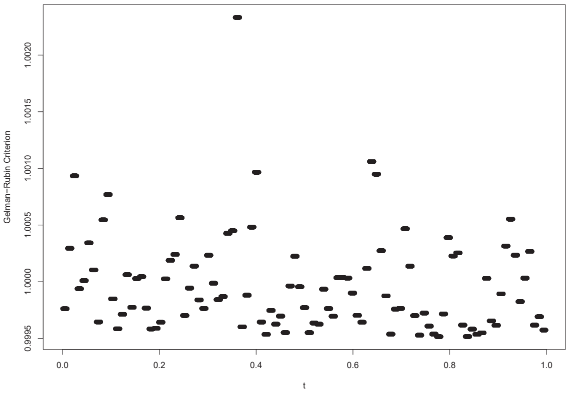


Figure A.57: The Gelman-Rubin Criterion for 10 Trees

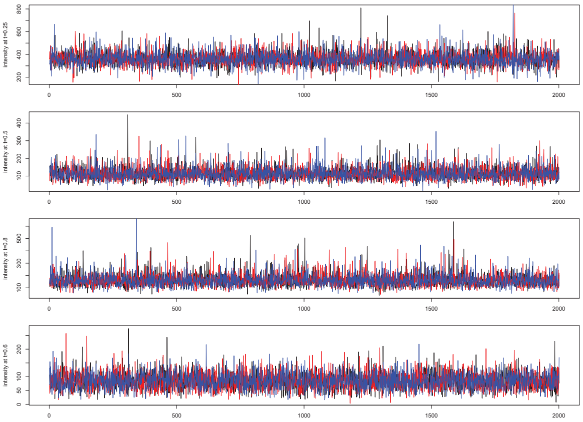


Figure A.58: Trace plots for 10 Trees

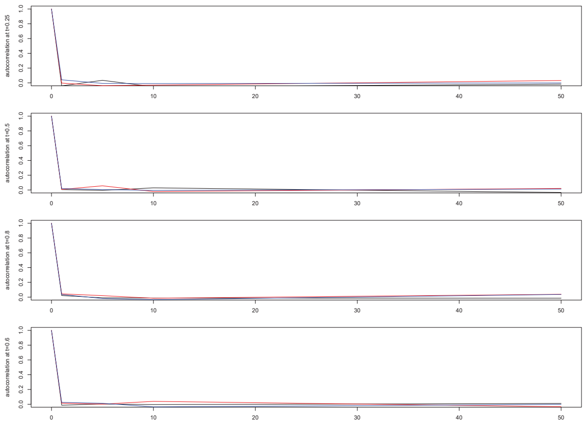


Figure A.59: Autocorrelation plots for 10 Trees

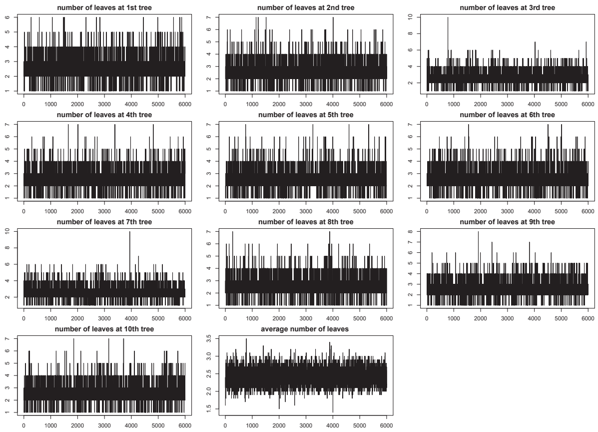


Figure A.60: Average number of leaves at trees

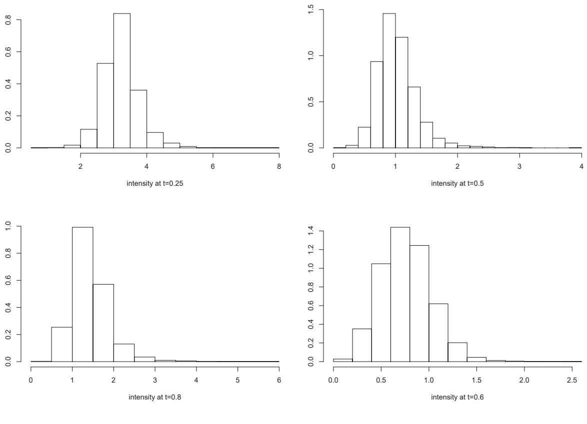


Figure A.61: Density of the estimated intensity for 10 Trees

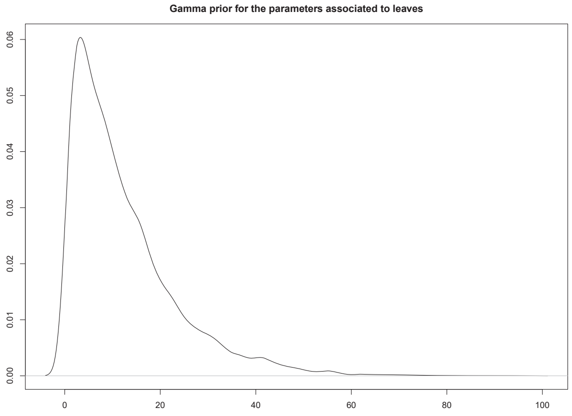


Figure A.62: Prior for 10 Trees

### A.6.2 Earthquakes Data

| Convergence Criterion ($R_{hat}$) | |
|---|---|
| Number of Trees | $R_{hat} < 1.1$ |
| 3 | 100% |
| 4 | 100% |
| 5 | 100% |
| 6 | 100% |
| 8 | 100% |
| 10 | 100% |

Table A.14: The percentage of testing points having convergence criterion smaller than 1.1.

#### A.6.2.1 10 Trees

We run 3 parallel chains each for 100000 iterations keeping every 50th sample.

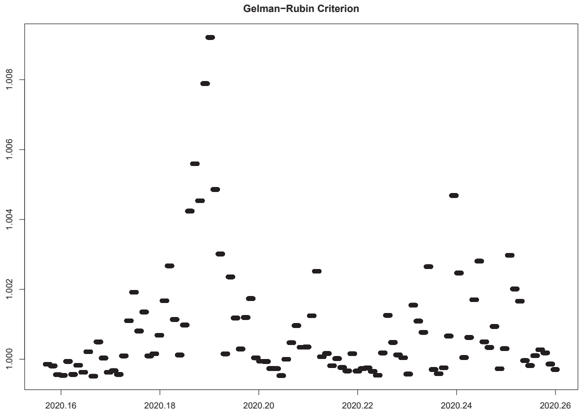


Figure A.63: The Gelman-Rubin Criterion for 10 Trees

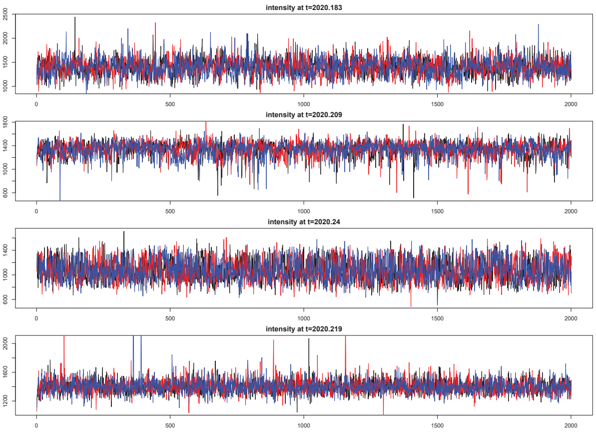


Figure A.64: Trace plots for 10 Trees

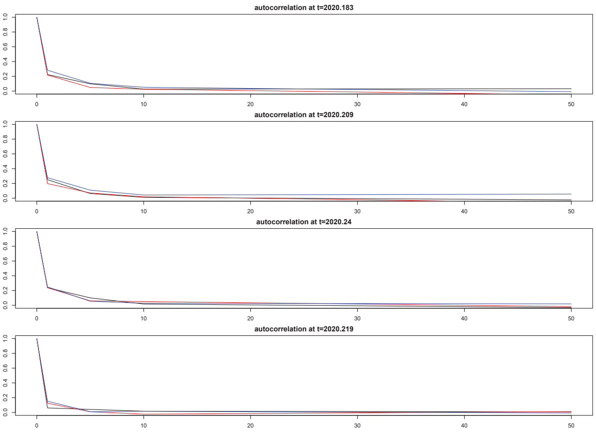


Figure A.65: Autocorrelation plots for 10 Trees

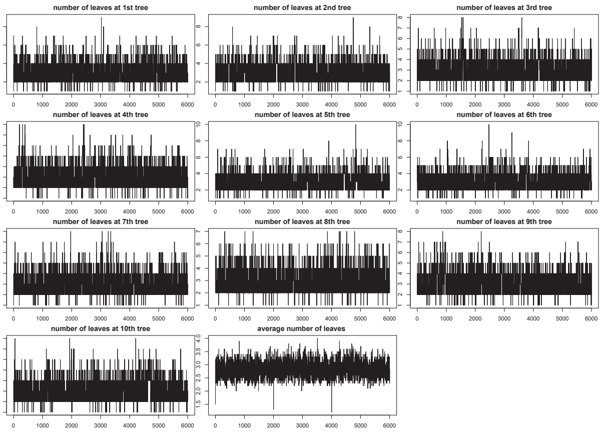


Figure A.66: Average number of leaves at trees

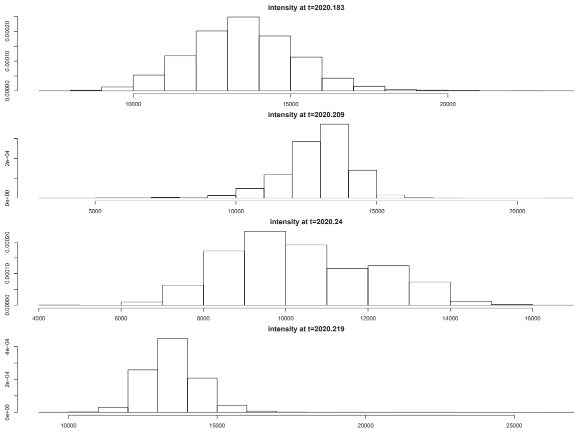


Figure A.67: Density of the estimated intensity for 10 Trees

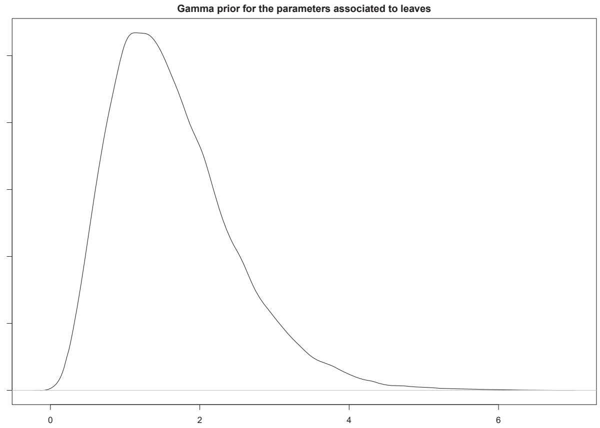


Figure A.68: Prior for 10 Trees

### A.6.3 Mapples

| Convergence Criterion ($R_{hat}$) | | |
|---|---|---|
| Number of Trees | $R_{hat} < 1.1$ | $R_{hat} < 1.2$ |
| 3 | 75% | 96% |
| 4 | 88% | 93% |
| 5 | 95% | 99% |
| 7 | 99.7% | 100% |
| 8 | 95% | 98% |
| 9 | 100% | 100% |
| 10 | 97.4% | 99.6% |
| 12 | 99.6% | 100% |

Table A.15: The percentage of testing points having convergence criterion smaller than 1.1 and 1.2, respectively.

#### A.6.3.1 5Trees

We run 3 parallel chains each for 300000 iterations keeping every 150th sample.

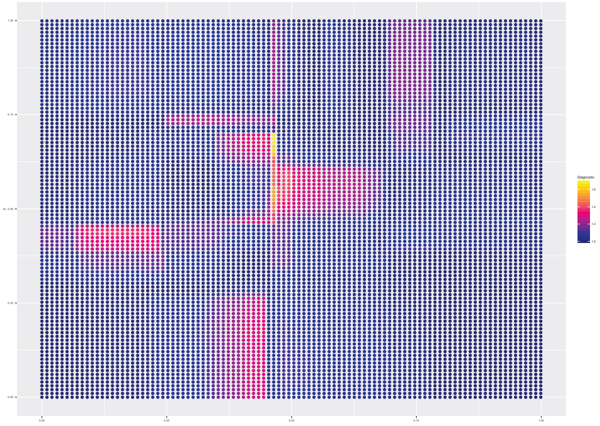


Figure A.69: The Gelman-Rubin Criterion for 5 Trees

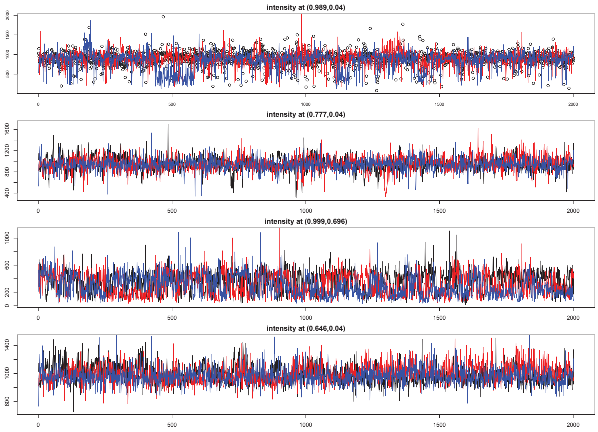


Figure A.70: Trace plots for 5 Trees

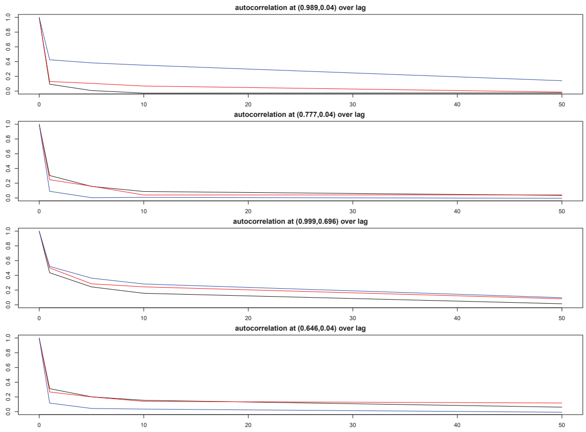


Figure A.71: Autocorrelation plots for 5 Trees

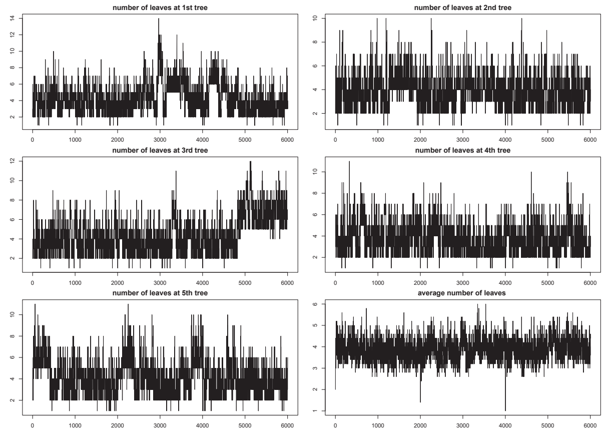


Figure A.72: Average number of leaves at trees

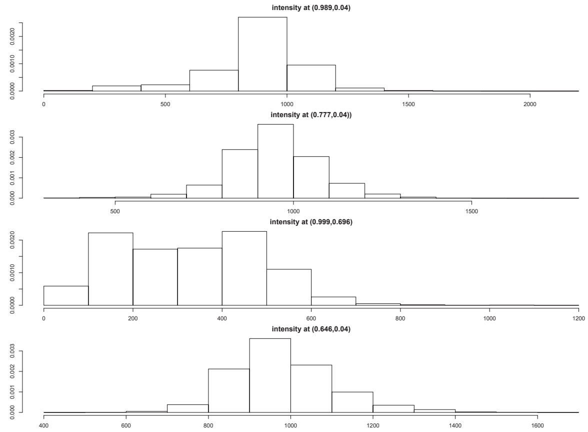


Figure A.73: Density of the estimated intensity for 5 Trees

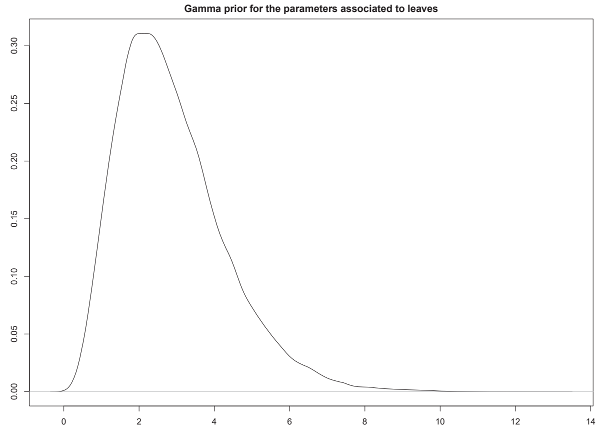


Figure A.74: Prior for 5 Trees

#### A.6.3.2 10 Trees

We run 3 parallel chains each for 300000 iterations keeping every 150th sample.

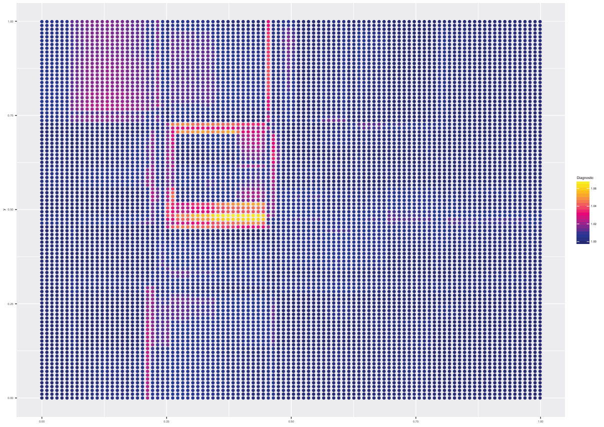


Figure A.75: The Gelman-Rubin Criterion for 10 Trees

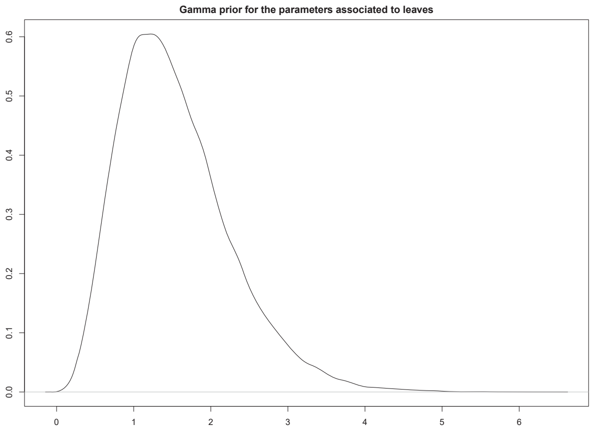


Figure A.76: Prior for 10 Trees

## A.6.4 Redwood

### A.6.4.1 5 Trees

We run 3 parallel chains each for 300000 iterations keeping every 150th sample.

