Imperial College London

Department of Infectious Disease Epidemiology, School of Public Health

Statistical methods for characterising the severity of an emerging pathogen: case studies of the COVID-19 pandemic

by

Iwona Ewa Hawryluk

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy of Imperial College London

July 17, 2024

Abstract

Epidemiological modellers face significant challenges during outbreaks of novel pathogens, particularly the need for rapid and precise modelling with limited and delayed data. The COVID-19 pandemic revealed the insufficiency of current methods in addressing these challenges and highlighted the need for more innovative approaches to infectious disease modelling. Motivated by these issues, this thesis aims to aid epidemiologists in building more robust models by developing novel Bayesian methods that enhance the accuracy and reliability of epidemiological models, especially for emerging pathogens such as SARS-CoV-2.

Specific contributions of this thesis are as follows: Chapter 2 introduces a new method for estimating marginal likelihoods using thermodynamic integration, facilitating model selection through Bayes factors. Applied to COVID-19 data, this approach reveals the pitfalls of model selection and emphasises the importance of rigorous methods. In Chapter 3, a range of probability density functions is fitted to hospitalisation distributions for COVID-19 patients in Brazil, providing crucial inputs for epidemic models. Spatial heterogeneity in hospitalisation times is explored, offering insights into regional variations in disease dynamics. Chapter 4 shifts focus to data quality, investigating reporting delays in COVID-19 mortality data for Brazil. A novel method using Gaussian Processes is proposed to correct reporting delays, enabling real-time monitoring of epidemiological trends with greater accuracy. Building on these methodological advancements, Chapter 5 explores the impact of regularisation in a renewal-equation-based R_t model, demonstrating the importance of informative priors in accurately estimating importations during emerging epidemics.

The methods proposed here aim to assist infectious disease modellers in rapidly responding to emerging threats using Bayesian statistical tools. Overall, this work combines traditional epidemiological approaches with modern statistical and machine learning methods to address the challenges faced during the COVID-19 pandemic, providing a framework for improving infectious disease modelling in future outbreaks.

Contents

List of Figures

List of Tables

Acknowledgements

None of this work would have been possible without the constant support of my supervisors. Thanks to Sam for always being a source of guidance, having answers to my questions, and offering fresh ideas to try out. And thanks to Tom, who wasn't just my supervisor but also a friend. Thank you for always believing in my abilities more than I did, making me feel confident and comfortable to ask questions, and involving me in numerous interesting opportunities outside the scope of my PhD.

I also want to express my gratitude to other senior members of the group, whose inputs were fundamental in shaping my projects: Swapnil, Seth, and Oliver. Thank you for your patience, invaluable contributions, and insightful feedback.

Special thanks to the people I worked with on the various Brazil modelling projects during the pandemic, especially Tom, Henrique, Ricardo, and Charlie. Your dedication, friendliness, and teamwork made even the most stressful and busy times manageable and the work smoother. Your support and collaboration were truly invaluable.

I would have gotten nowhere without the help and support from my academic mentors: Liza, Ettie, and Xenia. Thank you for advising me on how to navigate the complexities of academic life, and more importantly, for becoming my good friends.

It was a pleasure to be part of the Machine Learning and Global Health group — even before it officially became that. Thanks to all the members for building a friendly, approachable, and insanely knowledgeable community of researchers and for always making me feel welcome at our many offices, in London, Oxford, Bristol, Copenhagen, and Singapore.

Thanks to my MRes/PhD cohort: Tori, Olivia, Sam, Lanre, and Jack. I couldn't have picked a better

group of peeps to share the joys and pains of doing a PhD. Thank you to Nora and Nora for being my sports, arts, and spritz companions.

The support and encouragement of my friends from back home meant the world to me throughout this journey. Massive thanks to Szczury: Kasia, Malwina, Pola, and Hania, and Ziemniaczki: Marta, Aga, Iśka, and Michał. You've all been there for me for more than half of my life, for better and for worse. Dziękuję! I love you all.

Thanks to Zdeněk for helping me unwind, relax, and have fun during the last year of my PhD. Děkuji!

Needless to say, I would have never been able to get this far without the support of my family. Thank you for encouraging me to reach for the stars.

Finally, I'd like to thank the countless individuals who contribute to the development and maintenance of freely available software and libraries, especially Stan and Numpyro, which have been indispensable to my research journey. Their commitment and active engagement on forums are invaluable to the scientific community and deserve more recognition and appreciation.

Copyright Declaration

Statement of originality

I declare that all the work presented in this thesis is my own, and any contributions from others are duly acknowledged through references or description provided below.

Chapter 2 has been published as **Hawryluk, I.**, Mishra, S., Flaxman, S., Bhatt, S., Mellan, T.A. (2023) "Application of referenced thermodynamic integration to Bayesian model selection." PLOS ONE, 18(8): e0289889. https://doi.org/10.1371/journal.pone.0289889. The code accompanying the paper is available at https://github.com/mrc-ide/referenced-TI. Part of this work has been done by myself and my co-authors during my MRes degree at Imperial College London in 2020. This work was then continued during my PhD, including substantial changes in the writing and carrying out additional analyses and writing and publishing the paper.

Chapter 3 has been published as **Hawryluk, I.**, Mellan, T.A., Hoeltgebaum, H., Mishra, S., Schnekenberg, R.P., Whittaker, C., Zhu, H., Gandy, A., Donnelly, C.A., Flaxman, S., and Bhatt, S (2020) "Inference of COVID-19 epidemiological distributions from Brazilian hospital data", Journal of The Royal Society Interface. Royal Society, 17(172). https://doi.org/10.1098/rsif.2020.0596. The code accompanying the paper is available at https://github.com/mrc-ide/Brazil_COVID19_distributions. Dr Thomas A. Mellan carried out the calculation of R_t for the nation-wide and state-specific hospitalisation densities estimates, and is appropriately acknowledged in the chapter where the figure he generated is shown. Dr Henrique Helfer Hoeltgebaum, Dr Ricardo P. Schnekenberg and Dr Thomas A. Mellan have provided the code for extracting and processing the data from the Brazilian hospitalisations database.

Chapter 4 has been published as **Hawryluk, I.**, Hoeltgebaum, H., Mishra, S., Miscouridou, X., Schnekenberg, R.P., Whittaker, C., Vollmer, M., Flaxman, S., Bhatt, S., Mellan, T.A. (2021) "Gaussian process nowcasting: application to COVID-19 mortality reporting." Proceedings of the Thirty-Seventh Confer-

ence on Uncertainty in Artificial Intelligence in Proceedings of Machine Learning Research, 161, 1258-1268. https://proceedings.mlr.press/v161/hawryluk21a.html. The code accompanying the paper is available at https://github.com/ihawryluk/GP_nowcasting. Dr Thomas A. Mellan carried out the calculation of R_t for the pre-nowcasted and nowcasted data and is appropriately acknowledged in the chapter where the figures he generated are shown.