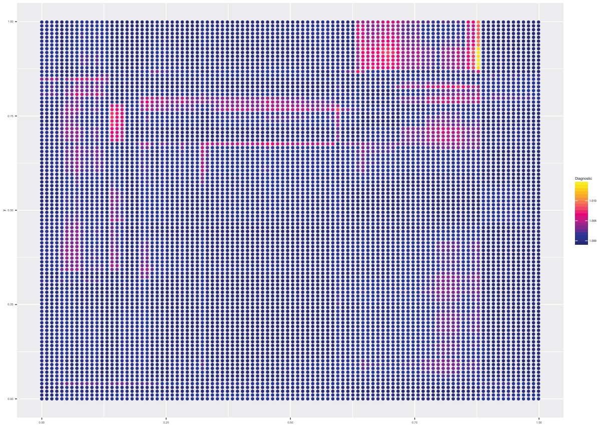


Figure A.77: The Gelman-Rubin Criterion for 5 Trees

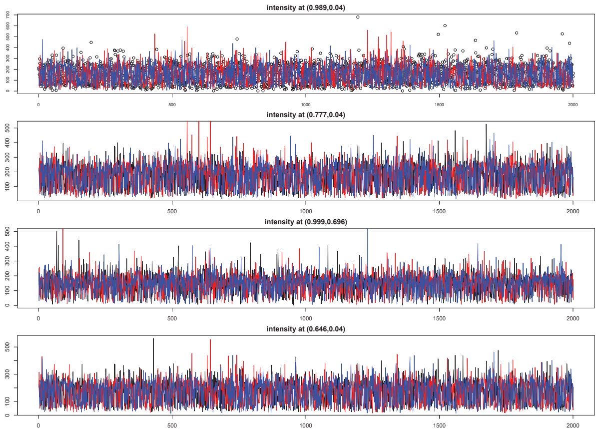


Figure A.78: Trace plots for 5 Trees

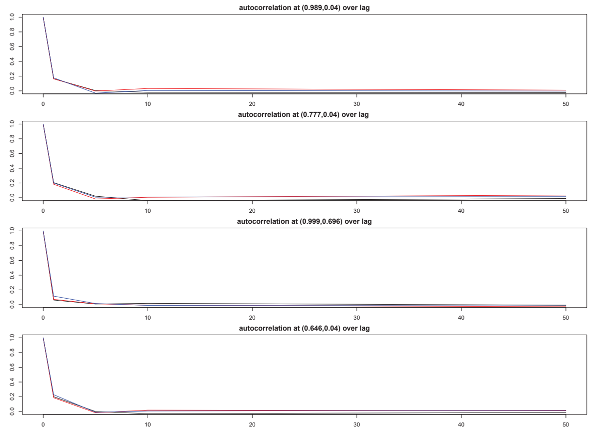


Figure A.79: Autocorrelation plots for 5 Trees

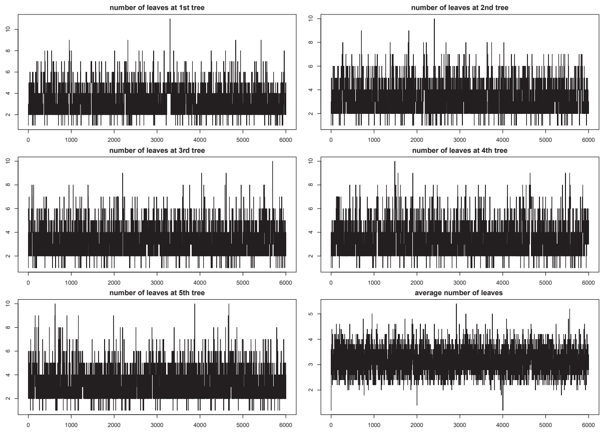


Figure A.80: Average number of leaves at trees

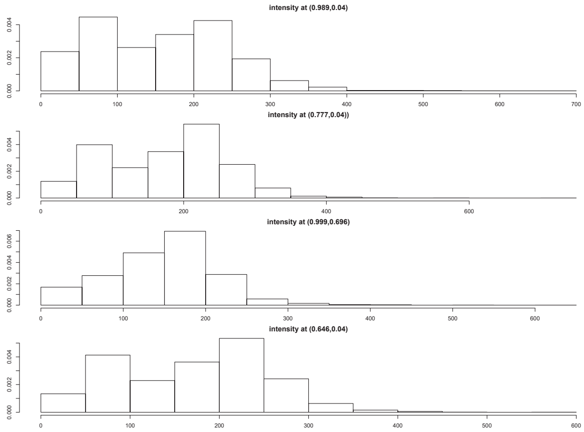


Figure A.81: Density of the estimated intensity for 5 Trees

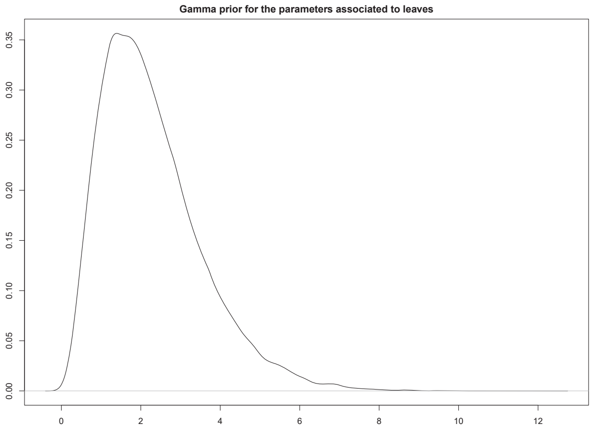


Figure A.82: Prior for 5 Trees

#### A.6.4.2 10 Trees

We run 3 parallel chains each for 300000 iterations keeping every 150th sample.

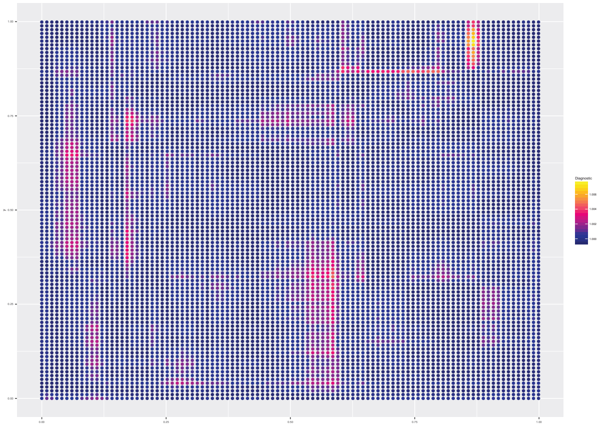


Figure A.83: The Gelman-Rubin Criterion for 10 Trees

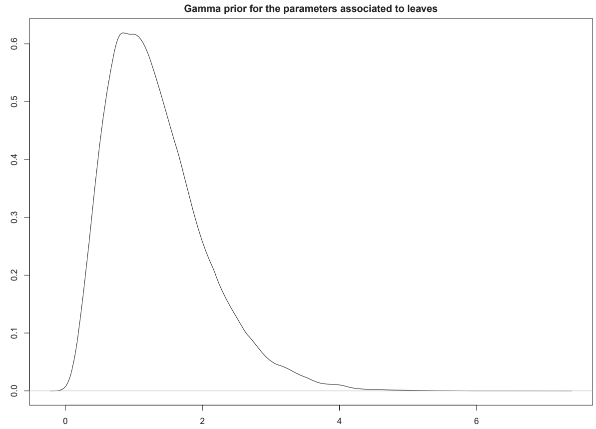


Figure A.84: Prior for 10 Trees

# Appendix B

# Appendix of Chapter 5

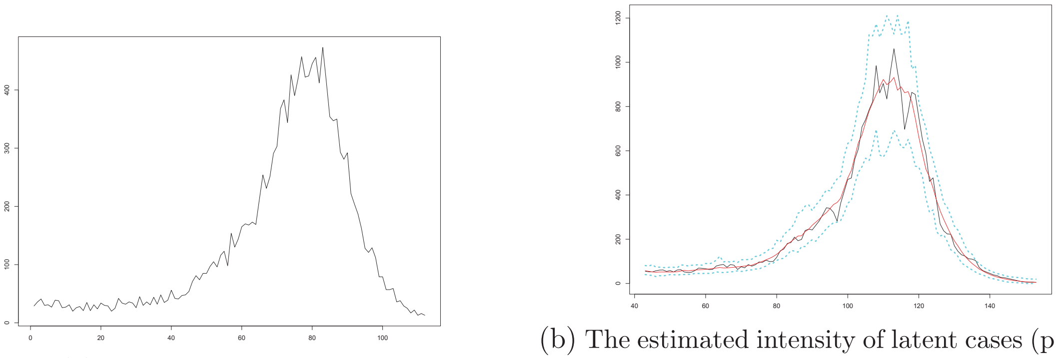
We illustrate here an additional scenario in which the process is triggered by 1023 infectious and the times of their infections, $\mathcal{H}_0$, are uniformly allocated in 21 days ($[0, 21)$) with a day being the time unit. We generate daily latent and observed cases according to the model equations 5.2-5.6 for the next 133 days ($[21, 154)$) given $\mathcal{H}_0$, $v = 0.002$, $d = 53.6$, $\beta = 0.5$ and $R_1 = 1.67$. We are interested in inferring the latent cases in $[42, 154)$ with $\mathcal{H}_0$ being the set of times of latent infections in $[21, 42)$. Using the generated observed cases in $[28, 49)$, we estimate the latent infections in $[21, 42)$ as follows: The latent cases on the day $i$ are equal to the number of observed events on the day $(i + 1)$ times $1/\beta$ and are spread uniformly in $[i, i + 1)$ for $21 \leq i < 42$. We assume $\alpha = 0.5$, $b = 1.5$, $d_{min} = 50$, $d_{max} = 60$, $v_{min} = 0.0001$ and $v_{max} = 0.5$. The ground truth is characterized by $\mathcal{H}_0$ consisting of 1698 seeds, however, while the estimated seeds are 1622. There are 14934 observed cases in $[42, 154)$.

We deal with 112 hidden states $\{X_n\}_{n=1}^{112}$. Each state $X_n$ is associated with the latent cases falling during the day $\mathcal{T}_n$ and the parameter $R_n$ associated with that day. We infer the latent intensity $\lambda^N(t)$ and the weights $\{R_n\}_{n=1}^{112}$ as well as the daily latent cases via the particle sample derived by drawing samples from the smoothing density with lag equal to 21.

Figure B.1 illustrate the estimated latent intensity, the estimated daily hidden cases, the estimated reproduction number, and the 99% CIs and the ESS using 10000 particles. Table B.1 confirms the convergence of posterior estimates concerning the number of particles.

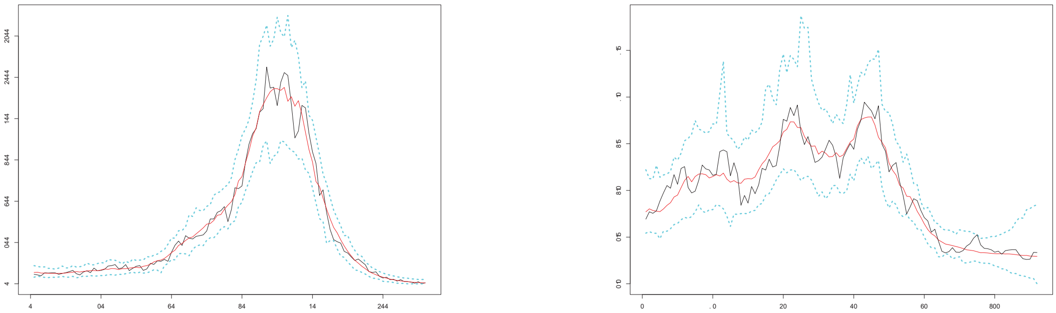
| Convergence of the posterior estimates | | | |
|---|---|---|---|
| MCSE | $N = 10000$ | $N = 15000$ | $N = 20000$ |
| $R$ | 0.001548 | 0.001253 | 0.001091 |
| $Y$ | 0.355208 | 0.284769 | 0.247152 |

Table B.1: MCSEs of posterior means of weights (R) and daily hidden cases (Y).



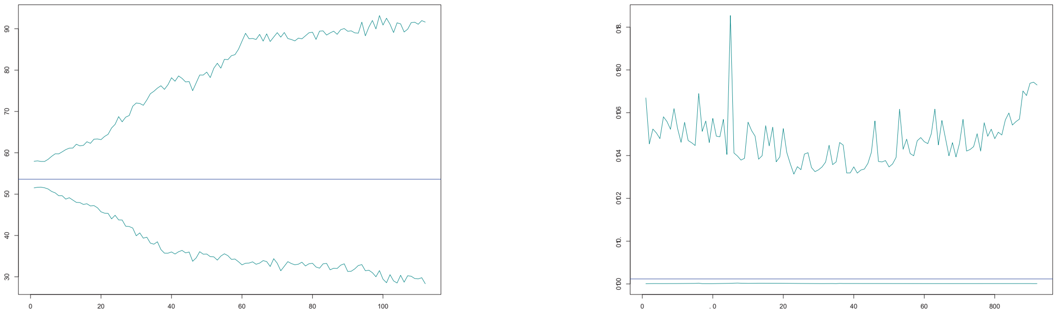
(a) The daily observed cases.

(b) The estimated intensity of latent cases (posterior median (red line) ; 99% CI (cyan line)) and the true intensity (black line).



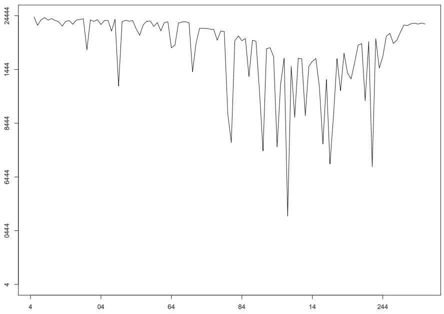
(c) The estimated daily latent cases (posterior median (red line) ;99% CI (cyan line)) and the true daily hidden cases (black line).

(d) The estimated weights $\{R_n\}_{n=1}^{112}$ (posterior median (red line); 99% CI (cyan line)) and the true values (black line).



(e) The 99% CI of $d$.

(f) The 99% CI of $v$.



(g) ESS

Figure B.1: The latent intensity, the reproduction number, the daily latent cases, the 99% CIs of time-constant parameters and the Effective Sample Size (ESS) plotted against time.

# Appendix C

# Appendix of Chapter 6

In this appendix, we present a simulation study for 4, 6 and 9 age groups on synthetic data, and an additional simulation study for the local authorities: Leicester and Ashford to illustrate the performance of the KDPF (Algorithm 16) for inferring the weights $\{\{\gamma_{na}\}_{n=1}^{k}\}_a$, the intensity of latent cases $\lambda^N(t,a)$, the instantaneous reproduction number, $R_a(t)$ and the hidden cases per age group $a$ over various numbers of groups. We introduce an alternative method to initialize the size of the susceptible population per age group at the beginning of the process.

## C.1 Simulation Analysis

### C.1.1 Four age groups

ONS shows that 47.19% of the population is aged under 30 years (hereafter 0-29), 25.42% aged 30 to 49 years (30-49), 18.92% aged 50 to 69 years (50-69), and 8.47% aged 70 years and over (70+) in Leicester [ONS, 2021]. We coarse the age groups of the contact matrix for reopening schools [Jarvis et al., 2020] and get the matrix:

$$\mathrm{m} = \begin{bmatrix} 5.82 & 1.93 & 1.05 & 0.24 \\ 2.75 & 1.60 & 1.00 & 0.36 \\ 1.54 & 1.09 & 1.27 & 0.43 \\ 0.60 & 0.71 & 0.76 & 0.87 \end{bmatrix}.$$

The process is triggered by 4963 infectious; Their infection times are uniformly allocated in 21 days ($[0,21)$).

We generate weekly latent infections and observed cases according to the model equations for weeks -2 – 17 ($[0,161)$) given $v = 0.004$, $d = 25.57$, $\beta = 0.5$, $\gamma_{-2,0-29} = 0.49$, $\gamma_{-2,30-49} = 0.47$, $\gamma_{-2,50-69} = 0.10$ and $\gamma_{-2,70+} = 0.38$. We consider that about 36.46% of the population is susceptible at the beginning of week -2 ; 129073 susceptibles ($0-29: 60893,\ 30-49:\ 32837,\ 50-69:\ 24414,\ 70+: 10929$).

We are interested in inferring the latent infections in weeks $1-16$ with $\mathcal{H}_0$ being the set of latent infection times in weeks -2 – 0. We assume $\alpha = 0$, $b = 0.5$, $d_{min} = 20$, $d_{max} = 30$, $v_{min} = 0.0001$ and $v_{max} = 0.5$. Using the generated observed cases in weeks -1 – 1 as described above, we estimate the latent infections with their associated age groups in weeks -2 – 0 as follows: The latent infections of age group

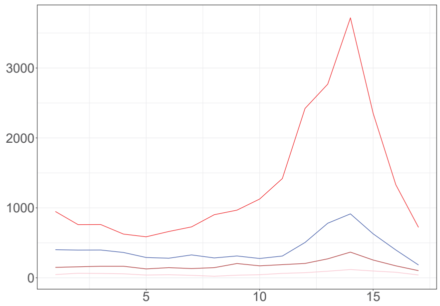
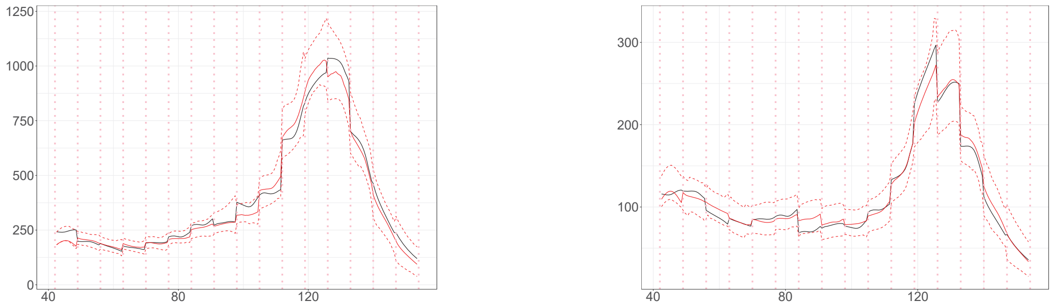


Figure C.1: The weekly observed cases aged 0-29 (red line), 30-49 (blue line), 50-69 (brown line) and 70+ (pink line).

$a_v$ on the week $i$ is equal to the number of reported events in age $a_v$ occurred on the week $(i+1)$ times $1/\beta$, and are spread uniformly in $[(i+2)*7+21, (i+3)*7+21)$ for $-2 \leq i \leq 0$.

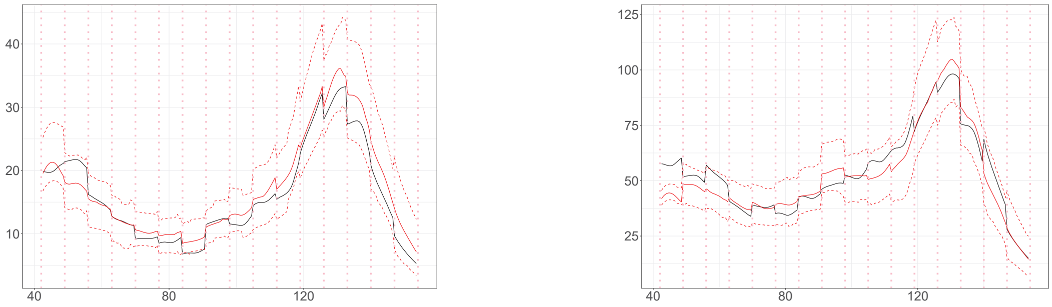
The ground truth is characterized by $\mathcal{H}_0$ consisting of 7987 infectious $(0-29:\ 4539,\ 30-49:\ 2309,\ 50-69:\ 281,\ 70+:\ 768)$, and 121176 susceptibles $(0-29:\ 56354,\ 30-49:\ 30528,\ 50-69:\ 24133,\ 70+:\ 10161)$ at the beginning of week 1. The estimated $\mathcal{H}_0$ consists of 7754 infectious $(0-29:\ 4404,\ 30-49:\ 2254,\ 50-69:\ 392,\ 70+:\ 704)$, and the estimated susceptibles are 121319 $(0-29:\ 56489,\ 30-49:\ 30583,\ 50-69:\ 24022,\ 70+:\ 10225)$, respectively. The observed cases in weeks $1-17$ are 33977 $(0-29:\ 22978,\ 30-49:\ 7046,\ 50-69:\ 1015,\ 70+:\ 3118)$ (Figure C.1).

Figures C.2-C.5 show the estimated intensities of latent infections and weekly latent infections per age group, the estimated weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ and instantaneous reproduction numbers $\{R_a\}_a$, and the ESS for 30000 particles. The majority of the time, the estimated weekly latent infections, intensities of latent infections, and instantaneous reproduction number per age group approach well the ground truth. The 95% CIs do not cover the ground truth of the weekly latent infections aged 50-69 corresponding to the 2nd and 7th week and those of 70+ corresponding to the 1st week, the intensity of latent infections aged 70+ during 1st week and 50-69 during 7th week, the weights $\gamma_{1,0-29}$, $\gamma_{7,50-69}$ and $\gamma_{1,70+}$, and the instantaneous reproduction number of age groups 0-29, 30-49 and 70+ during the 1st week, and 50-69 during the 8th week. Not covering the ground truth during the first weeks might be observed because we have considered $\gamma_{1,70+} \sim \text{Uniform}(0, 0.5)$ while the ground truth is higher than 0.5 as well as the heuristic approach to initialize the set of infectious and susceptibles per age group at the beginning of the process. The simulation analysis also shows that the algorithm achieves more accurate estimations of aggregate weekly latent infections and intensity of latent infections concerning all age groups. Table C.1 confirms the convergence of posterior estimates of weights and weekly hidden cases per age group concerning the number of particles. The 95% Credible Intervals (CIs) of the time-constant parameters include the actual values of the parameters.



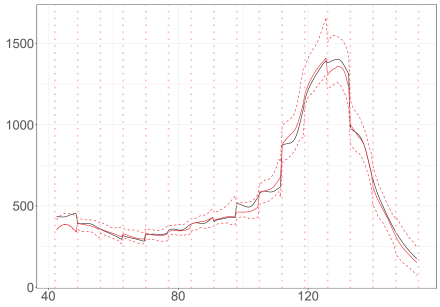
(a) The estimated intensity of latent infections aged 0-29.

(b) The estimated intensity of latent infections aged 30-49.



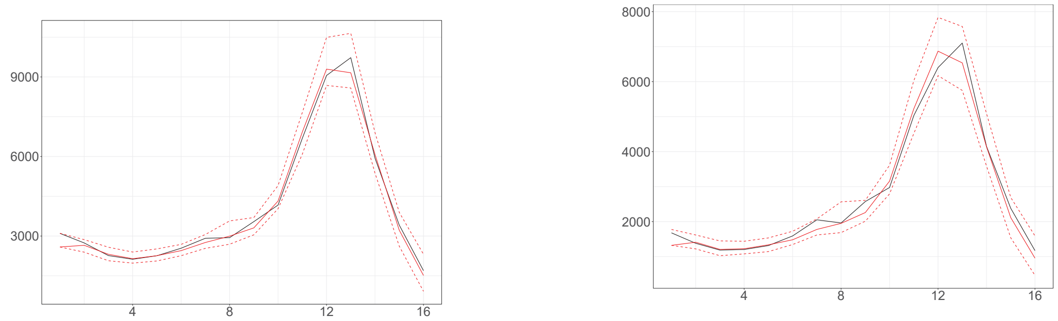
(c) The estimated intensity of latent infections aged 50-69.

(d) The estimated intensity of latent infections aged 70+.



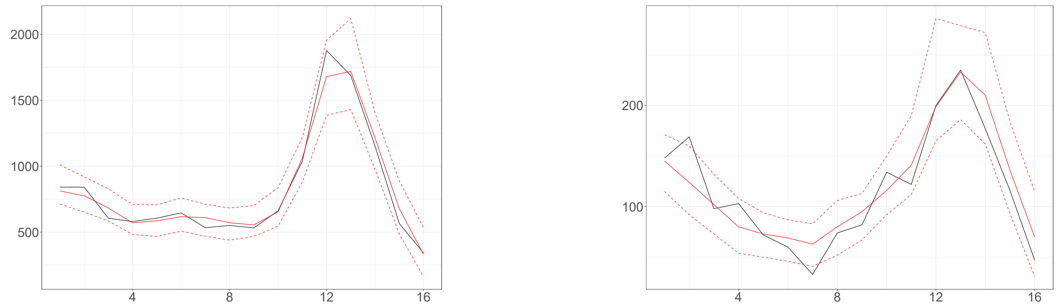
(e) The aggregate estimated intensity of latent infections.

Figure C.2: The ground truth (black line), and the estimated intensities of latent infections for the synthetic scenario using 4 age groups (posterior median (red line) ; 95% CI (red dashed line)) plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



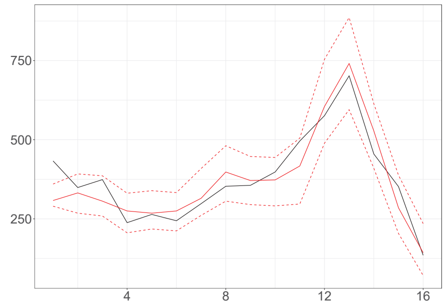
(a) The aggregate estimated weekly latent infections.

(b) The estimated weekly latent infections aged 0-29.



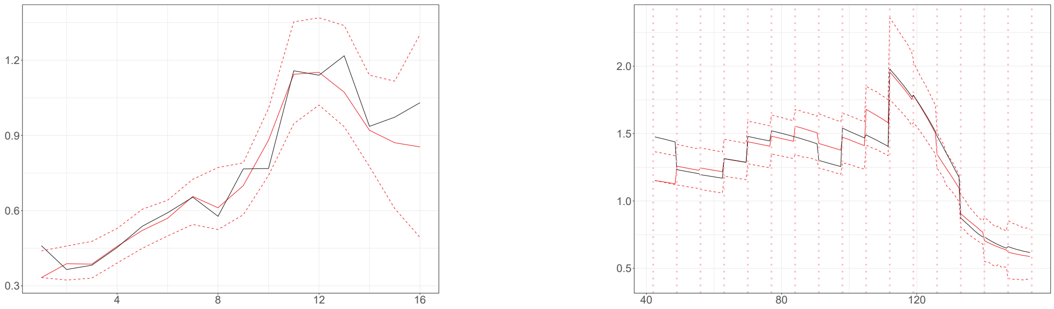
(c) The estimated weekly latent infections aged 30-49.

(d) The estimated weekly latent infections aged 50-69.



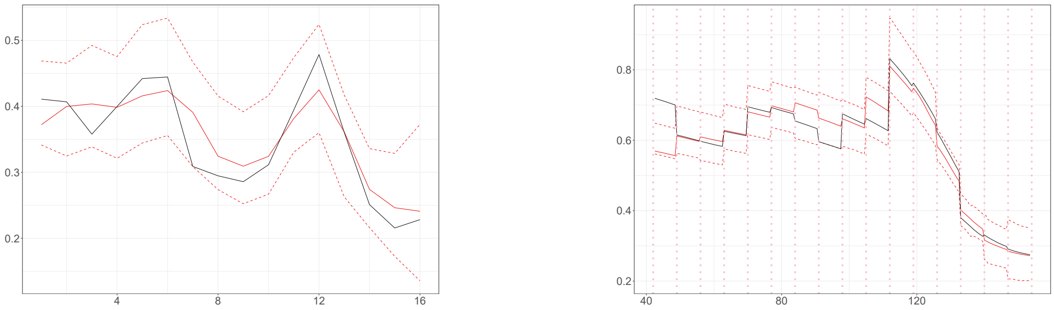
(e) The estimated weekly latent infections aged 70+.

Figure C.3: The ground truth (black line), and the weekly latent infections for the synthetic scenario using 4 age groups (posterior median (red line) ; 95% CI (red dashed line)) plotted against subintervals.



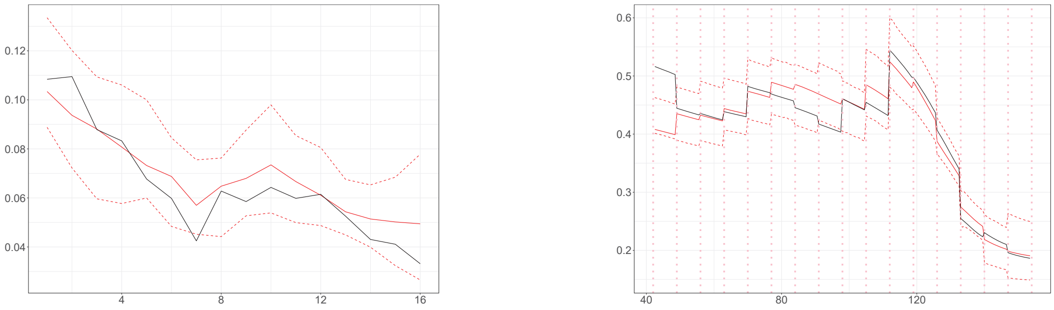
(a) $\{\gamma_{i,0-29}\}_{i=1}^{16}$

(b) $R_{0-29}$



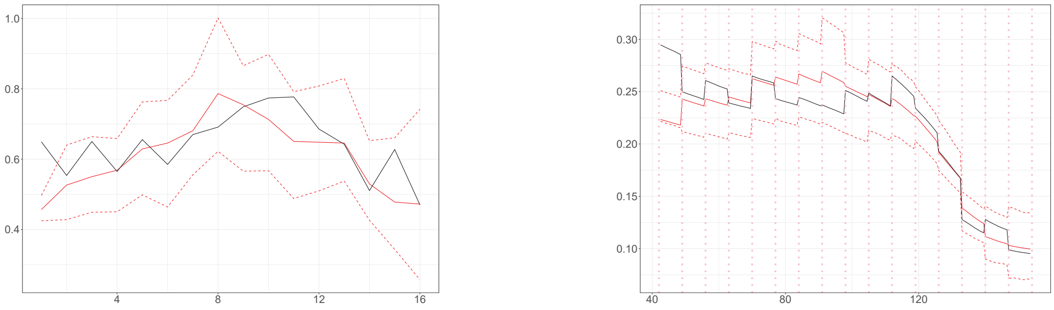
(c) $\{\gamma_{i,30-49}\}_{i=1}^{16}$

(d) $R_{30-49}$



(e) $\{\gamma_{i,50-69}\}_{i=1}^{16}$

(f) $R_{50-69}$



(g) The estimated weights $\{\gamma_{i,70+}\}_{i=1}^{16}$.

(h) $R_{70+}$

Figure C.4: The estimated weights $\{\gamma_{na}\}_a$ and instantaneous reproduction numbers $\{R_a\}_a$ (posterior median (red line); 95% CI (red dashed line)) and the true values (black line) for the synthetic scenario using 4 groups plotted against subintervals and time, respectively. The vertical dotted lines show the beginning of each week in the period we examine.

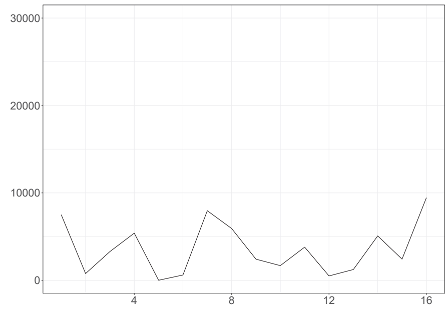


Figure C.5: The ESS for the synthetic scenario using 4 age groups plotted against subintervals.

| Convergence of the posterior estimates | | | | |
|---|---|---|---|---|
| $MCSE$ | $N = 10000$ | $N = 20000$ | $N = 30000$ | $N = 40000$ |
| $\overline{\gamma_1}$ | 0.000741 | 0.000565 | 0.000438 | 0.000428 |
| $\overline{\gamma_2}$ | 0.00043 | 0.0003 | 0.000243 | 0.000208 |
| $\overline{\gamma_3}$ | 0.000101 | 0.000077 | 0.00006 | 0.000053 |
| $\overline{\gamma_4}$ | 0.000811 | 0.00054 | 0.000444 | 0.000381 |
| $\overline{Y_1}$ | 2.4775 | 1.83933 | 1.373194 | 1.415707 |
| $\overline{Y_2}$ | 2.423559 | 2.119945 | 2.007936 | 1.897117 |
| $\overline{Y_3}$ | 0.181335 | 0.127363 | 0.102718 | 0.094132 |
| $\overline{Y_4}$ | 0.462724 | 0.302392 | 0.260452 | 0.213064 |

Table C.1: MCSEs of the average of weights and weekly hidden infections, estimated by the posterior means for the synthetic scenario using 4 age groups.

## C.2 Independent processes $\{\gamma(t,a)\}_a$ on age groups

Particle filters can have an exponential cost in the dimensionality of the hidden state to be stable [Beskos et al., 2017]. Increasing the number of age groups raises the dimensionality of the hidden state and, by extension, the computational resources required to keep the uniform-in-time convergence results dependent on recent states [Kantas et al., 2015]. A simple approach to mitigate the dimensionality problem would be to consider the process $\gamma(t,a) = c_a q(t)$, where $c_a$ is a given constant associated with age group $a$ and $q(t)$ a random process independent of age groups.

Here, we present a simulation analysis considering six age groups in Leicester [ONS, 2021], and nine age groups adopting the demographic features in Ashford [ONS, 2021] with independent processes $\{\gamma(t,a)\}_a$ on age groups, that is, $\gamma(t,a) = \gamma(t)$, $\forall$ age group $a$.

### C.2.1 Six age groups

We coarse the age groups of the contact matrix for reopening schools [Jarvis et al., 2020]. The process is triggered by 1630 infectious; Their infection times are uni-

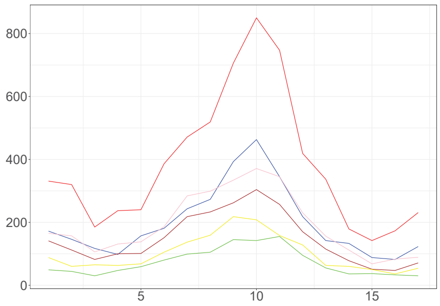


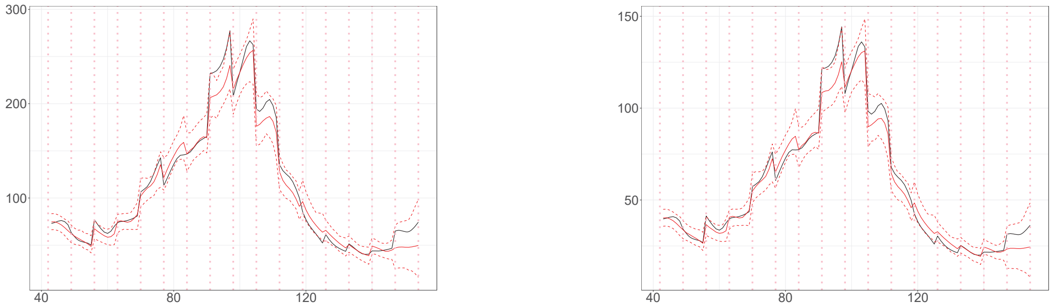
Figure C.6: The weekly observed cases aged 0-29 (red line), 30-39 (blue line), 40-49 (pink line), 50-59 (brown line), 60-69 (yellow line) and 70+ (green line) considering weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ independent on age groups $a$ plotted against time.

formly allocated in 21 days ($[0, 21)$). We generate weekly latent infections and observed cases according to the model equations for weeks -2 $-$ 17 ($[0, 161)$) given $v = 0.003$, $d = 15.1$, $\beta = 0.5$ and $\gamma_{-2} = 0.4$. We consider that about 71% of the population is susceptible at the beginning of week -2 ; 252406 susceptibles ($0-29:\ 119117,\ 30-39:\ 35549,\ 40-49:\ 28611,\ 50-59: 26462,\ 60-69: 21293,\ 70+:\ 21374$).

We are interested in inferring the latent infections in weeks $1-16$ with $\mathcal{H}_0$ being the set of latent infection times in weeks -2 $-$ 0 derived. We assume $\alpha = 0$, $b = 0.5$, $d_{min} = 10$, $d_{max} = 20$, $v_{min} = 0.0001$ and $v_{max} = 0.5$.

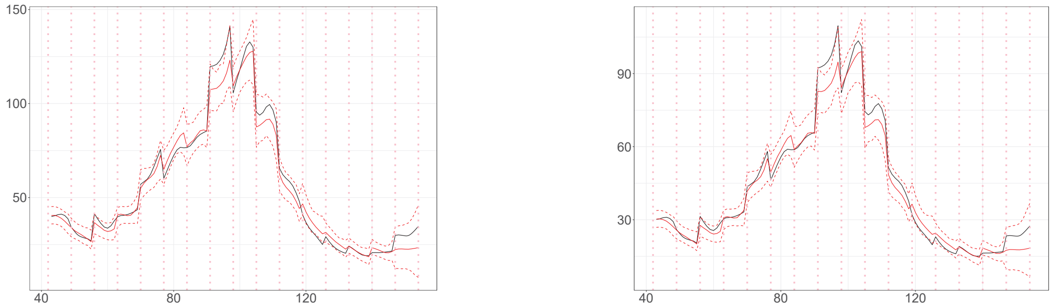
The ground truth is characterized by $\mathcal{H}_0$ consisting of 5513 infections ($0-29:\ 1870,\ 30-39:\ 992,\ 40-49:\ 1070,\ 50-59:\ 765,\ 60-69:\ 481,\ 70+:\ 335$), and, 246893 susceptibles ($0-29:\ 117247,\ 30-39:\ 34557,\ 40-49:\ 27541,\ 50-59: 25697,\ 60-69:\ 20310,\ 70+:\ 21039$) at the beginning of week 1. The estimated $\mathcal{H}_0$ consists of 4880 infections ($0-29:\ 1739,\ 30-39:\ 816,\ 40-49:\ 890,\ 50-59:\ 730,\ 60-69:\ 414,\ 70+:\ 294$), and, the estimated susceptibles are 247526 ($0-29:\ 117381,\ 30-39:\ 34773,\ 40-49:\ 27721,\ 50-59:\ 25732,\ 60-69: 20879,\ 70+:\ 21080$), respectively. The observed cases in weeks $1-17$ are 18555 ($< 30:\ 6473,\ 30-39:\ 3374,\ 40-49:\ 3254,\ 50-59:\ 2493,\ 60-69:\ 1720,\ 70+: 1241$) (Figure C.6).

Figures C.7-C.11 show the estimated intensities of latent infections, the estimated weekly latent infections, the estimated weights $\{\gamma_n\}_{n=1}^{16}$ and the estimated instantaneous reproduction numbers $\{R_a\}_{a\in\mathcal{A}}$ for each age group and the ESS for 20000 particles. We observe that the 95% credible intervals do cover the ground truth most of the time. Table C.2 confirms the convergence of posterior estimates of weights and weekly latent infections per age group concerning the number of particles. The 95% Credible Intervals (CIs) of the time-constant parameters include the actual values of the parameters.



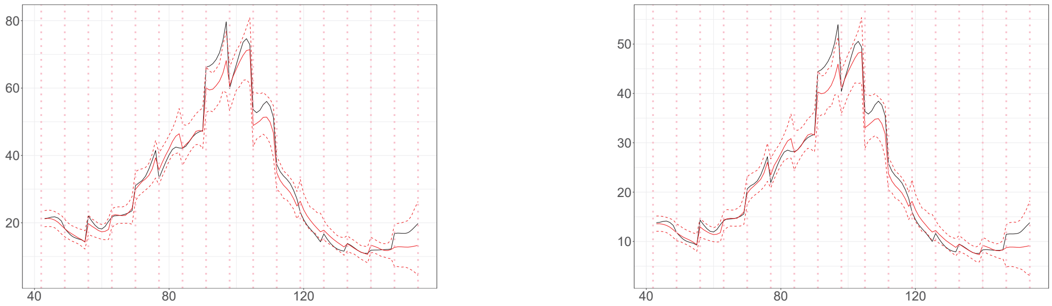
(a) The estimated intensity of latent infections aged 0-29.

(b) The estimated intensity of latent infections aged 30-39.



(c) The estimated intensity of latent infections aged 40-49.

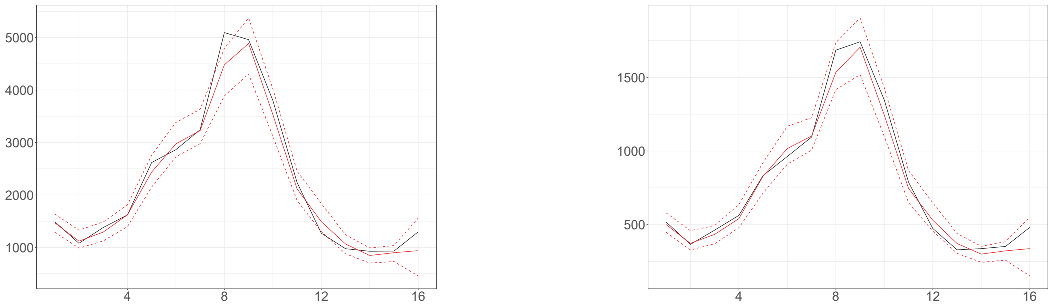
(d) The estimated intensity of latent infections aged 50-59.



(e) The estimated intensity of latent infections aged 60-69.

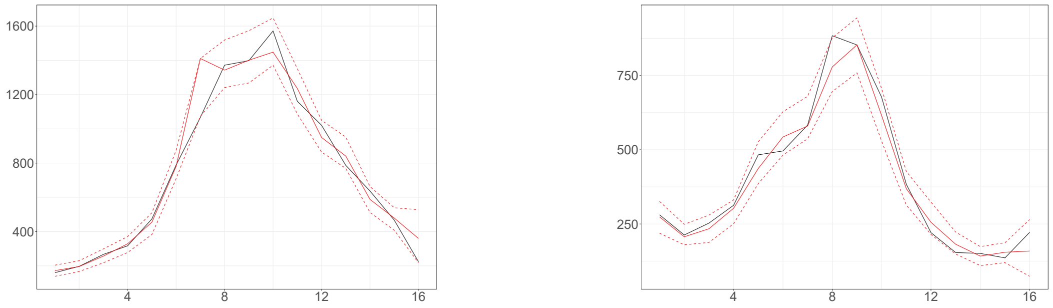
(f) The estimated intensity of latent infections aged 70+.