The code accompanying Chapter 5 is available at https://github.com/ihawryluk/importations

Chapter 1

Introduction

1.1 COVID-19 pandemic

In December 2019 in Wuhan, Hubei Province in China, the health facilities reported several clusters of patients with pneumonia of unknown origin [Zhu et al., 2020]. The cases were linked to a local seafood and wet animal market. Upon the epidemiological investigation of the Chinese Centre for Disease Control, the sequenced samples from the patients revealed an infection with a novel coronavirus, named 2019 novel coronavirus or 2019-nCoV [Li et al., 2020b, Zhu et al., 2020]. The new pathogen spread quickly across Wuhan and then to other parts of China. Despite introduction of the travel restrictions in China, quarantines and lockdowns, the virus eventually spread worldwide, facilitated by international travel. The World Health Organisation (WHO) declared it a Public Health Emergency of International Concern on 30$^{\text{th}}$ January 2020. On the 11$^{\text{th}}$ Feb, the WHO officially named the disease caused by the novel virus COVID-19 [WHO, 2019]. On the same day, the International Committee on Taxonomy of Viruses, responsible for developing the classification of viruses and taxon nomenclature, announced a new name for the virus — severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), due to its genetic similarity to the coronavirus responsible for the 2003 SARS outbreak [Gorbalenya et al., 2020]. Finally, the WHO announced a global COVID-19 pandemic on 11$^{\text{th}}$ March 2020.

The rapid spread of SARS-CoV-2 prompted governments worldwide to implement various public health measures, including lockdowns, social distancing, mask mandates, and widespread testing, in an effort to slow transmission and reduce the burden on healthcare systems. Scientists worldwide also joined

forces trying to contribute to stopping the evolving pandemic and saving lives: from the development of treatment and vaccines, through analyses of the virus to understand its biology, testing the efficacy of different types of protective equipment, to modelling the spread of the virus given different hypothetical scenarios including introductions of non-pharmaceutical interventions.

The overarching topic of this thesis is the latter — proposing new statistical methods which can aid the epidemiological modeller in characterising the severity and spread of an emerging pathogen with case studies from the COVID-19 pandemic.

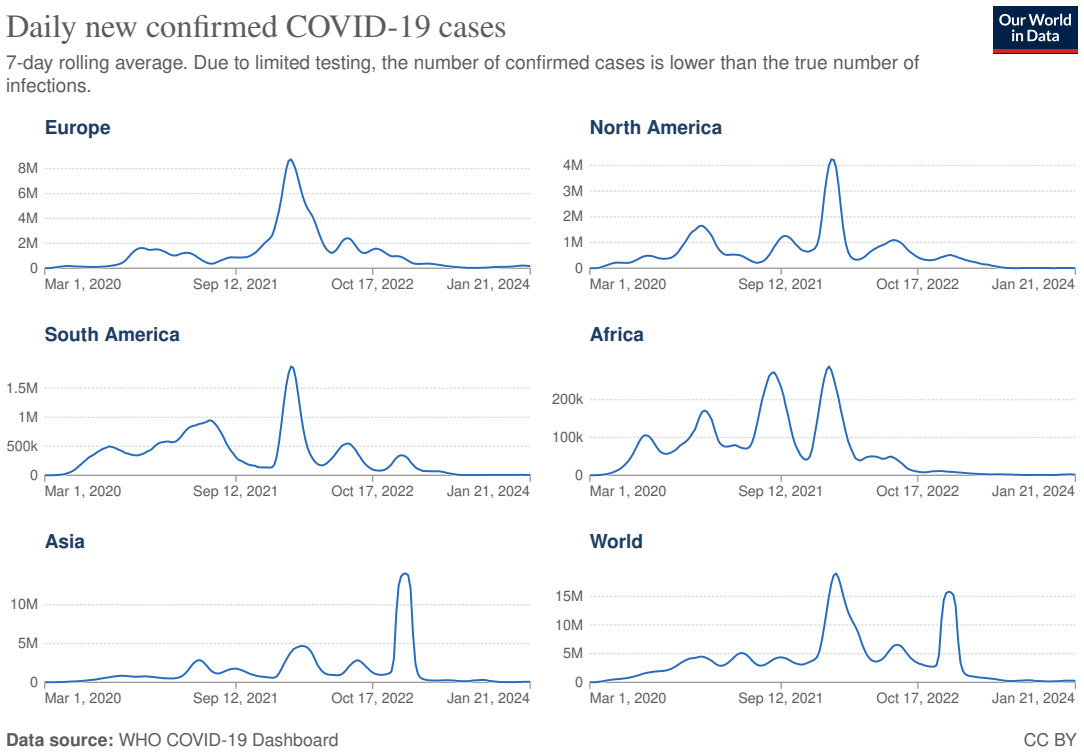


Figure 1.1: COVID-19 cases in Europe, North America, South America, Africa, Asia and worldwide. Figure obtained from `https://ourworldindata.org/covid-cases` on 30-01-2024.

1.1.1 Epidemiology of COVID-19

SARS-CoV-2 is a coronavirus (*Coronaviridae* family) [Gorbalenya et al., 2020]. Coronaviruses are common animal and human pathogens, typically causing upper or lower respiratory tract infections, depending on the virus type. Three zoonotic coronaviruses have spilt over from animal reservoirs in the last two decades: severe acute respiratory syndrome virus (SARS-CoV), Middle-East respiratory syndrome virus (MERS-

CoV) and SARS-CoV-2. All three of those infect mainly the lower respiratory tract and can lead to fatal acute respiratory distress syndrome [Gorbalenya et al., 2020, Lamers and Haagmans, 2022].

SARS-CoV-2 shares 79% sequence similarity with SARS-CoV across the entire genome. The symptoms of severe COVID-19, the disease caused by the SARS-CoV-2 virus, are similar to those of the SARS-CoV infection. Despite the phylogenetic similarity, SARS-CoV is more virulent, but its outbreak in 2003 was halted much quicker with public health interventions than the SARS-CoV-2 outbreak. This is likely due to easier identifiability of cases (majority presented symptoms) and the fact that SARS-CoV-2 is thought to replicate much quicker in the initial phases of the infection [Lamers and Haagmans, 2022].