Figure C.7: The ground truth (black line), and the estimated intensities of latent infections (posterior median (red line) ; 95% CI (red dashed line)) for the synthetic scenario using 6 age groups and weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ independent on age groups $a$ plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



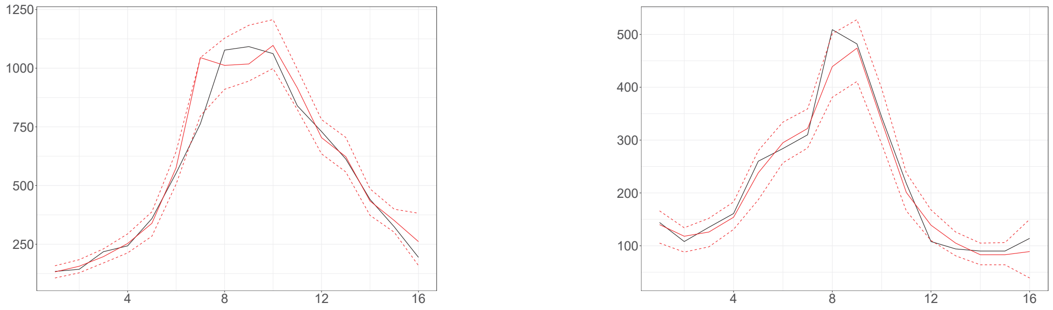
(a) The aggregate estimated weekly latent infections

(b) Aged 0-29



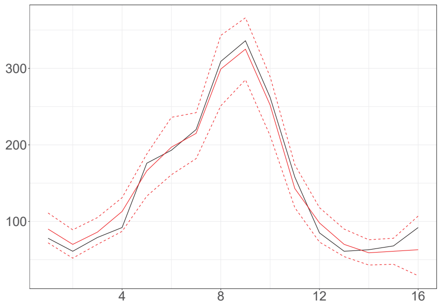
(c) Aged 30-39

(d) Aged 40-49



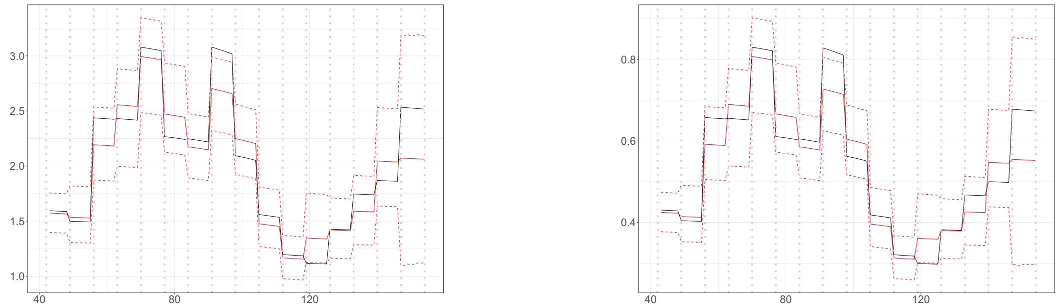
(e) Aged 50-59

(f) Aged 60-69

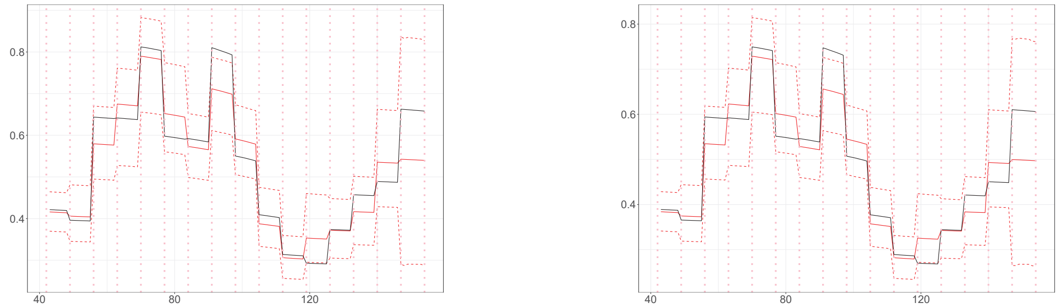


(g) Aged 70+

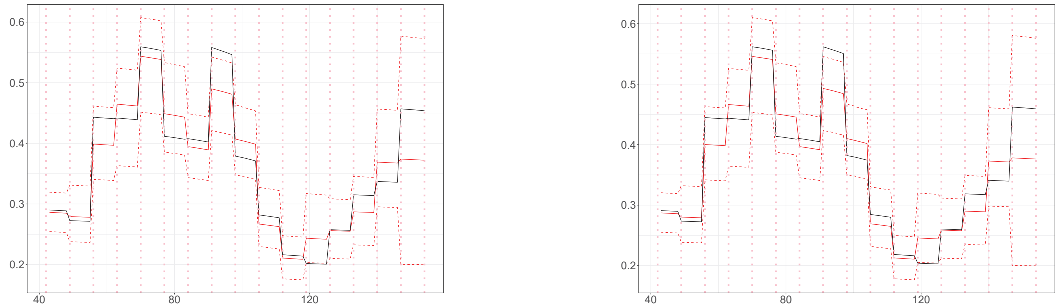
Figure C.8: The ground truth (black line), and the weekly latent infections (posterior median (red line) ; 95% CI (red dashed line)) for the synthetic scenario using 6 age groups and weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ independent on age groups $a$ plotted against subintervals.



(a) Aged 0-29. (b) Aged 30-39.



(c) Aged 40-49. (d) Aged 50-59.



(e) Aged 60-69. (f) Aged 70+.

Figure C.9: The ground truth (black line), and the instantaneous reproduction numbers (posterior median (red line) ; 95% CI (red dashed line)) for the synthetic scenario using 6 age groups and weights independent on age groups $a$ plotted against subintervals. The vertical dotted lines show the beginning of each week in the period we examine.

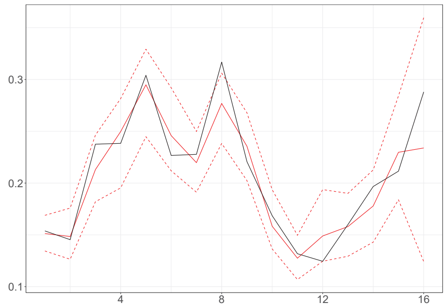


Figure C.10: The estimated weights $\{\gamma_n\}_{n=1}^{16}$ (posterior median (red line); 95% CI (red dashed line)) and the true values (black line) for the synthetic scenario using 6 groups plotted against subintervals.

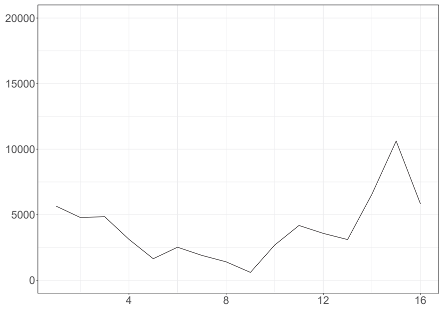


Figure C.11: The ESS for the synthetic scenario using 6 age groups and weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ independent on age groups $a$ plotted against subintervals.

| Convergence of the posterior estimates | | | |
|---|---|---|---|
| $MCSE$ | $N = 10000$ | $N = 20000$ | $N = 30000$ |
| $\overline{\gamma}$ | 0.0002 | 0.0001 | 0.0001 |
| $\overline{Y_1}$ | 0.58 | 0.42 | 0.37 |
| $\overline{Y_2}$ | 0.32 | 0.24 | 0.21 |
| $\overline{Y_3}$ | 0.31 | 0.24 | 0.20 |
| $\overline{Y_4}$ | 0.26 | 0.18 | 0.16 |
| $\overline{Y_5}$ | 0.19 | 0.14 | 0.12 |
| $\overline{Y_6}$ | 0.15 | 0.1 | 0.09 |

Table C.2: MCSEs of the average of weights and weekly hidden infections, estimated by the posterior means for the synthetic scenario using 6 age groups and weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ independent on age groups $a$.

### C.2.2 Nine age groups

ONS shows that 6.69% of the population is aged under 5 years (0-4), 9.69% aged 5 to 11 years (5-11), 7.36% aged 12 to 17 years (12-17), 23.45% aged 18 to 29 years (18-29), 14.09% aged 30 to 39 years (30-39), 11.33% aged 40 to 49 years (40-49), 10.48% aged 50 to 59 years (50-59), 8.44% aged 60 to 69 years (60-69) and 8.47% over 70 years (70+) in Ashford [ONS, 2021]. We consider the matrix:

$$\mathrm{m} = \begin{bmatrix} 1.56 & 2.21 & 1.86 & 0.97 & 1.77 & 1.05 & 0.51 & 0.28 & 0.14 \\ 1.51 & 4.39 & 2.97 & 0.63 & 1.19 & 1.18 & 0.58 & 0.24 & 0.15 \\ 1.61 & 3.7 & 3.95 & 1.33 & 0.97 & 1.53 & 1.22 & 0.34 & 0.21 \\ 0.38 & 0.36 & 0.60 & 1.21 & 0.74 & 0.58 & 0.60 & 0.40 & 0.31 \\ 0.75 & 0.73 & 0.47 & 0.80 & 0.92 & 0.68 & 0.53 & 0.40 & 0.30 \\ 0.48 & 0.78 & 0.81 & 0.68 & 0.74 & 0.85 & 0.70 & 0.37 & 0.44 \\ 0.22 & 0.36 & 0.61 & 0.71 & 0.54 & 0.66 & 0.78 & 0.47 & 0.35 \\ 0.15 & 0.19 & 0.21 & 0.55 & 0.51 & 0.43 & 0.59 & 0.72 & 0.53 \\ 0.06 & 0.09 & 0.10 & 0.34 & 0.30 & 0.40 & 0.34 & 0.41 & 0.87 \end{bmatrix}.$$

The process is triggered by 982 infectious; Their infection times are uniformly allocated in 21 days ($[0, 21)$). We generate weekly latent infections and observed

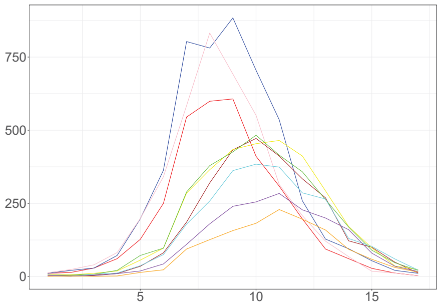


Figure C.12: The weekly observed cases aged 0-4 (red line), 5-11 (blue line), 12-17 (pink line), 18-29 (brown line), 30-39 (yellow line), 40-49 (green line), 50-59 (cyan line), 60-69 (purple line) and 70+ (orange line) considering weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ independent on age groups $a$ plotted against time.

cases according to the model equations for weeks -2 $-$ 17 ($[0, 161)$) given $\upsilon = 0.007$ and $d = 15.3$. We consider that about 99% of the population is susceptible at the beginning of week -2 ; 130036 susceptibles ($0-4:$ 7716, $5-11:$ 12276, $12-17:$ 9892, $18-29:$ 16026, $30-39:$ 15514, $40-49:$ 16645, $50-59:$ 19075, $60-69:$ 13881, $70+:$ 19011).

We are interested in inferring the latent infections in weeks $1-16$ with $\mathcal{H}_0$ being the set of latent infection times in weeks -2 $-$ 0 derived. We assume $\alpha = 0$, $b = 0.5$, $d_{min} = 10$, $d_{max} = 20$, $\upsilon_{min} = 0.0001$ and $\upsilon_{max} = 0.5$.

The ground truth is characterized by $\mathcal{H}_0$ consisting of 567 infections ($0-4:$ 95, $5-11:$ 127, $12-17:$ 143, $18-29:$ 39, $30-39:$ 43, $40-49:$ 37, $50-59:$ 33, $60-69:$ 30, $70+:$ 20), and, 129469 susceptibles ($0-4:$ 7621, $5-11:$ 12149, $12-17:$ 9749, $18-29:$ 15987, $30-39:$ 15471, $40-49:$ 16608, $50-59:$ 19042, $60-69:$ 13851, $70+:$ 18991) at the beginning of week 1. The estimated $\mathcal{H}_0$ consists of 588 infections ($0-4:$ 60, $5-11:$ 106, $12-17:$ 124, $18-29:$ 52, $30-39:$ 62, $40-49:$ 50, $50-59:$ 52, $60-69:$ 46, $70+:$ 36), and, the estimated susceptibles are 129448 ($0-4:$ 7656, $5-11:$ 12170, $12-17:$ 9768, $18-29:$ 15974, $30-39:$ 15452, $40-49:$ 16595, $50-59:$ 19023, $60-69:$ 13835, $70+:$ 18975), respectively. The observed cases in weeks $1-17$ are 27470 ($0-4:$ 3358, $5-11:$ 4970, $12-17:$ 4102, $18-29:$ 2843, $30-39:$ 3232, $40-49:$ 3148, $50-59:$ 2562, $60-69:$ 1871, $70+:$ 1384) (Figure C.12).

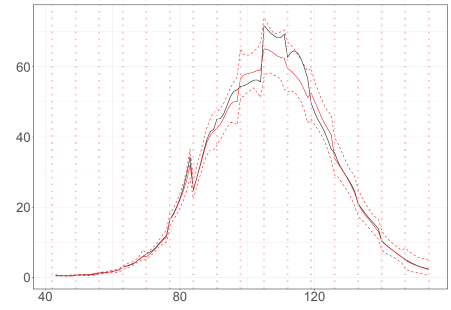
Figures C.13-C.17 show the estimated intensities of latent infections, the estimated weekly latent infections, the estimated weights $\{\gamma_n\}_{n=1}^{16}$ and the estimated instantaneous reproduction numbers $\{R_a\}_{a\in\mathcal{A}}$ for each age group and the ESS for 30000 particles. We observe that the 95% credible intervals do cover the ground truth most of the time. Table C.3 confirms the convergence of posterior estimates of weights and weekly latent infections per age group concerning the number of particles. The 95% Credible Intervals (CIs) of the time-constant parameters include the actual values of the parameters.

(a) 0-4. (b) 5-11.

(c) 12-17. (d) 18-29.

(e) 30-39. (f) 40-49.

(g) 50-59. (h) 60-69.



(i) 70+.

Figure C.13: The ground truth (black line), and the estimated intensities of latent infections (posterior median (red line) ; 95% CI (red dashed line)) for the synthetic scenario using 9 age groups and weights independent on age groups $a$ plotted against time.

(a) 0-4.

(b) 5-11.

(c) 12-17.

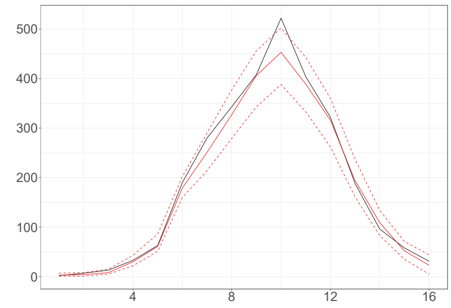
(d) 18-29.

(e) 30-39.

(f) 40-49.

(g) 50-59.

(h) 60-69.



(i) 70+.

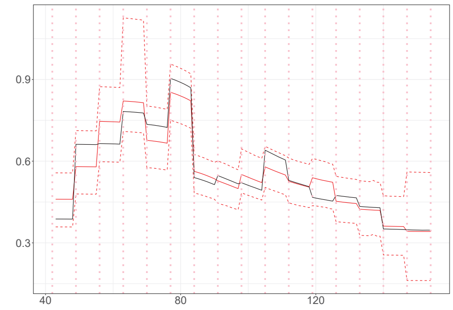
Figure C.14: The ground truth (black line), and the weekly latent infections (posterior median (red line) ; 95% CI (red dashed line)) for the synthetic scenario using 6 age groups and weights independent on age groups $a$ plotted against subintervals.

(a) 0-4. (b) 5-11.

(c) 12-17. (d) 18-29.

(e) 30-39. (f) 40-49.

(g) 50-59. (h) 60-69.



(i) 70+.

Figure C.15: The ground truth (black line), and the instantaneous reproduction numbers (posterior median (red line) ; 95% CI (red dashed line)) for the synthetic scenario using 9 age groups and weights independent on age groups $a$ plotted against subintervals.

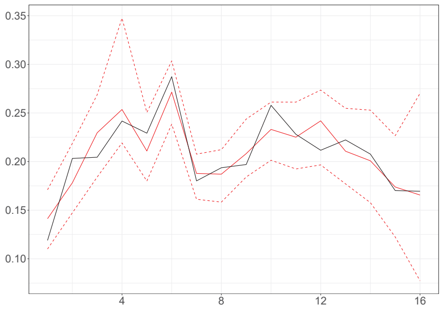


Figure C.16: The estimated weights $\{\gamma_n\}_{n=1}^{16}$ (posterior median (red line); 95% CI (red dashed line)) and the true values (black line) for the synthetic scenario using 9 groups plotted against subintervals.

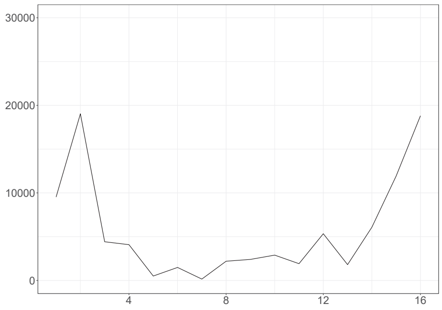


Figure C.17: The ESS for the synthetic scenario using 9 age groups and weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ independent on age groups $a$ plotted against subintervals.

| Convergence of the posterior estimates | | | |
|---|---|---|---|
| $MCSE$ | $N = 10000$ | $N = 20000$ | $N = 30000$ |
| $\overline{\gamma}$ | 0.0002 | 0.0001 | 0.0001 |
| $\overline{Y_1}$ | 0.26 | 0.2 | 0.17 |
| $\overline{Y_2}$ | 0.38 | 0.31 | 0.25 |
| $\overline{Y_3}$ | 0.35 | 0.25 | 0.22 |
| $\overline{Y_4}$ | 0.24 | 0.19 | 0.16 |
| $\overline{Y_5}$ | 0.26 | 0.18 | 0.16 |
| $\overline{Y_6}$ | 0.27 | 0.19 | 0.18 |
| $\overline{Y_7}$ | 0.24 | 0.17 | 0.15 |
| $\overline{Y_8}$ | 0.19 | 0.14 | 0.12 |
| $\overline{Y_9}$ | 0.15 | 0.1 | 0.09 |

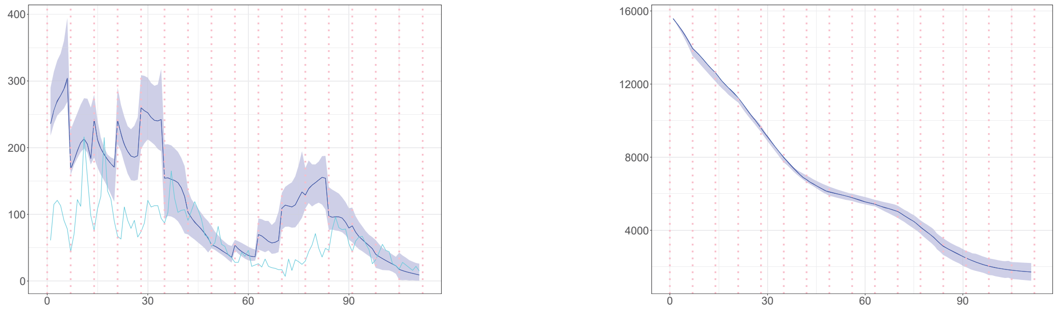
Table C.3: MCSEs of the average of weights and weekly hidden infections, estimated by the posterior means for the synthetic scenario using 9 age groups and weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ independent on age groups $a$.

## C.3 Real Data

### C.3.1 Ashford

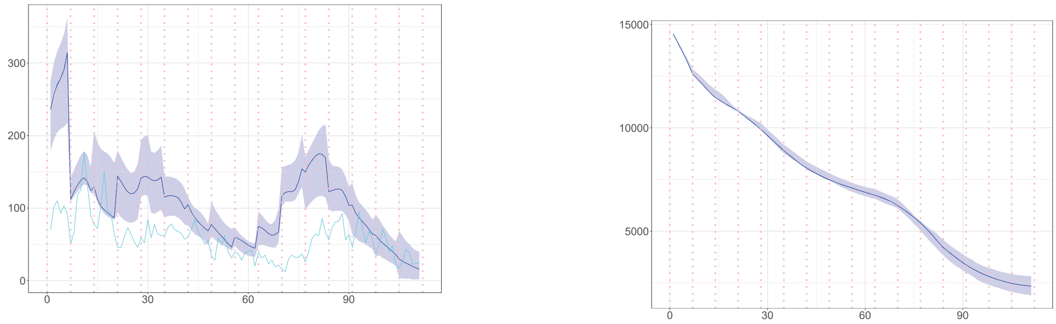
| Proposed Method | | | | | |
|---|---|---|---|---|---|
| Reported infections | Posterior Mean | Posterior Median | 95% CIs | True Number ($\mathcal{T}_{17}$) | True Number ($\mathcal{T}_{16}$) |
| aggregated | 327 | 318 | (195, 469) | 408 | |
| aged 0-29 | 67 | 62 | (22, 118) | 85 | 133 |
| aged 30-49 | 105 | 98 | (40, 178) | 107 | 197 |
| aged 50-69 | 98 | 91 | (34, 172) | 134 | 180 |
| aged 70+ | 57 | 54 | (20, 99) | 82 | 96 |

Table C.4: The true number of reported infections in $\mathcal{T}_{16}$ and $\mathcal{T}_{17}$, and the posterior median, the posterior mean and the 95% CIs of the estimated infections in $\mathcal{T}_{17}$ in Ashford.