The main transmission routes for SARS-CoV-2 are respiratory droplets and aerosols. The mean incubation period lasts about 3-5 days [Galmiche et al., 2023, Wu et al., 2022]. The lighter infection can be asymptomatic or cause common cold-like symptoms, such as cough, sore throat, congestion, fever, headache, fatigue, and diarrhoea. In severe cases, shortness of breath appears, followed by a progressive respiratory failure [Guan et al., 2020, Lamers and Haagmans, 2022]. The common risk factors for severe COVID-19 include older age, male sex and obesity, whereas the most common co-morbidities are heart failure, cardiac arrhythmia, diabetes, kidney failure, chronic pulmonary disease, and hypertension [Lamers and Haagmans, 2022].

1.1.2 Modelling of COVID-19

Mathematical modelling plays a pivotal role in epidemiology, offering invaluable insights into the transmission dynamics and potential impacts of infectious diseases' spread. Modelling is extensively utilised for endemic diseases, in order to control their prevalence, estimate the effects of treatment roll-outs, and infer the relations between the pathogen circulation and other factors, like climate or urbanisation. During the outbreaks, the modelling is an indispensable tool for understanding transmission routes, assessing intervention strategies, and predicting disease spread. It is also essential for informing timely and effective public health responses.

Mathematical modelling of the spread of COVID-19 at the beginning of the pandemic was extremely important to the policymakers deciding on restrictions aiming to curb the spread of the virus and protect the most vulnerable populations. During the early stages of the COVID-19 pandemic, numerous mathemat-

ical models emerged to grasp the complexities of the novel coronavirus's spread. These early modelling endeavours aimed to unravel the intricate transmission dynamics, assess potential outcomes, and guide public health interventions [Ferguson et al., 2020, Wu et al., 2020b]. Key parameters were meticulously examined, including the basic and effective reproduction numbers R_0 and R_t (the basic epidemiological parameters are explained later in Section 1.2) [Li et al., 2020b], serial interval [Wu et al., 2020b] or infection fatality rate [Verity et al., 2020].

These estimated parameters played a foundational role in constructing models that simulated various possible scenarios of the disease dynamics under an unmitigated epidemic or through introducing non-pharmaceutical interventions such as isolating infected individuals and their contacts, social distancing, schools and businesses closure, shielding vulnerable people or tiered lockdowns [Davies et al., 2020b, 2021b, Ferguson et al., 2020, Flaxman et al., 2020, Hellewell et al., 2020, Kucharski et al., 2020]. These simulated scenarios were aimed to help policymakers and health authorities formulate strategies to mitigate the impact of the pandemic.

Main challenges

Unlike the endemic disease with established patterns, during the outbreaks and with the emergence of novel pathogens, the modelling presents unique challenges to the infectious disease modellers. In the COVID-19 pandemic, those difficulties were mainly caused by the lack of data on the epidemiology of the novel virus and the rapidly evolving situation worldwide. More specifically, some of the main challenges included:

- Limited data: At the outset of the pandemic, there was limited data available on the novel coronavirus, including its transmission dynamics, infection fatality rate, incubation period, and other key parameters [Ferguson et al., 2020, Kucharski et al., 2020, Salje et al., 2020a, Verity et al., 2020]. An additional challenge in this area was caused by many infections being symptom-less or causing flu-like symptoms, which led to cases being unidentified and not recorded [Li et al., 2020b]. Since the testing capacity at the beginning of the pandemic was low and offered only to the most severe cases, the true numbers of infections in these early phases remained unknown. Many of the first modelling studies on SARS-CoV-2 were based on limited data collected from the Diamond Princess cruise ship [Mizumoto et al., 2020, Verity et al., 2020]. While this data was undoubtedly invaluable

at the time, it had to be used with caution and several assumptions, as the population and behaviour of passengers and staff of the cruise ship do not resemble the typical population. Moreover, as the outbreak of the disease took place in China, all the data available to researchers originated from one specific country. This lack of data made it challenging to develop accurate models.

- Uncertainty in parameters: The point made above directly led to a large uncertainty in any parameters estimated by the models. Wide, uninformative priors had to be applied, combined with the inability to quantify the true numbers of infections. This often resulted in equally wide uncertainty around the outputs of the models, such as R_0 or R_t [Gostic et al., 2020], which were of the most interest to policymakers [Davies et al., 2020a, Ferguson et al., 2020, Kucharski et al., 2020, Prem et al., 2020]. Additionally, modellers were asked about the effectiveness of hypothetical non-pharmaceutical interventions, which had never been applied on such a scale in the modern world. This caused modellers to make a lot of assumptions, which had to be thoroughly tested and often led to different answers [Baral et al., 2021, Chinazzi et al., 2020, Ferguson et al., 2020, Kissler et al., 2020, Kulldorff et al., 2020, Salje et al., 2020a].

- Effectiveness of interventions: Another challenge in estimating the effectiveness of interventions, possibly the most important issue for policymakers at the time, was the dependency of the effectiveness on a variety of factors that were impossible to predict. For example, effectiveness depended on the compliance of the people, the timing of the introduced interventions, the efficacy of face masks and other protective equipment, and a plethora of other factors, making it difficult to quantify their impact accurately [Ferguson et al., 2020, Flaxman et al., 2020, Kissler et al., 2020].

- Data quality and reporting bias: Variations in testing capacity, reporting practices, and case definitions across regions and countries introduced challenges in interpreting epidemiological data [Flaxman et al., 2020, Mellan et al., 2020, Wu et al., 2020b]. Modellers had to account for biases and uncertainties in surveillance data to develop reliable models. Additionally, due to the critical situation in many hospitals, data relating to hospitalisations and COVID-19 deaths often suffered from reporting delays.

- Heterogeneity in transmission: Transmission of the virus varied across different populations, regions, and settings, leading to challenges in modelling its spread accurately. Factors such as population density, demographics, and healthcare infrastructure influenced transmission dynamics and required consideration in models [Davies et al., 2020a, Kissler et al., 2020, Prem et al., 2020].

- Rapid spread of the virus globally: The rapid spread of the virus and its evolving dynamics presented challenges for modellers. They had to devise models capable of capturing the intricate interplay between factors such as human behaviour, public health interventions, and the virus's biology. As the pandemic progressed, the complexity of the global situation and the emergence of new variants of concern necessitated modellers to continually update and refine their models to incorporate the latest information and insights. [Faria et al., 2021, Volz et al., 2021].

In the subsequent sections of this chapter, I will introduce definitions, explanations and examples of epidemiological quantities, the basics of Bayesian statistics, and some statistical models used in epidemiology. These are concepts that the reader of this thesis should be familiar with to understand the methodologies proposed in the following chapters.

1.2 Basic epidemiological quantities

In this section, I introduce the definitions of the epidemiological quantities used in infectious disease modelling, which are used throughout this thesis.

Basic and effective reproduction numbers

One of the key parameters in infectious disease models is the **basic reproduction number R_0**. It is used to measure the transmission potential of a pathogen and is defined as an average number of secondary cases caused by one infected individual in a fully susceptible population [Anderson and May, 1991, Diekmann et al., 1990].

In practice however, the more common version of this parameters is the **effective reproduction number R_t** [Anderson and May, 1991, Gostic et al., 2020]. Similarly to R_0, R_t denotes the average number of secondary cases caused by a single individual, although the index t refers to the specific time of the epidemic. This is crucial, as the reproductive number will vary during the course of the epidemic, whereas the R_0 is considered a fixed and pathogen-specific value relating to the susceptible population. The variation in the value of R_t can be caused by numerous factors changing over time. Those include behaviour-related factors, such as differences in connectivity patterns due to e.g. migration, season or social distancing re-

strictions or increased personal hygiene awareness due to public health campaigns (washing hands, wearing face covering). Other aspects influencing the R_t are biology-related, such as increased immunity due to vaccinations or recovery from the infection, or mutations leading to variants of interest or concern.

Basic and effective reproduction numbers are central to infectious disease modelling, as their value indicates whether the epidemic will die out or grow. If $R_t < 1$, then every infected person infects on average less than 1 other person, which means that the epidemic is "under control" and is likely to die out without additional interventions. On the other hand, if $R_t > 1$, the disease is likely to spread within the population until it reaches its threshold or final size, unless interventions are executed in order to curb this spread [Diekmann et al., 1990].

Prevalence and incidence

In epidemiology, **prevalence** refers to a number or fraction of the population that is suffering from a given condition (e.g. diabetes) or is infected with a given pathogen (e.g. influenza) at a specific time or a period of time [Anderson and May, 1991]. It is used to gain insight into the current burden of the disease in the population of interest. This value can increase or decrease over time, as opposed to the **cumulative number of cases**, which denotes the total number of cases of the disease since the beginning of the outbreak or over a specified time frame.

Another key measure of the rate at which the disease or a health condition is spreading is **incidence**. Incidence denotes the number of new infections in the population of interest at a specific time [Anderson and May, 1991].

Generation time and serial intervals

The **generation time** and **serial interval** are used to describe the time between the successive generations of cases, however, their meaning and calculation are slightly different. The **generation time** is defined as the time between the infection of the primary case (infector) and the infection of the secondary case (infectee) [Wallinga and Lipsitch, 2007]. The **serial interval** denotes the time between the symptoms onset of the primary and secondary cases [Anderson and May, 1991].

For the illustration of the generation time, incubation period and serial interval, see Fig. 1.2

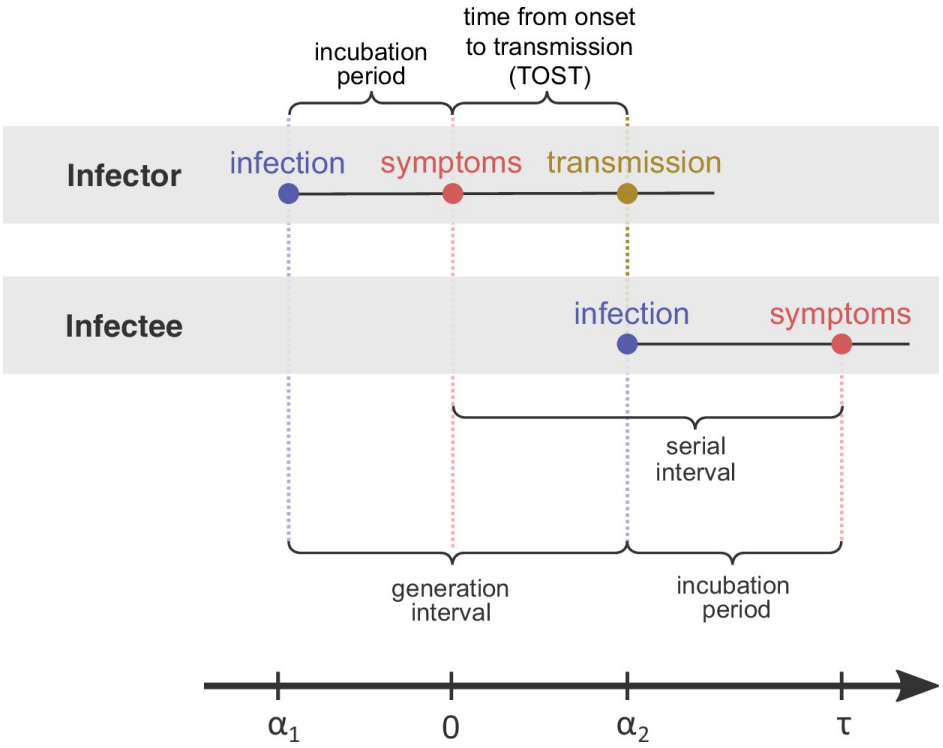


Figure 1.2: Illustration of generation time and serial interval. α_1 denotes the time of infection of the infector, α_2 is the time of infection of the infectee. Generation time is equal to $\alpha_2 - \alpha_1$, whereas the serial interval is equal to τ. Figure from Sender et al. [2022].

1.3 Bayesian modelling

The focus of this thesis is improving methods used in the Bayesian modelling of infectious diseases. Here I will introduce terminology relating to Bayesian statistics necessary to understand the methods developed in the remaining chapters. For a more detailed introduction to the Bayesian modelling and analysis, the reader is referred to e.g. Gelman et al. [2013], Lambert [2018], McElreath [2016], Rasmussen and Williams [2006].

1.3.1 Bayesian inference

The Bayesian inference is based on the following equation:

$$p(\theta|y) \propto p(\theta)p(y|\theta) \tag{1.1}$$

where y denotes the observed data and θ is a set of the model's parameters. Here, $p(\theta)$ is the model's **prior probability**, $p(y|\theta)$ is the **likelihood** and $p(\theta|y)$ denotes the **posterior**. The symbol $\propto$ means that the posterior is proportional to the product of the prior and likelihood.

The above is derived from the *Bayes theorem*, where additionally we have the normalising denominator:

$$P(\theta|y) = \frac{P(\theta)P(y|\theta)}{P(y)} \tag{1.2}$$

Here, the $\propto$ sign can be replaced with $=$, as the denominator is present on the right-hand side. This denominator is a constant value, also called a **normalising constant** (or **marginal likelihood** or **model evidence**). While not necessary for the inference itself, a normalising constant can be used for model selection (as explained later in Section 1.4.2).