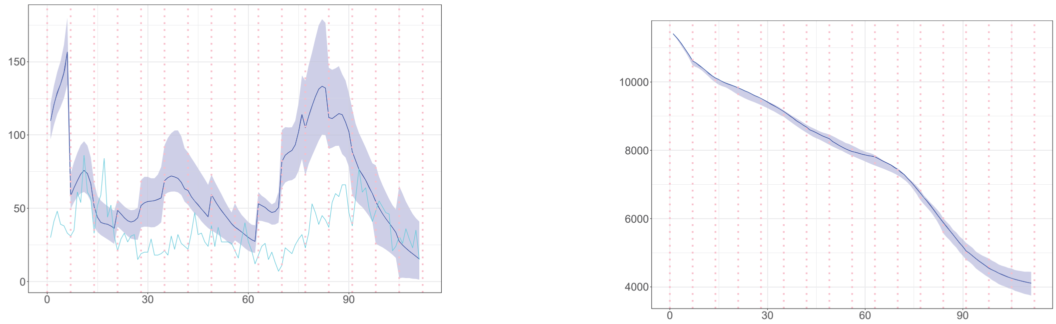
(a) The estimated intensity of latent cases aged 0-29.

(b) The estimated susceptibles aged 0-29.



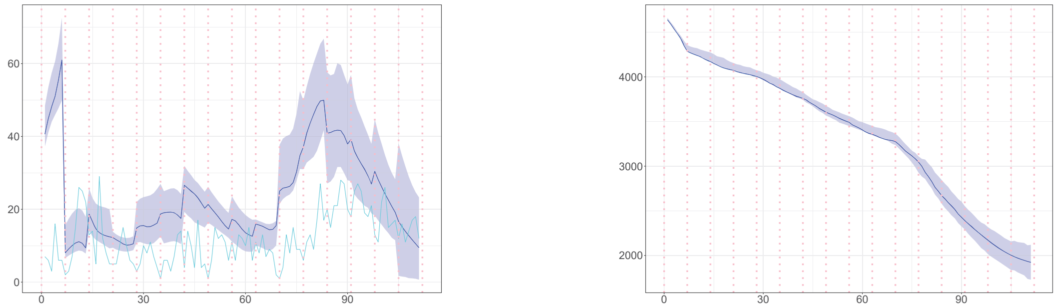
(c) The estimated intensity of latent cases aged 30-49.

(d) The estimated susceptibles aged 30-49.



(e) The estimated intensity of latent cases aged 50-69.

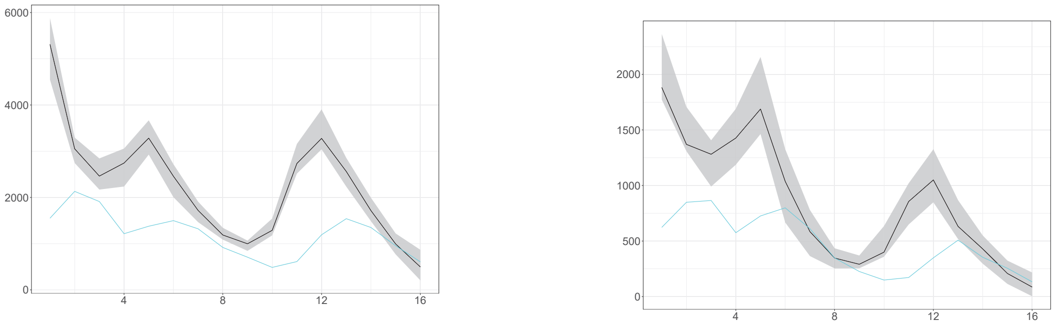
(f) The estimated susceptibles aged 50-69.



(g) The estimated intensity of latent cases aged 70+.

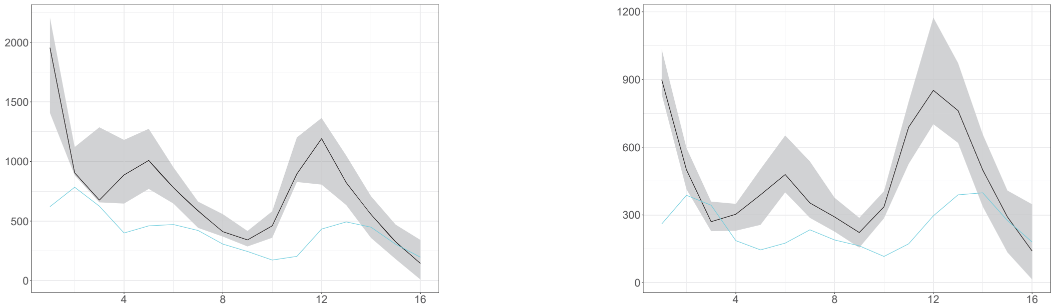
(h) The estimated susceptibles aged 70+.

Figure C.18: The estimated latent intensity and susceptibles (posterior median (blue line) ; 95% CI (ribbon)) and the daily observed cases (cyan line) in Ashford plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



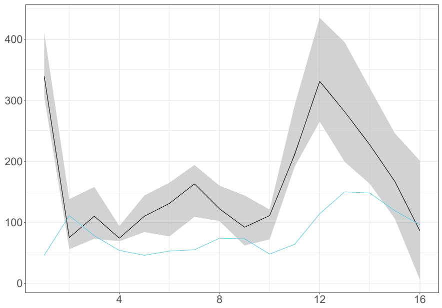
(a) The estimated aggregated weekly hidden cases.

(b) The estimated weekly hidden cases aged 0-29.



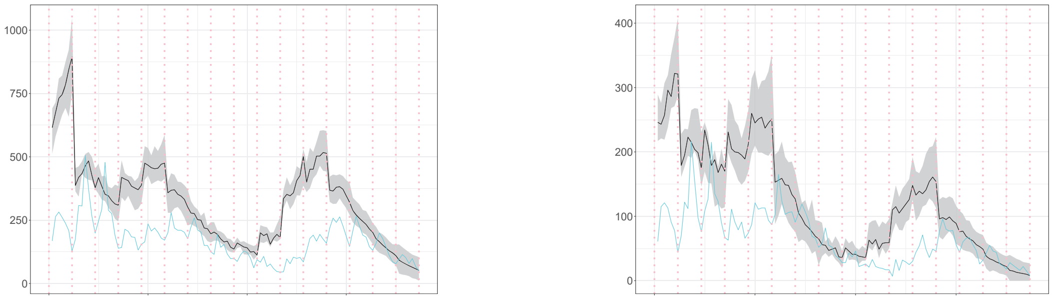
(c) The estimated weekly hidden cases aged 30-49.

(d) The estimated weekly hidden cases aged 50-69.

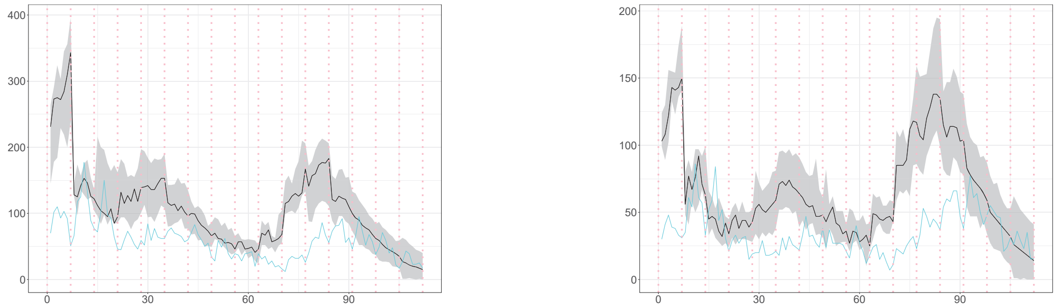


(e) The estimated weekly hidden cases aged 70+.

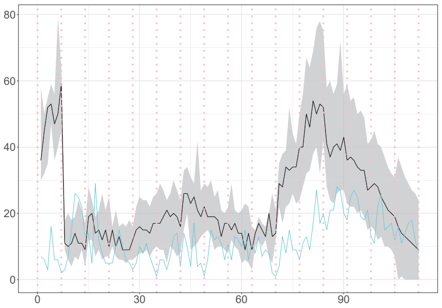
Figure C.19: The estimated weekly latent cases (black line; 95% CI (ribbon)) and the weekly observed cases (cyan line) in Ashford plotted against time.



(a) The estimated aggregated daily hidden cases. (b) The estimated daily hidden cases aged 0-29.

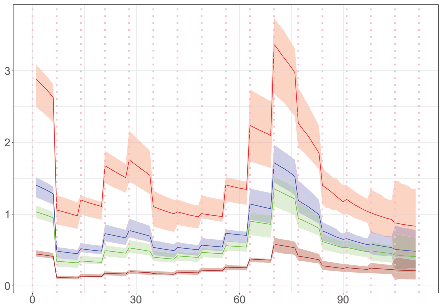


(c) The estimated daily hidden cases aged 30-49. (d) The estimated daily hidden cases aged 50-69.

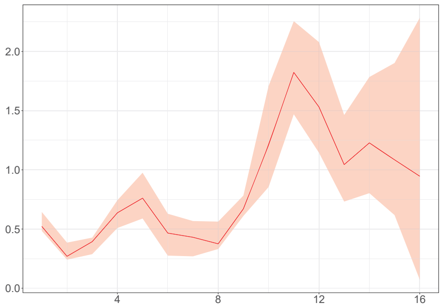


(e) The estimated daily hidden cases aged 70+.

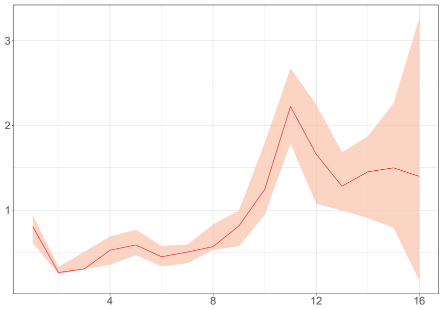
Figure C.20: The estimated daily latent cases (posterior median (black line); 95% CI (ribbon)) and the daily observed cases (cyan line) in Ashford plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



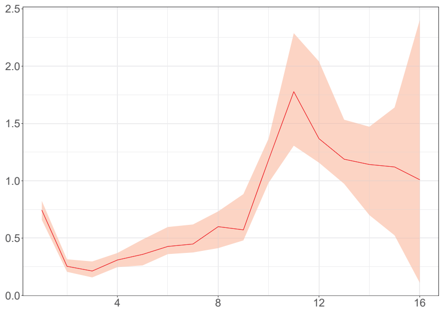
(a) The per age group instantaneous reproduction numbers.



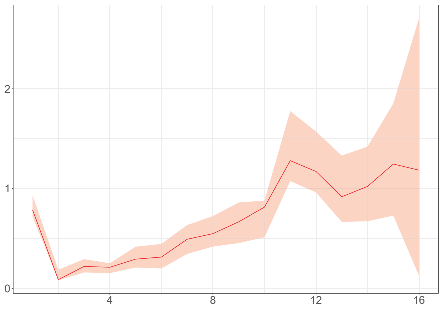
(b) The estimated weights $\{\gamma_{i,0-29}\}_{i=1}^{16}$.



(c) The estimated weights $\{\gamma_{i,30-49}\}_{i=1}^{16}$.

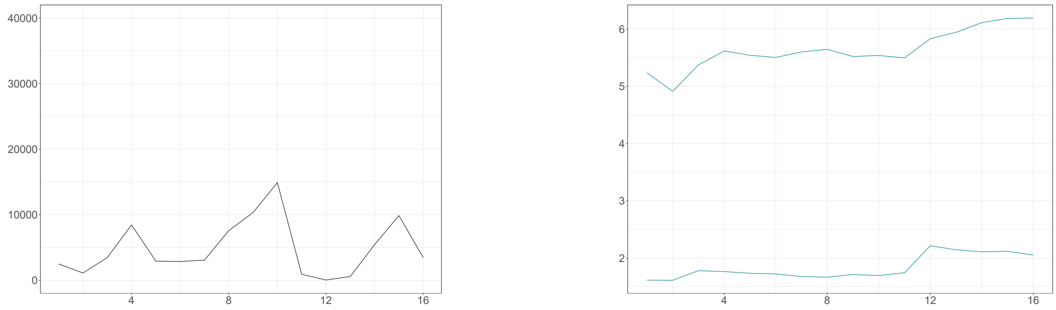


(d) The estimated weights $\{\gamma_{i,50-69}\}_{i=1}^{16}$.



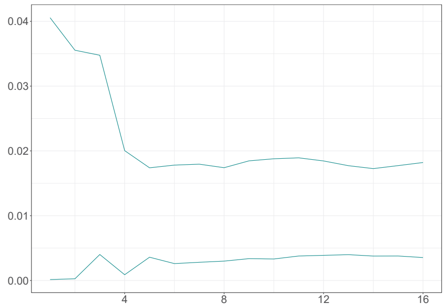
(e) The estimated weights $\{\gamma_{i,70+}\}_{i=1}^{16}$.

Figure C.21: The posterior median estimate of instantaneous reproduction number per age group (0-29 (red line), 30-49 (blue line), 50-69 (green line) and 70+ (brown line)), the posterior median estimate of weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ (red line) and the 95% CIs (ribbon) for Ashford plotted against time.



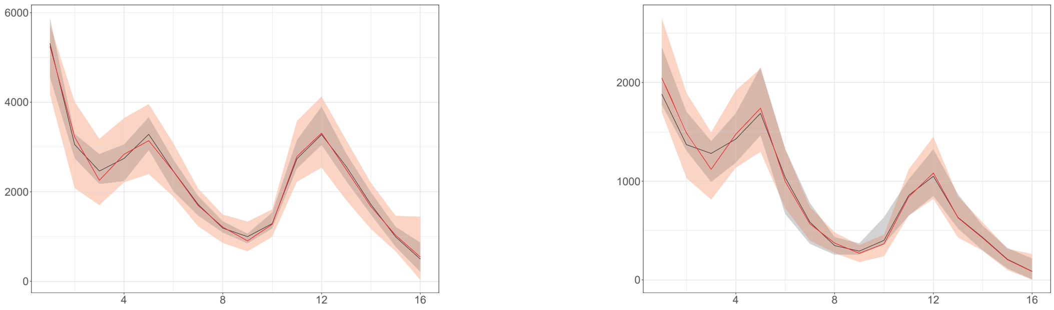
(a) ESS.

(b) The 95% CIs of $d$.



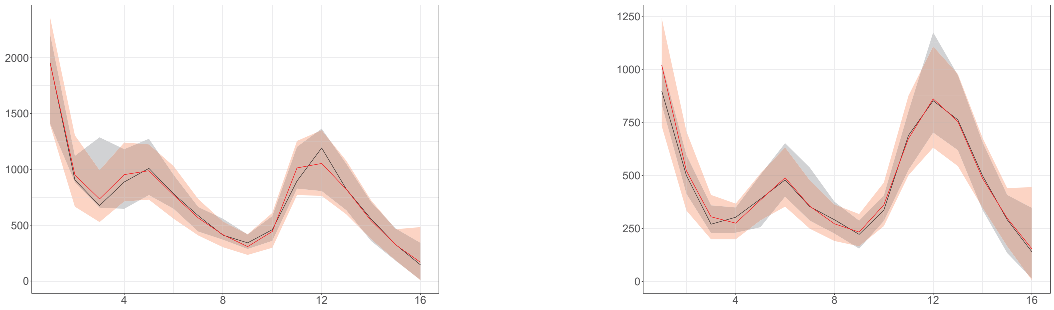
(c) The 95% CIs of $v$.

Figure C.22: ESS and 95% CIs of time-constant parameters for Ashford plotted against subintervals.



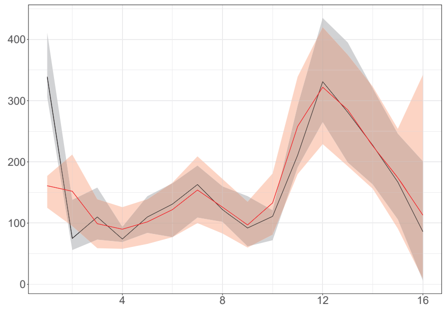
(a) Aggregated weekly hidden cases.

(b) Aged 0-29.



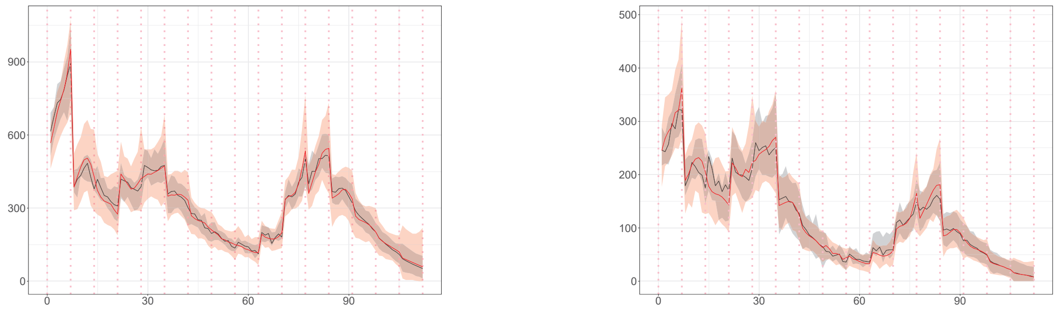
(c) Aged 30-49.

(d) Aged 50-69.

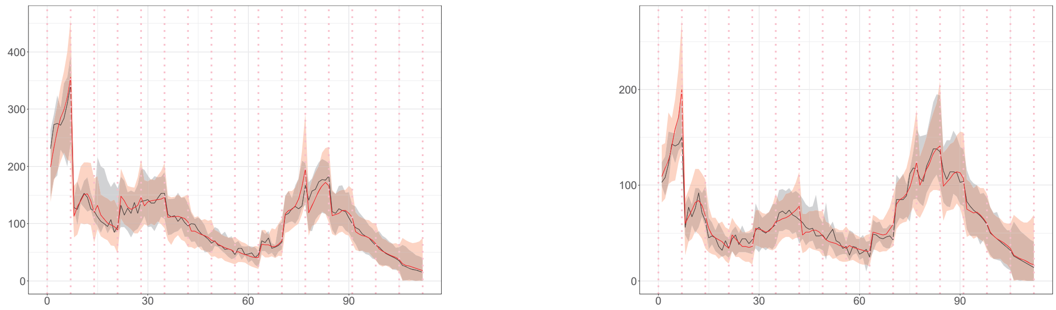


(e) Aged 70+.

Figure C.23: The posterior median estimate of weekly hidden cases of model A (black line) and model U (red line), and the 95% CIs (ribbon) in Ashford plotted against time.



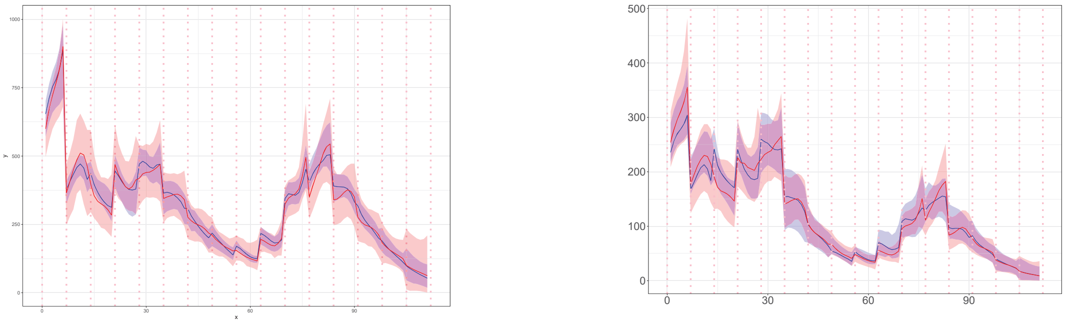
(a) Aggregated daily hidden cases. (b) Aged 0-29.



(c) Aged 30-49. (d) Aged 50-69.

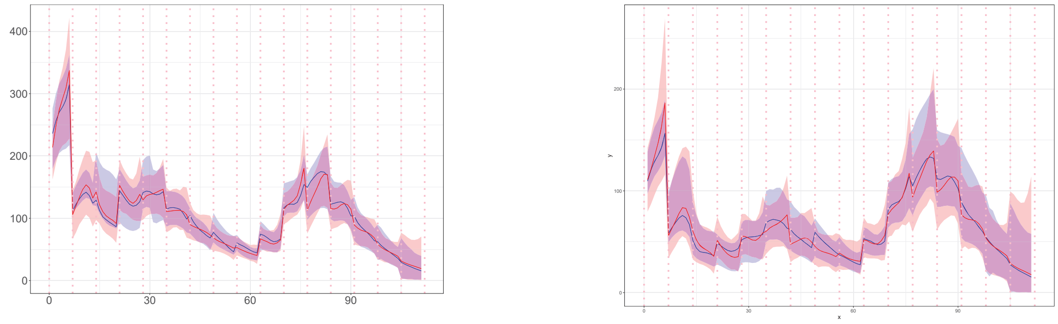
(e) Aged 70+.