1.3.2 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) provides a versatile approach to addressing the challenging tasks of sampling from multi-dimensional distributions for obtaining numeric results. It is particularly valuable in Bayesian statistics for approximating complex posterior distributions. MCMC methods allow researchers to draw samples from a distribution when direct sampling is impractical or impossible. These obtained samples can be directly employed for tasks such as parameter inference and prediction.

There exist many sampling methods under the umbrella of MCMC. The simpler ones, such as the random walk Metropolis algorithm or Gibbs sampling may require a very long time to converge to the target distribution. Hamiltonian Monte Carlo (HMC) represents an MCMC technique designed to overcome the challenges of many common MCMC approaches, such as random walk patterns or sensitivity to the correlation of the parameters. HMC uses the first-order gradient to inform a sequence of steps, allowing HMC to achieve faster convergence, even for high-dimensional distributions. Nowadays, the most up-to-date Probabilistic Programming Languages (Section 1.3.4) come with the state-of-the-art No-U-Turn-Sampler (NUTS) [Hoffman and Gelman, 2014], which allows the users to perform HMC without specifying the step size or the number of steps.

1.3.3 Model diagnostics

To assess the performance of the sampling algorithms, identify potential issues with convergence, sampling efficiency or mixing, as well as verify the reliability of the posterior distributions, diagnostic tools for MCMC can be used. Some of the most common tools are explained below.

Trace plots display the values of the MCMC sampled model's parameters (y-axis) over iterations (x-axis) in each of the MCMC chains. Examining the plots, we can discover whether the algorithm has converged to a target distribution. Additionally, trace plots can suggest whether the algorithm explored sufficient parameter space (i.e. good mixing).

$\hat{R}$ or **Gelman-Rubin statistics**, compares the within-chain and between-chain variability to assess convergence [Gelman and Rubin, 1992, Vehtari et al., 2021]. A desirable value of $\hat{R}$ is below 1.01.

Effective Sample Size (ESS) quantifies the number of independent samples in the MCMC output. It accounts for auto-correlation and provides a more accurate estimate of the actual sample size for inference.

Density plots of the model's parameters posteriors show the sampled values of the parameters in the form of histograms or density plots. Examining them allows us to identify multi-modality or skewness in the distributions, as well as assess the uncertainty and reliability of the range of sample values.

The diagnostic tools mentioned above pertain mostly to the sampling algorithms. Additionally, to assess the correctness of the model choice and its fit to the data, practitioners often carry out prior- and posterior-predictive checks.

Prior predictive checks are used to assess if the chosen priors are reasonable and whether the data could be produced by the model without incorporating any information coming from the likelihood. To perform a prior predictive check, one must sample values from the priors of the model without using the observed data. These sample values can be then compared to the observed data to gauge insights into the correctness of the selected priors.

Posterior predictive checks are similar, however, instead of sampling from the prior distributions, the values are sampled using the fitted model, which is based on the likelihood and the observed data. These samples can be then compared to the observed data to assess if the two sets align, and if the patterns of the observed dataset were reproduced.

For details about the diagnostic checks and the *Bayesian workflow*, I refer the reader to the comprehensive paper on this topic by Gelman et al. 2020.

1.3.4 Probabilistic Programming Languages

The increasing interest in probabilistic (Bayesian) modelling in a variety of fields gave rise to the development of **Probabilistic Programming Languages** (PPLs). They are a category of programming languages and libraries designed to facilitate probabilistic modelling and inference. PPLs provide a framework for expressing Bayesian models and performing inference in an automated way. Their outputs allow the practitioners to reason about uncertainty, make predictions, and diagnose the fitted models.

The first language considered a PPL is BUGS (Bayesian inference Using Gibbs Sampling) developed by Gilks and Spiegelhalter in the early 1990s [Gilks et al., 1994]. Since then, many new PPLs have been invented. Amongst the most popular ones, we can list Stan [Carpenter et al., 2017], PyMC3 [Salvatier et al., 2016], Edward [Tran et al., 2016], Turing.jl [Ge et al., 2018] or Pyro [Bingham et al., 2018] (and its alternative backend NumPyro [Phan et al., 2019]).

In this thesis, the code for the Bayesian models used for analyses in chapters 2, 3 and 4 are written using Stan (with PyStan interface [Stan Development Team]), code for Chapter 5 and part of 2 is written in NumPyro.

1.3.5 Kullback-Leibler divergence

Kullback-Leibler (KL) divergence also known as **relative entropy**, measures the distance between two probability distribution [Kullback and Leibler, 1951]. It has wide applications in information theory and can quantify e.g. how much information is gained when comparing statistical models. Mathematically, the KL divergence between two probability densities p and q is defined as:

$$D_{\mathrm{KL}}(p\|q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \tag{1.3}$$

1.4 Model selection

Model selection is an important part of the Bayesian workflow [Ding et al., 2018, Gelman et al., 2020]. As all models are merely an approximation of the world around us, we can rarely be certain about the mechanics of the phenomena we are trying to model. Therefore, an advisable step of any Bayesian analysis is modifying the model of interest, e.g. by changing the priors or likelihood functions, in order to find the model that explains the observed data, generalises to the new unseen data or provides reliable predictions. The definition of "the best" model is not straightforward – good models should be accurate while not overfitting the data, they should be robust and stable, interpretable and not overly complex. Because of these many conditions on which the models should be compared to each other, the task of model selection is not an easy one. Many metrics and mechanisms have been developed to aid the model selection process. Those include for example DIC, BIC, cross-validation, proper scoring, Bayes factors and many more [Ding et al., 2018]. Chapter 2 pertains specifically to the model selection process, and more details about this can be found there. Additionally, some of the most common methods of model selection are explained below.

1.4.1 Information Criteria

AIC — Akaike Information Criterion is an estimator of prediction error [Akaike, 1974]. Let us denote the number of parameters of the model with k and the maximised value of the likelihood function with $\hat{L}$, then the AIC value is defined as:

$$AIC = 2k - 2\ln(\hat{L}) \tag{1.4}$$

The idea behind the AIC is that it balances the goodness of fit (through the log-likelihood) with the complexity of the model (penalised by the number of parameters). When comparing models, the one minimalising the AIC value is the most favourable.

BIC — Bayesian Information Criterion is similar to AIC, but penalises the free parameters more strongly [Schwarz, 1978].

$$BIC = k\ln(n) - 2\ln(\hat{L}) \tag{1.5}$$

Importantly, BIC is not intended to predict out-of-sample model performance and there are arguments against calling it a 'Bayesian' method [Gelman et al., 2014].

DIC — Deviance Information Criterion is a generalisation of AIC more widely used in Bayesian modelling [Spiegelhalter et al., 2002]. If we define the deviance $D(\theta)$ as

$$D(\theta) = -2\log(p(y|\theta)) + C \tag{1.6}$$

where y is the data, θ are the parameters of the model, $p(y|\theta)$ is the likelihood, and C is a constant that cancels out so can be neglected, then we can define DIC as

$$DIC = D(\bar{\theta}) + 2p_D \tag{1.7}$$

In the equation above, $\bar{\theta}$ is an expectation of θ and $2p_D = D\bar{(\theta)} - D(\bar{\theta})$ refers to an effective number of parameters. The main difference between the AIC or BIC and DIC is that for DIC we no longer require the maximum likelihood, but we can use the readily available MCMC samples instead to compute the $\bar{\theta}$ [Ding et al., 2018, Spiegelhalter et al., 2002].

WAIC — Widely Applicable Information Criterion (or Watanabe-Akaike information criterion) is closely related to DIC, and particularly popular in the context of Bayesian model selection [Watanabe, 2010]. As opposed to AIC or BIC, WAIC has the property of averaging over the posterior distribution rather than conditioning on a point estimate [Gelman et al., 2014, Vehtari et al., 2017].

1.4.2 Marginal Likelihood and Bayes factors

Marginal Likelihood [Gelman et al., 2013], also known as **model evidence** or **normalising constant**, is enclosed in the denominator part of the Bayes theorem:

$$p(\theta|y, m) = \frac{q(\theta|y, m)}{z(y|m)} \tag{1.8}$$

where the denominator (model evidence) is equal to:

$$z(y|m) = \int q(\theta|y, m)d\theta \tag{1.9}$$

Here we denote y – data, m – model, θ - vector of the model's parameters. $p(\theta)$ is the posterior distribution, and $q(\theta)$ the un-normalised density. For clarity, we will drop the indexing on y and m as it is implied.

Bayes factors (BFs) are a statistical measure used for quantifying the evidence for one hypothesis (or model) over another [Kass and Raftery, 1995]. Bayes factors can be calculated as the ratio of two marginal likelihoods z_1 and z_2 of two competing models M_1 and M_2:

$$BF_{21} = \frac{z_2}{z_1} \tag{1.10}$$

A common interpretation of BFs has been proposed by Kass and Raftery [1995] and is presented in Table 1.1. For simple models, such as those including conjugate priors or those with a low number of parameters, marginalised likelihood z can be obtained analytically or using numerical quadrature methods, e.g. trapezoidal rule. Similarly, for discrete parameters, the denominator of the Bayes rule turns into a computable sum. In general, however, calculating the marginal likelihoods is very difficult or impossible for complex models [Gelman and Meng, 1998], e.g. dynamic, spatial or Bayesian hierarchical models that often involve approximating very high-dimensional integrals. The general critique against Bayes factors is their sensitivity to the priors, however, some scientists argue that is a desired feature of a Bayesian method for model comparison.

BF_{21}	Evidence against M_1
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
>100	Decisive

Table 1.1: Interpretation of the Bayes factors values proposed in Kass and Raftery [1995].

1.5 Statistical models used in infectious disease modelling

In this section, I will introduce some basic statistical models used in the modelling of infectious disease dynamics. Historically, infectious disease epidemics have been modelled using a mathematical SIR model introduced in Kermack and McKendrick [1927]. The basis of the SIR model and its variations is a system of differential equations describing the dynamics of individuals between tree compartments: S - susceptibles, I - infected, R - recovered or removed. The original, basic form of this system of equations is given below, with parameters β denoting the transmission or contact rate and γ denoting the rate of recovery:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta \frac{S(t)I(t)}{N} \\
\frac{dI}{dt} &= \beta \frac{S(t)I(t)}{N} - \gamma I(t) \\
\frac{dR}{dt} &= \gamma I(t)
\end{aligned} \tag{1.11}$$

Here I will introduce some basic statistical models which are currently used in infectious disease modelling, and which I utilise in the further chapters of my thesis.

1.5.1 Random Walk

Random walk is a mathematical random process describing a path consisting of a number of steps defined by random factors in space [Codling et al., 2008]. Considering $X_1, X_2, ...$ such that $X_n \in \mathbb{R}^d$ and X_n are independent and identically distributed (iid) random variables, we can define a random walk as a sequence $(S_n)_{n \geq 0}$, such that:

$$\begin{aligned}
S_0 &= z, \\
S_n &= S_{n-1} + X_n, \ n \geq 1
\end{aligned} \tag{1.12}$$

where $z \in \mathbb{R}^d$ is the starting point, and X_n are the steps.

A random walk in which steps are normally-distributed, that is $X_n \sim \mathcal{N}(\mu, \sigma)$ is called a **Gaussian random walk**.

1.5.2 Autoregressive Model

An **autoregressive model** is a random process in which every successive step depends linearly on its previous values [Besag, 1974]. An $AR(p)$ denotes an autoregressive process of order p and is defined as

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t \tag{1.13}$$

where ϕ_i are the parameters of the model.

1.5.3 Gaussian Processes

Gaussian Process (GP) defines a collection of random variables, such that any finite set of those random variables has a multivariate normal distribution [Rasmussen and Williams, 2006]. It is a non-parametric stochastic process widely used in statistical modelling.

The defining properties of a Gaussian Process are *mean function* $m(x)$ and *covariance function* also called a *kernel* $k(x, x')$. For a real process $f(x)$ these can be defined as:

$$\begin{aligned} m(x) &= \mathbb{E}[f(x)], \\ k(x, x') &= \mathbb{E}[(f(x) - m(x))(f(x') - m(x)], \end{aligned} \tag{1.14}$$

and the GP will be then written as

$$f(x) \sim \mathrm{GP}(m(x), k(x, x')) \tag{1.15}$$

Most kernels include two parameters: *lengthscale* l and *variance* σ^2. The lengthscale controls the length over which the GP output is expected to vary — a shorter lengthscale will produce results that are more wiggly and rapidly changing, capturing short-range variations, whereas a longer lengthscale causes smoother, slower-changing functions, capturing long-range variation. The variance controls how much the functions can deviate from their mean.