Figure C.24: The posterior median estimate of daily hidden cases of model A (black line) and model U (red line), and the 95% CIs (ribbon) in Ashford plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



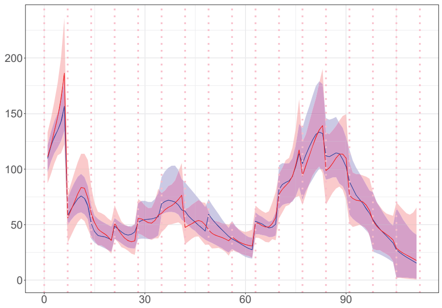
(a) Aggregated latent intensity.

(b) Aged 0-29.



(c) Aged 30-49.

(d) Aged 50-69.



(e) Aged 70+.

Figure C.25: The posterior median estimate of latent intensity of model A (blue line) and model U (red line), and the 95% CIs (ribbon) in Ashford plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.

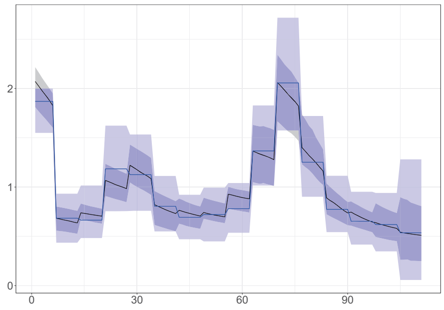
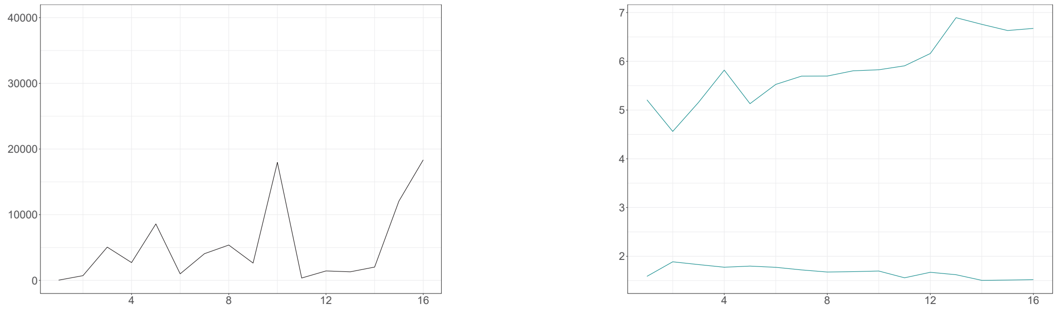


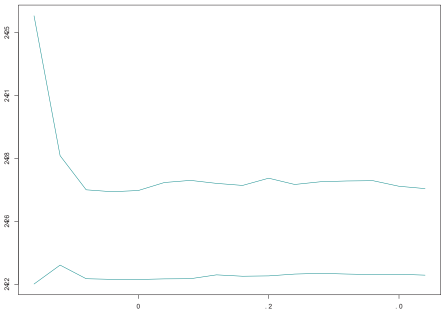
Figure C.26: The posterior median estimate of instantaneous reproduction number of model A (black line) and model U (blue line), and the 95% CIs (ribbon) in Ashford plotted against time.

### C.3.2 Kingston upon the Thames



(a) ESS.

(b) The 95% CIs of $d$.



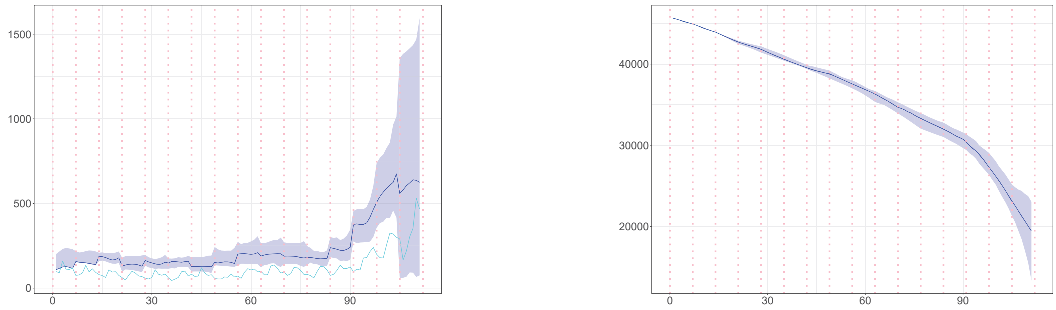
(c) The 95% CIs of $v$.

Figure C.27: ESS and 95% CIs of time-constant parameters for Kingston upon the Thames plotted against subintervals.

### C.3.3 Leicester

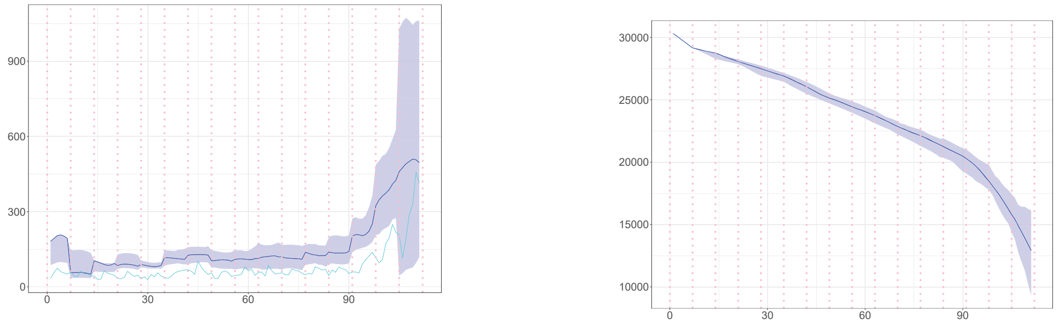
| Proposed Method | | | | | |
|---|---|---|---|---|---|
| Reported infections | Posterior Mean | Posterior Median | 95% CIs | True Number ($\mathcal{T}_{17}$) | True Number ($\mathcal{T}_{16}$) |
| aggregated | 4969 | 4793 | (2577, 7763) | 5794 | |
| aged 0-29 | 2275 | 2090 | (681, 4325) | 2367 | 1844 |
| aged 30-49 | 1659 | 1557 | (554, 3013) | 2054 | 1244 |
| aged 50-69 | 881 | 819 | (246, 1653) | 1112 | 528 |
| aged 70+ | 154 | 142 | (38, 290) | 261 | 87 |

Table C.5: The true number of reported infections in $\mathcal{T}_{16}$ and $\mathcal{T}_{17}$, and the posterior median, the posterior mean and the 95% CIs of the estimated infections in $\mathcal{T}_{17}$ in Leicester.



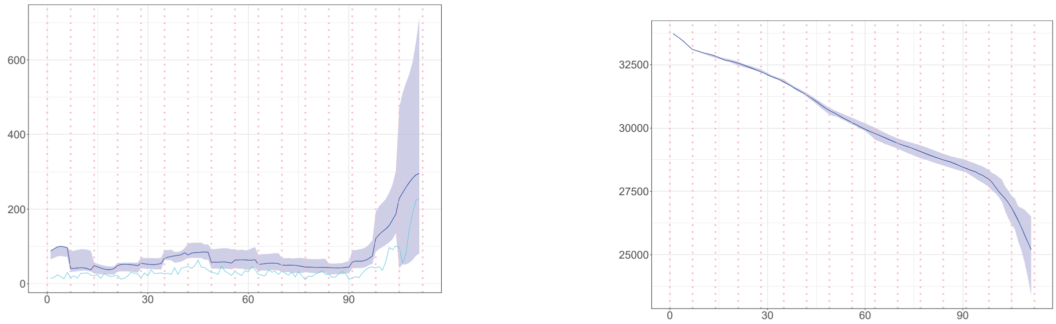
(a) The estimated intensity of latent cases aged 0-29.

(b) The estimated susceptibles aged 0-29.



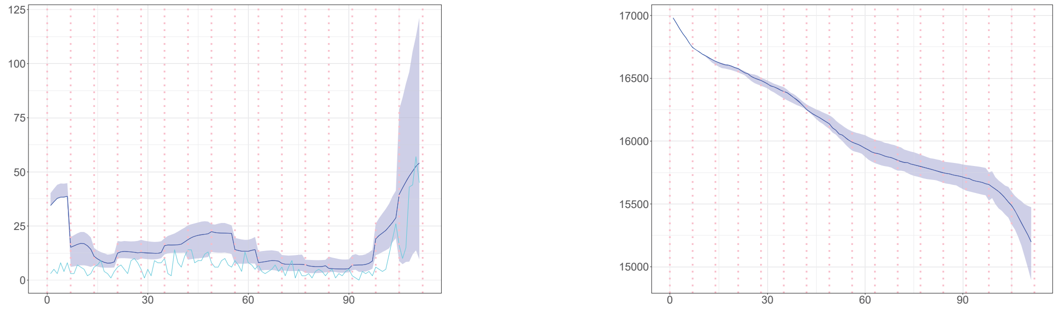
(c) The estimated intensity of latent cases aged 30-49.

(d) The estimated susceptibles aged 30-49.



(e) The estimated intensity of latent cases aged 50-69.

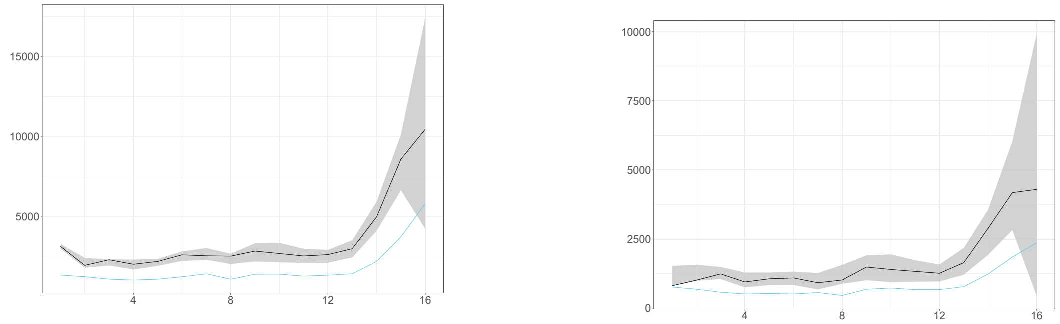
(f) The estimated susceptibles aged 50-69.



(g) The estimated intensity of latent cases aged 70+.

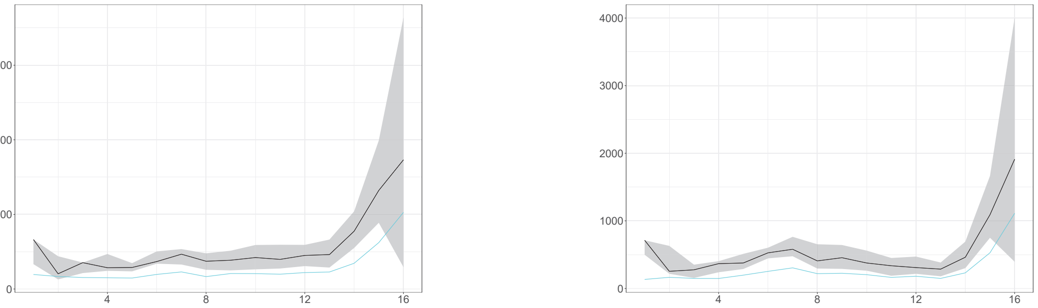
(h) The estimated susceptibles aged 70+.

Figure C.28: The estimated latent intensity and susceptibles (posterior median (blue line) ; 95% CI (ribbon)) and the daily observed cases (cyan line) in Leicester plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



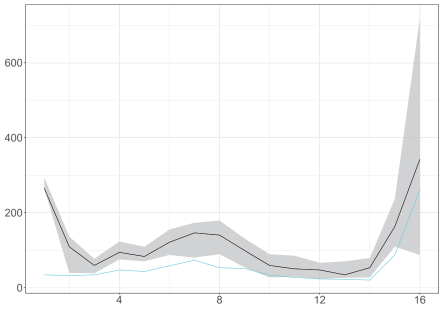
(a) The estimated aggregated weekly hidden cases.

(b) The estimated weekly hidden cases aged 0-29.



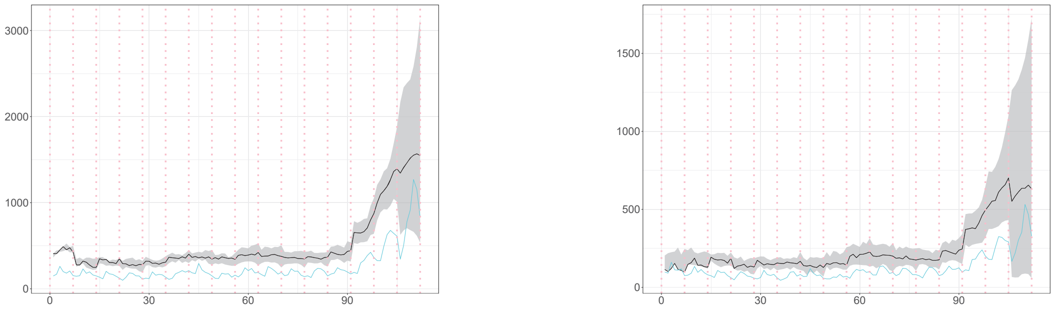
(c) The estimated weekly hidden cases aged 30-49.

(d) The estimated weekly hidden cases aged 50-69.

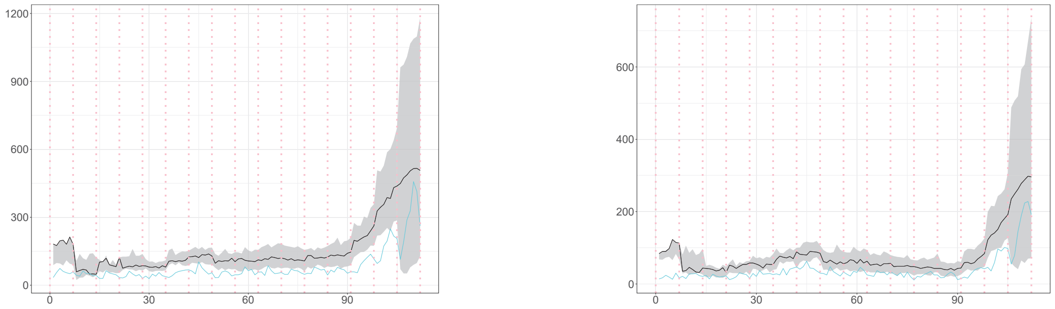


(e) The estimated weekly hidden cases aged 70+.

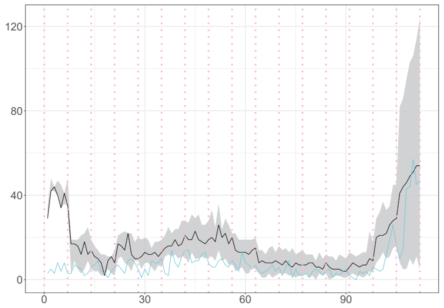
Figure C.29: The estimated weekly latent cases (black line; 95% CI (ribbon)) and the weekly observed cases (cyan line) in Leicester plotted against time.



(a) The estimated aggregated daily hidden cases. (b) The estimated daily hidden cases aged 0-29.

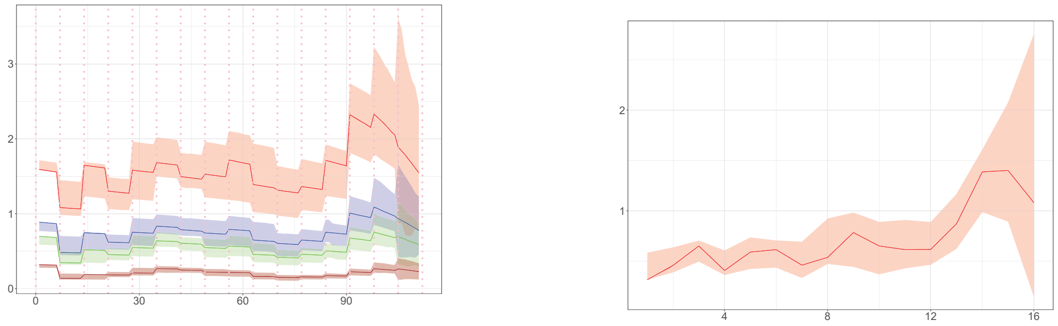


(c) The estimated daily hidden cases aged 30-49. (d) The estimated daily hidden cases aged 50-69.



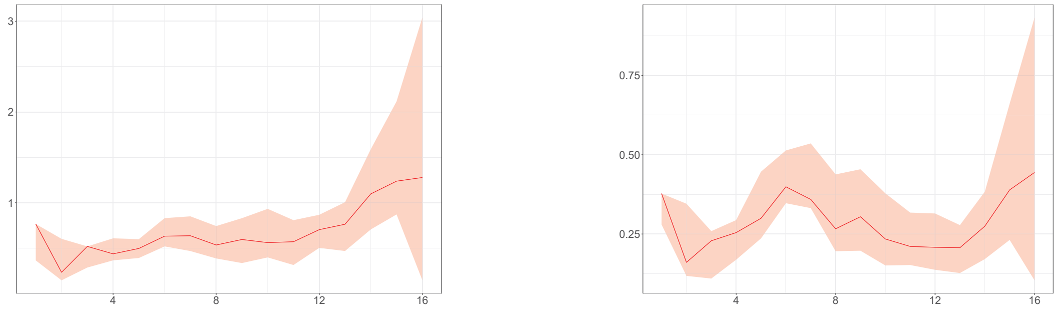
(e) The estimated daily hidden cases aged 70+.

Figure C.30: The estimated daily latent cases (posterior median (black line); 95% CI (ribbon)) and the daily observed cases (cyan line) in Leicester plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



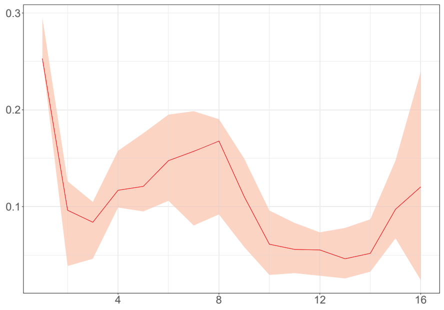
(a) The per age group instantaneous reproduction numbers.

(b) The estimated weights $\{\gamma_{i,0-29}\}_{i=1}^{16}$.



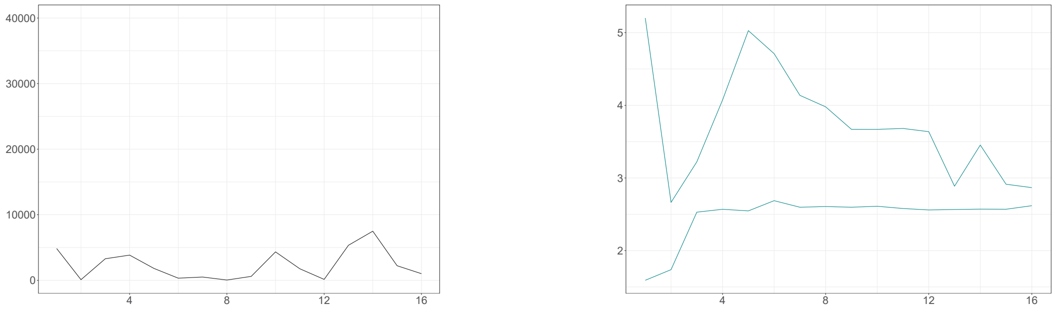
(c) The estimated weights $\{\gamma_{i,30-49}\}_{i=1}^{16}$.

(d) The estimated weights $\{\gamma_{i,50-69}\}_{i=1}^{16}$.



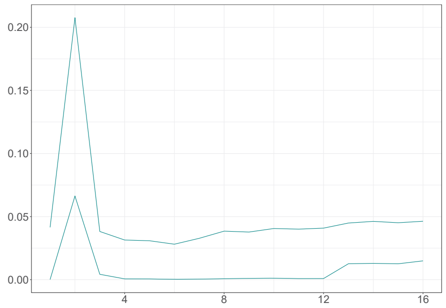
(e) The estimated weights $\{\gamma_{i,70+}\}_{i=1}^{16}$.

Figure C.31: The posterior median estimate of instantaneous reproduction number per age group (0-29 (red line), 30-49 (blue line), 50-69 (green line) and 70+ (brown line)), the posterior median estimate of weights $\{\{\gamma_{na}\}_{n=1}^{16}\}_a$ (red line) and the 95% CIs (ribbon) for Leicester plotted against time.



(a) ESS.

(b) The 95% CIs of $d$.