Some of the most common kernels in Gaussian Processes are:

- linear $k_{\mathrm{Lin}}(x, x') = \sigma_b^2 + \sigma_v^2(x - c)(x' - c)$, where c is a constant,
- squared-exponential a.k.a. radial basis function (RBF): $k_{\mathrm{SE}}(x, x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\ell^2}\right)$,
- periodic $k_{\mathrm{Per}}(x, x') = \sigma^2 \exp\left(-2\sin^2\left(\frac{\pi|x-x'|}{p}\right)\ell^2\right)$, where p is the length of one complete cycle,
- Matérn $k_{\mathrm{Mat}}(x, x'; \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{|x-x'|}{\ell}\right)^{\nu} K_\nu\left(\sqrt{2\nu}\frac{|x-x'|}{\ell}\right)$, where K_ν is a modified Bessel function of the second kind.

The examples of prior draws from the GPs with these kernels are given in Fig. 1.3.

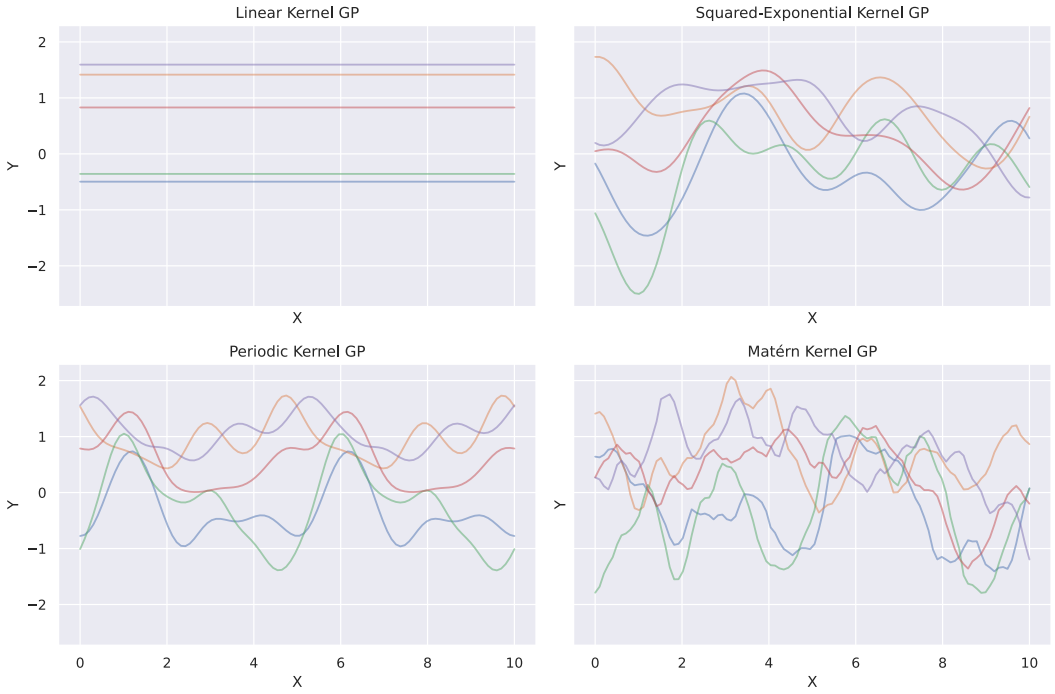


Figure 1.3: Examples of the prior draws from zero-mean GPs with linear, squared-exponential, periodic and Matérn kernels.

Many more kernel functions are available. It is also possible to combine multiple kernels by addition or multiplication (or a combination of both), or other methods preserving the property that a covariance matrix of the process is positive semi-definite.

Gaussian Processes are a powerful and flexible statistical tool which can aid in tackling problems such as regression analysis, forecasting, classification, clustering and many others. Since they are a probabilistic method, they allow incorporating uncertainty into the predictions and estimates. One of the main drawbacks of this method is that they typically are computationally expensive and non-scalable to large datasets. Many researchers have tackled this problem and proposed a range of improvements, from solving the computational issues to using approximate methods [Flaxman et al., 2015, Liu et al., 2020, Riutort-Mayol et al., 2022, Titsias, 2009, Wang et al., 2019].

1.6 Renewal Equation

Renewal equation is used ubiquitously in modelling infectious disease incidence [Cori et al., 2013, Fraser, 2007a, Pakkanen et al., 2023]. It provides a framework for describing the rate of change of new infections over time, taking into account the interactions between susceptible and infectious individuals.

This is typically modelled assuming that the mean incidence $I(t)$ at time t follows the renewal equation:

$$I(t) = \int_0^\infty \beta(t, \tau) I(t - \tau) d\tau \tag{1.16}$$

Here, the term $\beta(t, \tau)$ denotes the transmissibility of the pathogen at the calendar time t and time since infections τ, reflecting its viral load and the contact rates between the infectious and susceptible individuals [Fraser, 2007b].

Another formulation of the renewal equation used for modelling incidence is given in Pakkanen et al. [2023]:

$$I(t) = R_0 \int_0^\infty I(t - \tau) g(\tau) d\tau \tag{1.17}$$

in which $g(.)$ is the probability density function (PDF) of the generation interval.

The information provided in sections 1.2 - 1.6 was intended to give the theoretical and technical foundation for the projects presented in the remaining chapters. In the final section of the Introduction, I will summarise the aims of this thesis.

1.7 Thesis aims

The outbreak of the COVID-19 epidemic in China in December 2019 and its global spread in the following months, highlighted both the need for statistical modelling of infectious diseases and the weaknesses and limitations those models suffer from. Typically, the use of modelling in the field of infectious disease is carried out over extended periods of time, allowing to build on years of research into the epidemiology of the pathogen, with access to large and varied datasets. In the case of an outbreak of a new, rapidly spreading pathogen, for which no vaccine or effective cure is readily available, the modelling has to be executed fast, with very limited and often lagged data, as discussed earlier in Section 1.1.2. The challenges

faced by the epidemiological modellers in the beginning and during the COVID-19 pandemic emphasised the importance of collaborative work between scientists, public health authorities and even the public. The outbreak of COVID-19 also highlighted the many gaps in the current approaches to epidemic modelling.

Motivated by these issues posed by the statistical modelling of infectious diseases, the overarching aim of this thesis is the development of methods, which can aid epidemiologists in building more accurate and trustworthy models of emerging pathogens. The proposed methods are applied in the modelling of the SARS-CoV-2 epidemiology, as this pathogen emerged and was spreading worldwide during the work presented here was undertaken.

In Chapter 2, I explore a new method for estimating the marginal likelihoods using thermodynamic integration. This allows the use of Bayes factors to perform the model selection. First I explain how to use this approach on a few pedagogical examples, and finally, I employ it to present the pitfalls of model selection on a COVID-19 dataset.