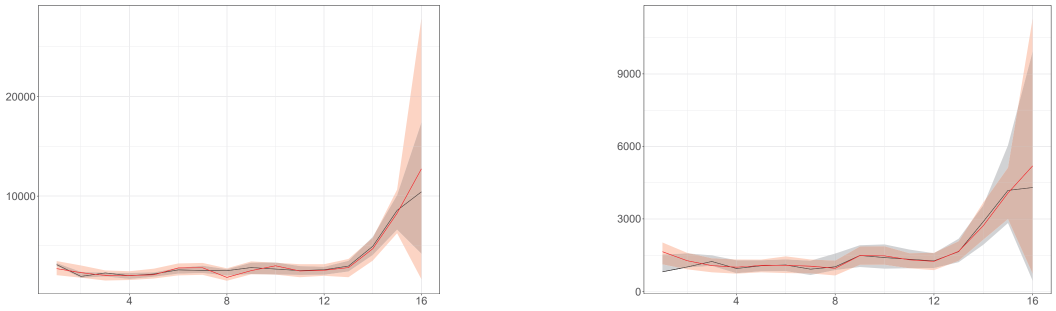


(c) The 95% CIs of $v$.

Figure C.32: ESS and 95% CIs of time-constant parameters for Leicester plotted against subintervals.

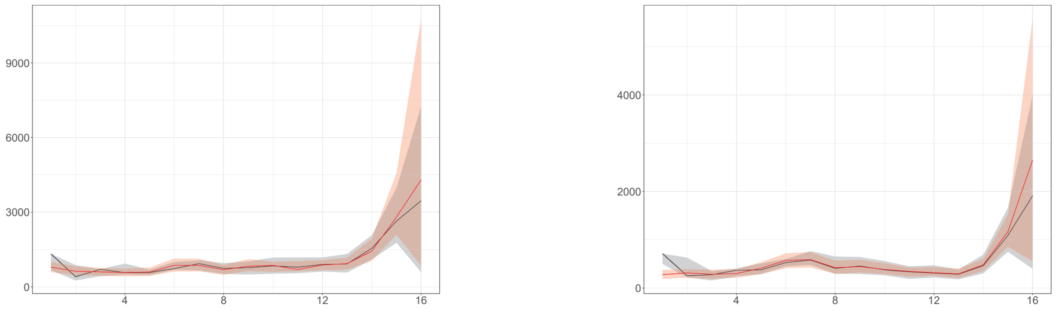
## C.4 Alternative method of estimating the susceptible population in Ashford and Kingston upon Thames

We use the percentage of the population aged $a$ with levels of antibodies against SARS-CoV-2 at or above a threshold of 179 nanograms per millilitre (ng/ml), denoted by $p_a$, available from the ONS [ONS, 2022], to initialize the number of susceptibles aged $a$ at the beginning of the first week. The percentage of people 0-15 having antibodies at or above the threshold is not available. We assume the percentage is the same as that of 16-24. Following this methodology, $(1 - p_a)N_a$ gives the susceptible population aged $a$ at the start of the first week. However, the estimated susceptible population is less than the total number of reported infections in Ashford and Kingston for some age bands. The antibodies might not fully protect against infection in December 2021; this could be due to declining immunity or immune escape (new variants being different from old variants and thus the previous infection being less protective against a new one). For this reason, we assume that a smaller percentage of the population aged $a$, given by $k_a p_a$ with $0 < k_a < 1$, has enough antibodies against a new infection in December 2021. We run Algorithm 16 for different values of $k_a$ and select the one that results in a 95% CI for the projected reported instances in week 17 ($\mathcal{T}_{17}$) that is narrower. Tables C.6-C.13 show the posterior mean, the posterior median and the 95% CI of the estimated reported infections in $\mathcal{T}_{17}$ for both local authorities. We require the susceptible population



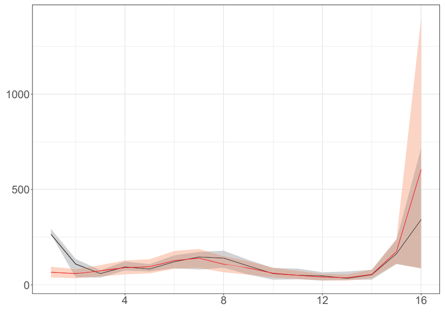
(a) Aggregated weekly hidden cases.

(b) Aged 0-29.



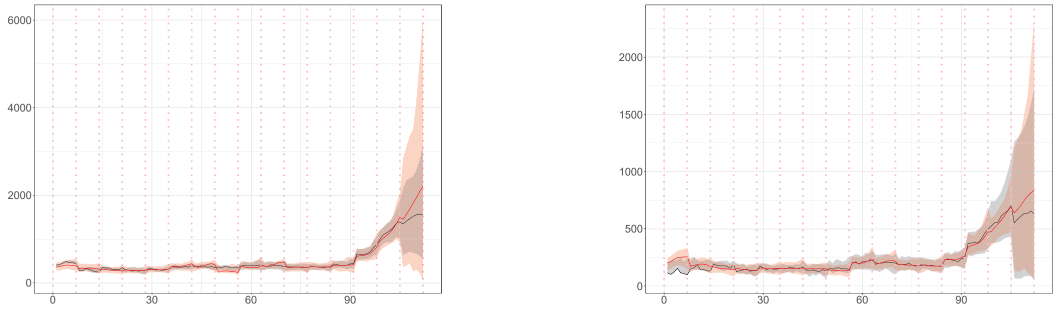
(c) Aged 30-49.

(d) Aged 50-69.



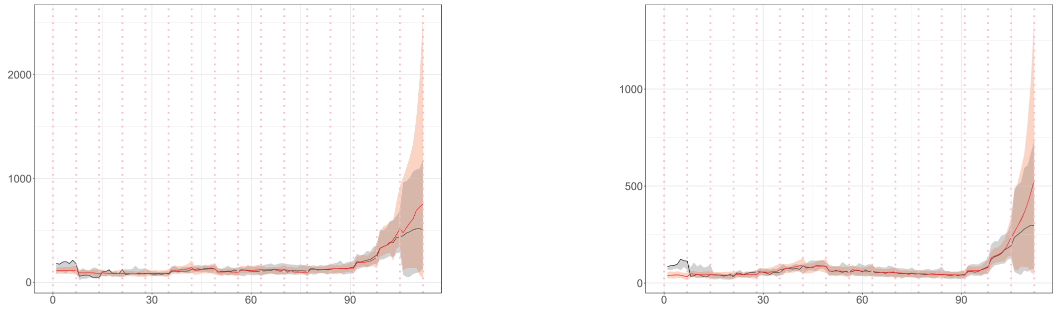
(e) Aged 70+.

Figure C.33: The posterior median estimate of weekly hidden cases of model A (black line) and model U (red line), and the 95% CIs (ribbon) in Leicester plotted against time.



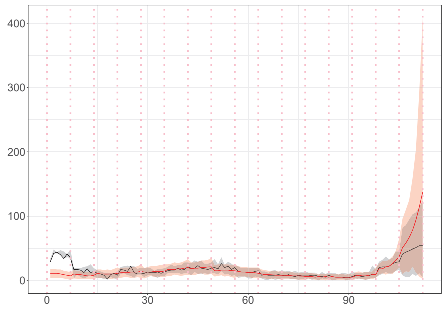
(a) Aggregated daily hidden cases.

(b) Aged 0-29.



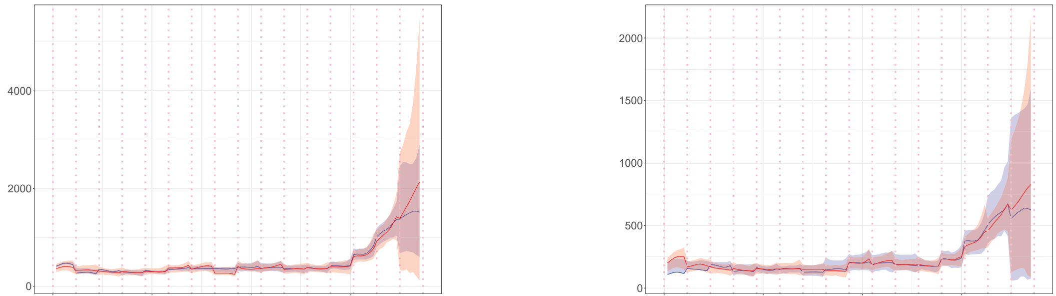
(c) Aged 30-49.

(d) Aged 50-69.



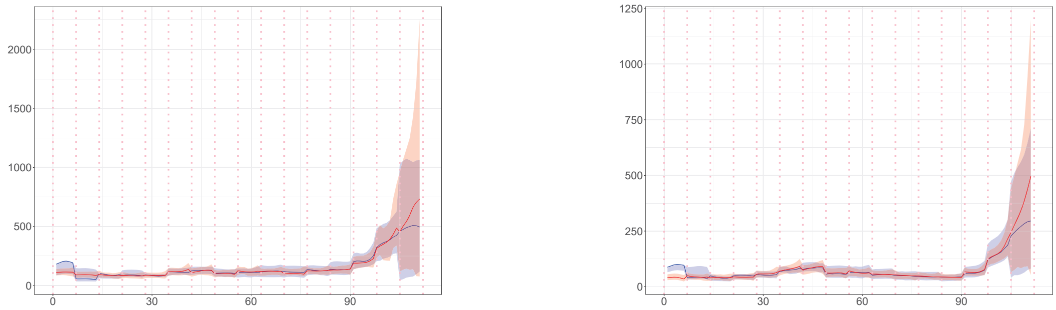
(e) Aged 70+.

Figure C.34: The posterior median estimate of daily hidden cases of model A (black line) and model U (red line), and the 95% CIs (ribbon) in Leicester plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.



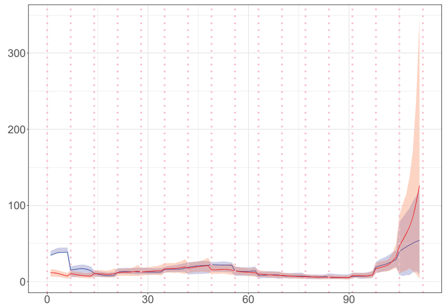
(a) Aggregated latent intensity.

(b) Aged 0-29.



(c) Aged 30-49.

(d) Aged 50-69.



(e) Aged 70+.

Figure C.35: The posterior median estimate of latent intensity of model A (blue line) and model U (red line), and the 95% CIs (ribbon) in Leicester plotted against time. The vertical dotted lines show the beginning of each week in the period we examine.

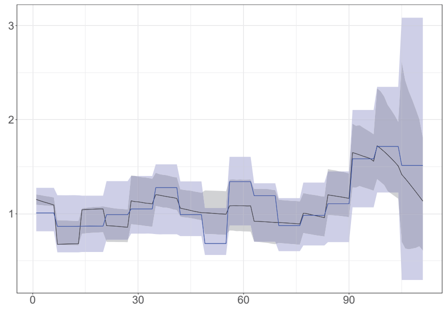


Figure C.36: The posterior median estimate of instantaneous reproduction number of model A (black line) and model U (blue line), and the 95% CIs (ribbon) in Leicester plotted against time.

to be at least twice as large as the reported cases for each age group to maintain consistency with our assumption that we see 50% of the infections. Summarizing the number of susceptibles aged $a$ at the beginning of the first week in Leicester (4/9/2021), Kingston upon Thames (11/12/2021) and Ashford (19/12/2021) are given by $(1-p_a)N_a$, $(1-0.4p_a)N_a$ and $(1-0.5p_a)N_a$, respectively.

| | Posterior estimates of the reported cases in $\mathcal{T}_{17}$. | | | |
|---|---|---|---|---|
| | True | Mean | Median | 95% CI |
| $Y_{0-29}$ | 85 | 70.3 | 63 | (21,133) |
| $Y_{30-49}$ | 107 | 115.38 | 106 | (44,209) |
| $Y_{50-69}$ | 134 | 101.59 | 94 | (35,180) |
| $Y_{70+}$ | 82 | 61.07 | 56 | (19,111) |

Table C.6: The true value, the posterior mean, the posterior median and the 95% CI of the reported cases per age group in $\mathcal{T}_{17}$ assuming $k_a = 0.4$ in Ashford.

| | Posterior estimates of the reported cases in $\mathcal{T}_{17}$. | | | |
|---|---|---|---|---|
| | True | Mean | Median | 95% CI |
| $Y_{0-29}$ | 85 | 67 | 63 | (25,114) |
| $Y_{30-49}$ | 107 | 113 | 107 | (49,186) |
| $Y_{50-69}$ | 134 | 106 | 101 | (49,176) |
| $Y_{70+}$ | 82 | 61.22 | 58 | (22,103) |

Table C.7: The true value, the posterior mean, the posterior median and the 95% CI of the reported cases per age group in $\mathcal{T}_{17}$ assuming $k_a = 0.5$ in Ashford.

| | Posterior estimates of the reported cases in $\mathcal{T}_{17}$. | | | |
|---|---|---|---|---|
| | True | Mean | Median | 95% CI |
| $Y_{0-29}$ | 85 | 85.36 | 67 | (11,249) |
| $Y_{30-49}$ | 107 | 110.39 | 102 | (35,199) |
| $Y_{50-69}$ | 134 | 101.58 | 93 | (33,192) |
| $Y_{70+}$ | 82 | 59.19 | 55 | (19,108) |

Table C.8: The true value, the posterior mean, the posterior median and the 95% CI of the reported cases per age group in $\mathcal{T}_{17}$ assuming $k_a = 0.6$ in Ashford.

| | Posterior estimates of the reported cases in $\mathcal{T}_{17}$. | | | |
|---|---|---|---|---|
| | True | Mean | Median | 95% CI |
| $Y_{0-29}$ | 85 | 69.19 | 64 | (22,123) |
| $Y_{30-49}$ | 107 | 106.42 | 101 | (43,178) |
| $Y_{50-69}$ | 134 | 102.15 | 97 | (41,175) |
| $Y_{70+}$ | 82 | 58.6 | 55 | (21,101) |

Table C.9: The true value, the posterior mean, the posterior median and the 95% CI of the reported cases per age group in $\mathcal{T}_{17}$ assuming $k_a = 0.7$ in Ashford.

| | Posterior estimates of the reported cases in $\mathcal{T}_{17}$. | | | |
|---|---|---|---|---|
| | True | Mean | Median | 95% CI |
| $Y_{0-29}$ | 139 | 153.81 | 146 | (70,248) |
| $Y_{30-49}$ | 246 | 245.8 | 236 | (124,383) |
| $Y_{50-69}$ | 187 | 162.53 | 155 | (79,255) |
| $Y_{70+}$ | 95 | 98.27 | 94 | (44,158) |

Table C.10: The true value, the posterior mean, the posterior median and the 95% CI of the reported cases per age group in $\mathcal{T}_{17}$ assuming $k_a = 0.4$ in Kingston.

| | Posterior estimates of the reported cases in $\mathcal{T}_{17}$. | | | |
|---|---|---|---|---|
| | True | Mean | Median | 95% CI |
| $Y_{0-29}$ | 139 | 147 | 138 | (62,246) |
| $Y_{30-49}$ | 246 | 241.28 | 229 | (116,390) |
| $Y_{50-69}$ | 187 | 164.13 | 156 | (73,269) |
| $Y_{70+}$ | 95 | 95.64 | 91 | (40,157) |

Table C.11: The true value, the posterior mean, the posterior median and the 95% CI of the reported cases per age group in $\mathcal{T}_{17}$ assuming $k_a = 0.5$ in Kingston.

| | Posterior estimates of the reported cases in $\mathcal{T}_{17}$. | | | |
|---|---|---|---|---|
| | True | Mean | Median | 95% CI |
| $Y_{0-29}$ | 139 | 154.1 | 142 | (52,277) |
| $Y_{30-49}$ | 246 | 241.46 | 223 | (94,427) |
| $Y_{50-69}$ | 187 | 171.12 | 155 | (56,322) |
| $Y_{70+}$ | 95 | 95.12 | 89 | (36,167) |

Table C.12: The true value, the posterior mean, the posterior median and the 95% CI of the reported cases per age group in $\mathcal{T}_{17}$ assuming $k_a = 0.6$ in Kingston.

| | Posterior estimates of the reported cases in $\mathcal{T}_{17}$. | | | |
|---|---|---|---|---|
| | True | Mean | Median | 95% CI |
| $Y_{0-29}$ | 139 | 146.13 | 131 | (46,278) |
| $Y_{30-49}$ | 246 | 222.5 | 203 | (74,409) |
| $Y_{50-69}$ | 187 | 157.8 | 144 | (51,291) |
| $Y_{70+}$ | 95 | 89.56 | 82 | (31,165) |

Table C.13: The true value, the posterior mean, the posterior median and the 95% CI of the reported cases per age group in $\mathcal{T}_{17}$ assuming $k_a = 0.7$ in Kingston.

# Bibliography

Ryan Prescott Adams, Iain Murray, and David JC MacKay. Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 9–16. ACM, 2009.

Arianna Agosto and Paolo Giudici. A Poisson autoregressive model to understand COVID-19 contagion dynamics. *Risks*, 2020.

Linda JS Allen. An introduction to stochastic epidemic models. In *Mathematical Epidemiology*, pages 81–130. Springer, 2008.

Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2010.

Andrea Apolloni, Chiara Poletto, and Vittoria Colizza. Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic. *BMC Infectious Diseases*, 13, 2013.

Elja Arjas and Dario Gasbarra. Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, pages 505–524, 1994.

Sergio Arregui, María José Iglesias, Sofía Samper, Dessislava Marinova, Carlos Martin, Joaquín Sanz, and Yamir Moreno. Data-driven model for the assessment of mycobacterium tuberculosis transmission in evolving demographic structures. *Proceedings of the National Academy of Sciences*, 2018.

Adrian Baddeley and Rolf Turner. spatstat: An R package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, 12:1–42, 2005.

Adrian Baddeley, Imre Bárány, and Rolf Schneider. Spatial point processes and their applications. *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13–18, 2004*, pages 1–75, 2007.

Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial point patterns: methodology and applications with R*. CRC press, 2015.

Fadoua Balabdaoui and Dirk Mohr. Age-stratified discrete compartment model of the COVID-19 epidemic with application to Switzerland. *Scientific Reports*, 10 (1), 2020.

Christopher D Barr and Frederic Paik Schoenberg. On the Voronoi estimator for the intensity of an inhomogeneous planar Poisson process. *Biometrika*, 97(4):977–984, 2010.

Robert Bartoszynski. Branching processes and the theory of epidemics. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 1967.

Guillaume Béraud, Sabine Kazmercziak, Philippe Beutels, Daniel Levy-Bruhl, Xavier Lenne, Nathalie Mielcarek, Yazdan Yazdanpanah, Pierre-Yves Boëlle, Niel Hens, and Benoit Dervaux. The french connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PlOS ONE*, 2015.

Andrea L Bertozzi, Elisa Franco, George Mohler, Martin B Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences*, 2020.

Alexandros Beskos, Dan Crisan, Ajay Jasra, Kengo Kamatani, and Yan Zhou. A stable particle filter for a class of high-dimensional state-space models. *Advances in Applied Probability*, 49:24–48, 2017.

Rico Blaser and Piotr Fryzlewicz. Random rotation ensembles. *The Journal of Machine Learning Research*, 2016.

Rico Blaser and Piotr Fryzlewicz. Regularizing axis-aligned ensembles via data rotations that favor simpler learners. *Statistics and Computing*, 2021.

Justin Bleich and Adam Kapelner. Bayesian Additive Regression Trees with Parametric Models of Heteroskedasticity. *arXiv preprint arXiv:1402.5397*, 2014.

Justin Bleich, Adam Kapelner, Edward I George, Shane T Jensen, et al. Variable selection for BART: an application to gene regulation. *The Annals of Applied Statistics*, 8(3):1750–1781, 2014.

Pierre Brémaud and Laurent Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996.

Anders Brix and Peter J Diggle. Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):823–841, 2001.

Raiha Browning, Deborah Sulem, Kerrie Mengersen, Vincent Rivoirard, and Judith Rousseau. Simple discrete-time self-exciting models can describe complex dynamic processes: A case study of COVID-19. *PLOS ONE*, 2021.

Angelo Canty and B. D. Ripley. *boot: Bootstrap R (S-Plus) Functions*, 2019. R package version 1.3-22.

Simon Cauchemez, Alain-Jacques Valleron, Pierre-Yves Boelle, Antoine Flahault, and Neil M Ferguson. Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452(7188):750–754, 2008.

Muge Cevik and Stefan D Baral. Networks of SARS-CoV-2 transmission. *Science*, 373(6551):162–163, 2021.

Muge Cevik, Julia L Marcus, Caroline Buckee, and Tara C Smith. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission dynamics should inform policy. *Clinical Infectious Diseases*, 73(Supplement_2):S170–S176, 2021.

Yu Chen, Jin Cheng, Yu Jiang, and Keji Liu. A time delay dynamic system with external source for the local outbreak of 2019-nCoV. *Applicable Analysis*, pages 1–12, 2020.

Wen-Hao Chiang, Xueying Liu, and George Mohler. Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *International Journal of Forecasting*, 2021.

Hugh A Chipman, Edward I George, and Robert E McCulloch. Bayesian CART model search. *Journal of the American Statistical Association*, 93(443):935–948, 1998.

Hugh A Chipman, Edward I George, Robert E McCulloch, et al. BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298, 2010.

Hugh A Chipman, Edward I George, Robert E McCulloch, and Thomas S Shively. mBART: Multidimensional Monotone BART. *Bayesian Analysis*, 1(1):1–30, 2021.

Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, 2013.

Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2022.

David R Cox and Valerie Isham. *Point processes*. CRC Press, 1980.

Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.

Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer, 2007.

Nicholas G Davies, Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, and Rosalind M Eggo. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, 26(8):1205–1211, 2020.

Tilman M Davies and Adrian Baddeley. Fast computation of spatially adaptive kernel estimates. *Statistics and Computing*, 28(4):937–956, 2018.

O le Polain de Waroux, Sandra Cohuet, Donny Ndazima, AJ Kucharski, Aitana Juan-Giner, Stefan Flasche, Elioda Tumwesigye, Rinah Arinaitwe, Juliet Mwanga-Amumpaire, Yap Boum, et al. Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: a survey in Southwest Uganda. *BMC Infectious Diseases*, 18(1):1–12, 2018.

Peter J Diggle, Barry Rowlingson, and Ting-li Su. Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics: The official Journal of the International Environmetrics Society*, 16(5):423–434, 2005.

Peter J Diggle et al. *Statistical analysis of spatial point patterns.* 2nd ed., Academic press, 2003.

Robert P Dobrow. *Introduction to stochastic processes with R.* John Wiley & Sons, 2016.

R. Douc and O. Cappe. Comparison of resampling schemes for particle filtering. In *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005.*, pages 64–69, 2005.

Arnaud Doucet, Adam M Johansen, et al. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704):3, 2009.

Ken TD Eames, Natasha L Tilston, Ellen Brooks-Pollock, and W John Edmunds. Measured dynamic social contact patterns explain the spread of H1N1V influenza. *PLOS Computational Biology*, 2012.

W John Edmunds, CJ O'callaghan, and DJ Nokes. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1384):949–957, 1997.

Akira Endo, Edwin Van Leeuwen, and Marc Baguelin. Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. *Epidemics*, 29:100363, 2019.

Juan V Escobar. A Hawkes process model for the propagation of COVID-19: Simple analytical results. *EPL (Europhysics Letters)*, 131(6):68005, 2020.

Xuhui Fan, Bin Li, Yi Wang, Yang Wang, and Fang Chen. The Ostomachion Process. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

Xuhui Fan, Bin Li, and Scott Sisson. The binary space partitioning-tree process. In *International Conference on Artificial Intelligence and Statistics*, pages 1859–1867. PMLR, 2018.

Xuhui Fan, Bin Li, and Scott Sisson. Binary space partitioning forest. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3022–3031. PMLR, 2019.

C Paddy Farrington, Mona N Kanaan, and Nigel J Gay. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50 (3):251–292, 2001.

Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College in London, 2020. DOI:10.25561/77482.

Paul EM Fine. The interval between successive cases of an infectious disease. *American Journal of Epidemiology*, 158(11):1039–1047, 2003.

Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020.

Piotr Fryzlewicz. *haarfisz: Software to perform Haar Fisz transforms*, 2010. URL `https://CRAN.R-project.org/package=haarfisz`. R package version 4.5.

Piotr Fryzlewicz and Guy P Nason. A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics*, 13(3):621–638, 2004.

Michele Garetto, Emilio Leonardi, and Giovanni Luca Torrisi. A time-modulated Hawkes process to model the spread of COVID-19 and the impact of countermeasures. *Annual Reviews in Control*, 2021.

John J Gart. The mathematical analysis of an epidemic with two kinds of susceptibles. *Biometrics*, pages 557–566, 1968.

Shufei Ge, Shijia Wang, Yee Whye Teh, Liangliang Wang, and Lloyd Elliott. Random tessellation forests. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472, 1992.

Andrew Gelman, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016, 2014.

Andrew Gelman, Daniel Lee, and Jiqiang Guo. Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, 40(5):530–543, 2015.

Megan L. Gelsinger, Maryclare Griffin, David S. Matteson, and Joseph Guinness. Log-Gaussian Cox Process Modeling of Large Spatial Lightning Data using Spectral and Laplace Approximations. *To appear in Annals of Applied Statistics*, 2022.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

M Gabriela M Gomes, Marcelo U Ferreira, Rodrigo M Corder, Jessica G King, Caetano Souto-Maior, Carlos Penha-Gonçalves, Guilherme Gonçalves, Maria Chikina, Wesley Pegden, and Ricardo Aguas. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *Journal of Theoretical Biology*, 2022.

GOV.UK. Ashford Reported Cases, 2022a. URL `https://coronavirus.data.gov.uk/details/download`. [Online; accessed 29-April-2022].

GOV.UK. The R value and growth rate, 2022b. URL `https://www.gov.uk/guidance/the-r-value-and-growth-rate`. [Online; last time accessed 24-November-2022].

GOV.UK. Kingston upon Thames Reported Cases, 2022c. URL `https://coronavirus.data.gov.uk/details/download`. [Online; accessed 21-April-2022].

GOV.UK. Leicester Reported Cases, 2022d. URL `https://coronavirus.data.gov.uk/details/download`. [Online; accessed 12-January-2022].

Shota Gugushvili, Frank van der Meulen, Moritz Schauer, and Peter Spreij. Fast and scalable non-parametric Bayesian inference for Poisson point processes. *arXiv preprint arXiv:1804.03616*, 2018.

Giorgio Guzzetta, Marco Ajelli, Zhenhua Yang, Stefano Merler, Cesare Furlanello, and Denise Kirschner. Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. *Journal of Theoretical Biology*, 2011.

Mark S Handcock, Adrian E Raftery, and Jeremy M Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354, 2007.

Theodore Edward Harris. *The theory of branching processes.* Springer Berlin, 1963.

Alan G Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):438–443, 1971a.

Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971b.

Alan G Hawkes. Hawkes processes and their applications to finance: a review. *Quantitative Finance*, 18(2):193–198, 2018.

Juha Heikkinen and Elja Arjas. Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scandinavian Journal of Statistics*, 25(3):435–450, 1998.

Niel Hens, Girma Minalu Ayele, Nele Goeyvaerts, Marc Aerts, Joel Mossong, John W Edmunds, and Philippe Beutels. Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight european countries. *BMC Infectious Diseases*, 9(1):1–12, 2009.

Jennifer L Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.

Janine Illian, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan. *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons, 2008.

Janine B. Illian, Sigrunn H. Sørbye, and Håvard Rue. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, 6(4):1499 – 1530, 2012.

Valerie Isham and Graham Medley. *Models for infectious human diseases: their structure and relation to data.* Cambridge University Press, 1996.

Christine Jacob. Branching processes: their role in epidemiology. *International Journal of Environmental Research and Public Health*, 7(3):1186–1204, 2010.

Christopher I Jarvis, Kevin Van Zandvoort, Amy Gimma, Kiesha Prem, Petra Klepac, G James Rubin, and W John Edmunds. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Medicine*, 18(1):1–10, 2020.

James Holland Jones. Notes on R0. Unpublished. Department of Anthropological Sciences: Stanford, CA, USA., 2007.
Available https://web.stanford.edu/ jhj1/teachingdocs/Jones-on-R0.pdf.

Nikolas Kantas, Arnaud Doucet, Sumeetpal S Singh, Jan Maciejowski, and Nicolas Chopin. On particle methods for parameter estimation in state-space models. *Statistical Science*, 30(3):328–351, 2015.

Adam Kapelner and Justin Bleich. bartMachine: Machine learning with Bayesian additive regression trees. *Journal of Statistical Software*, 70(4):1–40, 2016.

Minkyoung Kim, Dean Paini, and Raja Jurdak. Modeling stochastic processes in disease spread across a heterogeneous social system. *Proceedings of the National Academy of Sciences*, 116(2):401–406, 2019.

Bereket P Kindo, Hao Wang, and Edsel A Peña. Multinomial probit Bayesian additive regression trees. *Stat*, 5(1):119–131, 2016.

Shinsuke Koyama, Taiki Horie, and Shigeru Shinomoto. Estimating the time-varying reproduction number of COVID-19 with a state-space method. *PLOS Computational Biology*, 2021.

C Kresin, FP Schoenberg, and G Mohler. Comparison of the Hawkes and SEIR Models for the Spread of COVID-19 2020. *Advances and Applications in Statistics.*

Balaji Lakshminarayanan, Daniel Roy, and Yee Whye Teh. Particle Gibbs for Bayesian Additive Regression Trees. In *Artificial Intelligence and Statistics*, pages 553–561. PMLR, 2015.

Stamatina Lamprinakou and Axel Gandy. Age-stratified epidemic model using a latent marked Hawkes process. *arXiv preprint arXiv:2208.09555*, 2022.

Stamatina Lamprinakou, Mauricio Barahona, Seth Flaxman, Sarah Filippi, Axel Gandy, and Emma J. McCoy. Bart-based inference for Poisson processes. *Computational Statistics & Data Analysis*, 2023a.

Stamatina Lamprinakou, Axel Gandy, and Emma McCoy. Using a latent hawkes process for epidemiological modelling. *PLOS ONE*, 2023b.

Patrick J Laub, Thomas Taimre, and Philip K Pollett. Hawkes processes. *arXiv preprint arXiv:1507.02822*, 2015.

Patrick J Laub, Young Lee, and Thomas Taimre. *The elements of Hawkes processes.* Springer, 2021.

Thomas J Leininger and Alan E Gelfand. Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. *Bayesian Analysis*, 12(1):1–30, 2017.

Jacob E Lemieux, Katherine J Siddle, Bennett M Shaw, Christine Loreth, Stephen F Schaffner, Adrianne Gladden-Young, Gordon Adams, Timelia Fink, Christopher H Tomkins-Tinch, Lydia A Krasilnikova, et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science*, 371(6529): eabe3261, 2021.

Kathy Leung, Mark Jit, Eric HY Lau, and Joseph T Wu. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Scientific Reports*, 7(1):1–12, 2017.

PA W Lewis and Gerald S Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

Bo Li, Thomas Bengtsson, and Peter Bickel. Curse-of-dimensionality revisited: Collapse of importance sampling in very large scale systems. *Technical Report, Department of Statistics, UC-Berkeley*, 2005.

Antonio R Linero. Bayesian regression trees for high-dimensional prediction and variable selection. *Journal of the American Statistical Association*, 113(522):626–636, 2018.

Antonio R Linero and Yun Yang. Bayesian regression tree ensembles that adapt to smoothness and sparsity. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(5):1087–1110, 2018.

Jane Liu and Mike West. Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice*, pages 197–223. Springer, 2001.

Alun L Lloyd. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1470):985–993, 2001.

Chris Lloyd, Tom Gunter, Michael Osborne, and Stephen Roberts. Variational inference for Gaussian process modulated Poisson processes. In *International Conference on Machine Learning*, pages 1814–1822, 2015.

Clive Loader. Local Regression and Likelihood Springer: New York. 1999.

Giancarlo De Luca, Kim Van Kerckhove, Pietro Coletti, Chiara Poletto, Nathalie Bossuyt, Niel Hens, and Vittoria Colizza. The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium. *BMC Infectious Diseases*, 18(1):1–16, 2018.

Mateus Maia, Keefe Murphy, and Andrew C Parnell. GP-BART: a novel Bayesian Additive Regression Trees approach using Gaussian processes. *arXiv preprint arXiv:2204.02112*, 2022.

Dominique Makowski, Mattan Ben-Shachar, and Daniel Lüdecke. bayestestr: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, 4(40):1541, 2019.

Luigi Marangi, Grazina Mirinaviciute, Elmira Flem, Gianpaolo Scalia Tomba, Giorgio Guzzetta, Birgitte Freiesleben de Blasio, and Piero Manfredi. The natural history of varicella zoster virus infection in Norway: Further insights on exogenous boosting and progressive immunity to herpes zoster. *PLOS ONE*, 2017.

Hongyuan Mei and Jason M Eisner. The neural Hawkes process: A neurally self-modulating multivariate point process. *Advances in Neural Information Processing Systems*, 30, 2017.

Jesper Møller and Rasmus Plenge Waagepetersen. *Statistical inference and simulation for spatial point processes.* CRC press, 2003.

Yamir Moreno, Romualdo Pastor-Satorras, and Alessandro Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 26(4):521–529, 2002.

Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5(3):e74, 2008.

Jared S Murray. Log-linear Bayesian additive regression trees for multinomial logistic and count regression models. *Journal of the American Statistical Association*, 116 (534):756–769, 2021.

Jessica R Murray and Jerry Svarc. Global positioning system data collection, processing, and analysis conducted by the us geological survey earthquake hazards program. *Seismological Research Letters*, 88(3):916–925, 2017.

Mark EJ Newman and Gesine Reinert. Estimating the number of communities in a network. *Physical Review Letters*, 117(7):078301, 2016.

Artem S Novozhilov. On the spread of epidemics in a closed heterogeneous population. *Mathematical Biosciences*, 215(2):177–185, 2008.

Yosihiko Ogata. Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, 155:471–507, 1999.

ONS. Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland, 2021. URL `https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland`. [Online; accessed 11-January-2022].

ONS. Antibodies against coronavirus (COVID-19), 2022. URL `https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/antibodies#antibodies`. [Online; accessed 27-April-2022].

Joonha Park and Edward L Ionides. Inference on high-dimensional implicit dynamic models using a guided intermediate resampling filter. *Statistics and Computing*, 30(5):1497–1522, 2020.

Prakash N Patil, Andrew TA Wood, et al. Counting process intensity estimation by orthogonal wavelet methods. *Bernoulli*, 10(1):1–24, 2004.

Sen Pei, Sasikiran Kandula, and Jeffrey Shaman. Differential effects of intervention timing on COVID-19 spread in the United States. *Science Advances*, 6(49): eabd6370, 2020.

Lorenzo Pellis, Simon Cauchemez, Neil M Ferguson, and Christophe Fraser. Systematic selection between age and household structure for models aimed at emerging epidemic predictions. *Nature Communications*, 11(1):1–11, 2020.

Roger Peng. Multi-dimensional point process models in R. *Journal of Statistical Software*, 8:1–27, 2003.

PHE-Group. Public Health England (PHE) Transmission Group, "Factors contributing to risk of SARS-CoV2 transmission in various settings", 2020. URL `https://www.gov.uk/government/publications/phe-factors-contributing-to-risk-of-sars-cov2-transmission-in-various-settings-26-november-2020`.

Michael K Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.

Matthew T Pratola et al. Efficient Metropolis–Hastings proposal mechanisms for Bayesian regression tree models. *Bayesian Analysis*, 11(3):885–911, 2016.

Tom Rainforth and Frank Wood. Canonical correlation forests. *arXiv preprint arXiv:1507.05444*, 2015.

Maria A Riolo, George T Cantwell, Gesine Reinert, and Mark EJ Newman. Efficient method for estimating the number of communities in a network. *Physical Review*, 96(3):032310, 2017.

Marian-Andrei Rizoiu, Swapnil Mishra, Quyu Kong, Mark Carman, and Lexing Xie. Sir-Hawkes: Linking epidemic models and Hawkes processes to model diffusions in finite populations. In *Proceedings of the 2018 world wide web conference*, pages 419–428, 2018.

Christian P Robert and George Casella. *Monte Carlo statistical methods*. Springer, 1999.

Christian P Robert and George Casella. *Introducing Monte Carlo methods with R*. Springer, 2010.

Veronika Rockova and Stephanie Pas. Posterior concentration for Bayesian Regression Trees and their Ensembles. *Annals of Statistics*, 08 2017.

Juan José Rodriguez, Ludmila I Kuncheva, and Carlos J Alonso. Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1619–1630, 2006.

Pejman Rohani, Xue Zhong, and Aaron A King. Contact network structure explains the changing epidemiology of pertussis. *Science*, 330(6006):982–985, 2010.

Lionel Roques, Etienne K Klein, Julien Papaix, Antoine Sar, and Samuel Soubeyrand. Using early data to estimate the actual infection fatality ratio from COVID-19 in France. *Biology*, 9(5):97, 2020.

Sheldon M Ross. *Stochastic processes*. John Wiley & Sons, 1995.

Veronika Ročková and Enakshi Saha. On Theory for BART. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 2839–2848. PMLR, 16–18 Apr 2019. URL `https://proceedings.mlr.press/v89/rockova19a.html`.

Veronika Ročková and Stéphanie van der Pas. Posterior concentration for Bayesian regression trees and forests. *The Annals of Statistics*, 48(4):2108 – 2131, 2020.

Sylvain Sardy and Paul Tseng. On the statistical analysis of smoothing by maximizing dirty Markov random field posterior distributions. *Journal of the American Statistical Association*, 99(465):191–204, 2004.

David W Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

Daniel M Sheinson, Jarad Niemi, and Wendy Meiring. Comparison of the performance of particle filter algorithms applied to tracking of a disease epidemic. *Mathematical Biosciences*, 255:21–32, 2014.

Christian Shelton, Zhen Qin, and Chandini Shetty. Hawkes process inference with missing data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Rodney A Sparapani, Brent R Logan, Robert E McCulloch, and Purushottam W Laud. Nonparametric survival analysis using Bayesian additive regression trees (BART). *Statistics in Medicine*, 35(16):2741–2753, 2016.

Theresa Stocks, Tom Britton, and Michael Höhle. Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany. *Biostatistics*, 21(3):400–416, 2020.

Mervyn Stone. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):44–47, 1977.

Geir Storvik, Alfonso Diz-Louis Palomares, Solveig Engebretsen, Gunnar Øyvind Isaksson Rø, Kenth Engø-Monsen, Aja Bråthen Kristoffersen, Birgitte Freiesleben de Blasio, and Arnoldo Frigessi. A sequential Monte Carlo approach to estimate a time varying reproduction number in infectious disease models: the COVID-19 case. *arXiv preprint arXiv:2201.07590*, 2022.

Alexei V Tkachenko, Sergei Maslov, Ahmed Elbanna, George N Wong, Zachary J Weiner, and Nigel Goldenfeld. Time-dependent heterogeneity leads to transient suppression of the COVID-19 epidemic, not herd immunity. *Proceedings of the National Academy of Sciences*, 118(17):e2015972118, 2021.

Tyler M Tomita, James Browne, Cencheng Shen, Jaewon Chung, Jesse L Patsolic, Benjamin Falk, Carey E Priebe, Jason Yim, Randal Burns, Mauro Maggioni, et al. Sparse projection oblique randomer forests. *Journal of Machine Learning Research*, 21(104), 2020.

S Triambak and DP Mahapatra. A random walk Monte Carlo simulation study of COVID-19-like infection spread. *Physica A: Statistical Mechanics and its Applications*, 574:126014, 2021.

MNM Van Lieshout. *Markov point processes and their applications.* World Scientific, 2000.

Jacco Wallinga, W John Edmunds, and Mirjam Kretzschmar. Perspective: human contact patterns and the spread of airborne infectious diseases. *Trends in Microbiology*, 7(9):372–377, 1999.

Jacco Wallinga, Peter Teunis, and Mirjam Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, 164(10):936–944, 2006.

MP Wand. Data-based choice of histogram bin width. *The American Statistician*, 51(1):59–64, 1997.

Jia Wangping, Han Ke, Song Yang, Cao Wenzhe, Wang Shengshu, Yang Shanshan, Wang Jianwei, Kou Fuyin, Tai Penggang, Li Jing, et al. Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *Frontiers in Medicine*, 7:169, 2020.

Mike West. Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(2):409–422, 1993.

Colin J Worby, Sandra S Chaves, Jacco Wallinga, Marc Lipsitch, Lyn Finelli, and Edward Goldstein. On the relative role of different age groups in influenza epidemics. *Epidemics*, 13:10–16, 2015.

Shuang-Hong Yang and Hongyuan Zha. Mixture of mutually exciting processes for viral diffusion. In *International Conference on Machine Learning*, pages 1–9. PMLR, 2013.

Junni L Zhang and Wolfgang K Härdle. The Bayesian additive classification tree applied to credit risk modelling. *Computational Statistics & Data Analysis*, 54 (5):1197–1205, 2010.

Shi Zhao, Biao Tang, Salihu S Musa, Shujuan Ma, Jiayue Zhang, Minyan Zeng, Qingping Yun, Wei Guo, Yixiang Zheng, Zuyao Yang, et al. Estimating the generation interval and inferring the latent period of COVID-19 from the contact tracing data. *Epidemics*, 36:100482, 2021.

P. Zheng, Peter A. Durr, and Peter J Diggle. Edge–correction for Spatial Kernel Smoothing Methods? When is it necessary? *GisVet Conference*, 2004.

Difan Zou, Lingxiao Wang, Pan Xu, Jinghui Chen, Weitong Zhang, and Quanquan Gu. Epidemic Model Guided Machine Learning for COVID-19 Forecasts in the United States. *medRxiv*, 2020. doi: 10.1101/2020.05.24.20111989.