Using this newly developed framework, in Chapter 3, I fit a range of probability density functions to hospitalisation distributions for COVID-19 patients in the Brazilian healthcare system and select the best-fitting ones. Those distributions, such as symptoms-onset-to-death-time, were needed as an integral input in the COVID-19 epidemic models, and their correctness was necessary for the appropriate surveillance and policy making. I obtain the mathematical densities on the national- and state-level and compare the outcomes, exploring the spatial heterogeneity of the hospitalisation times across Brazil.

In Chapter 4, I change the focus to the correctness of the data itself. Namely, I investigate the reporting delays in the COVID-19 mortality database for Brazil and propose a new method of correcting them using Gaussian Processes. The main aim of this work is to obtain a realistic real-time overview of the epidemiological situation, rather than relying on the data suffering from under-ascertainment and delays.

Building on this work on appropriate model building and selections, in Chapter 5 I investigate the identifiability in a renewal-equation-based R_t model. Using simulated datasets, I show examples of how regularisation through an informative prior can lead the model towards correct inference of the importations in the emerging epidemic, and how a wider prior can result in unrealistic estimates of the parameters, still providing a good fit to the observed data.

Finally, in Chapter 6 I summarise the findings of my PhD thesis and discuss their potential impact on

infectious disease epidemic modelling. I also talk about the limitations and possible further directions of this work.

Chapter 2

Application of referenced thermodynamic integration to Bayesian model selection

Evaluating normalising constants is important across a range of topics in statistical learning, notably Bayesian model selection. However, in many realistic problems, this involves the integration of analytically intractable, high-dimensional distributions, and therefore requires the use of stochastic methods such as thermodynamic integration (TI). In this chapter we apply a simple but under-appreciated variation of the TI method, here referred to as *referenced TI*, which computes a single model's normalising constant efficiently by using a judiciously chosen reference density. The advantages of the approach and theoretical considerations are set out, along with pedagogical 1 and 2D examples. The approach is shown to be useful in practice when applied to a real problem — to perform model selection for a semi-mechanistic hierarchical Bayesian model of COVID-19 transmission in South Korea involving the integration of a 200D density.

2.1 Introduction

Mathematical modelling of infectious diseases is widely used to understand the processes underlying pathogen transmission and inform public health policies. With advances in both computing power and availability of data, it is possible to build more complex, robust and accurate models. The increasing importance of epidemiological models requires synthesis with rigorous statistical methods. This synthesis is required to estimate necessary parameters, quantify uncertainty in predictions, and test hypotheses [Grassly and Fraser, 2008].

Fundamental to creating reliable mathematical models in epidemiology was the popularisation of Markov Chain Monte Carlo (MCMC) methods, which enable fitting complex models to datasets of various sizes. MCMC allows us to sample from complicated probability distributions, without having to know the normalising constant [Gelman and Meng, 1998]. However, to fit an epidemiological model to the data, first, we need to come up with a model. For example, to fit a multivariate linear regression, we need to choose the covariates and pick prior probability densities for every parameter in the regression equation. To fit a generalised linear model, we need to choose a link function. For a simple epidemiological example, consider fitting a probability density function to a distribution of onset-to-death time (see Chapter 3). We have a choice of multiple probability densities, such as gamma or log-normal, and additionally, the number of parameters depends on the hierarchical structure of the model.

In practice the decision to choose a model is often based on heuristics, relying on the knowledge and experience of the modeller, rather than through a systematic selection process [Stocks et al., 2018, Sun et al., 2015]. Several model selection methods are available, but those methods often come with a trade-off between accuracy and computational complexity. For example, widely used in epidemiology Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) (see 1.4.1) are easy to compute, but come with certain limitations [Grassly and Fraser, 2008]. Specifically, they do not take into account the parameters' uncertainty or the prior probabilities and might favour excessively complex models.

The marginal likelihood (see 1.4.2), or normalising constant of a model, is a feature central to the principles and pathology of Bayesian statistics. For example — given two models representing two competing hypotheses, the ratio of the normalising constants (known as the Bayes factor), describes the relative probability of the data having been generated by one hypothesis compared to the other. Consequently, at

a practical level, the estimation of normalising constants is an important topic for model selection in the Bayesian setting [Kass and Raftery, 1995].

In practice, estimating a normalising constant relies on "integrating out" or marginalising the parameters of the model to get the probability the associated hypothesis produced the data. But in general, this is difficult, because we cannot easily integrate arbitrary high-dimensional distributions — certainly analytical or quadrature-based methods are of little help directly. The widely used Laplace method allows an analytic approximation to the normalising constant, by approximating the integral of interest by an n-dimensional Gaussian integral [Tierney and Kadane, 1986]. The properties of Gaussian densities greatly facilitate mathematical analysis.

In general, practitioners turn to a range of more accurate approaches, typically based on statistical sampling. Specific examples include bridge sampling [Bennett, 1976, Meng and Wong, 1996a,b], stochastic density of states based methods [Habeck, 2012] nested sampling [Skilling, 2006], annealed importance sampling [Neal, 2001], integrated nested Laplace approximation [Rue et al., 2009], power posteriors [Friel and Pettitt, 2008, Lartillot and Philippe, 2006], steppingstone sampling [Xie et al., 2011], thermodynamic integration [Friel et al., 2017, Gelman and Meng, 1998, Kirkwood, 1935, Lartillot and Philippe, 2006] and others. For a comprehensive review of these methods, we refer to Friel and Wyse [2012], Gelman and Meng [1998], Llorente et al. [2023].

This work focuses on the latter — thermodynamic integration — in particular on the development of the practical details for efficient application for Bayesian model selection.

2.1.1 Thermodynamic Integration

By way of introduction for researchers unfamiliar with thermodynamic integration (TI), TI allows us to estimate the ratio of two normalising constants in a general and asymptotically exact way. Instead of marginalising the associated densities explicitly in terms of the high-dimensional integrals, using TI we only have to evaluate a 1-dimensional integral, where the integrand can easily be sampled with MCMC. To see how this works, consider two models labelled 1 and 2 with normalising constants z_1 and z_2. Each z is given by

$$z_i = \int q_i(\boldsymbol{\theta}) d\boldsymbol{\theta}\,, \;\; i \in \{1, 2\}\,, \tag{2.1}$$

where q_i is an un-normalised density for model M_i with parameters $\boldsymbol{\theta}$, that gives the model's Bayesian posterior density as

$$p_i(\boldsymbol{\theta}) = \frac{q_i(\boldsymbol{\theta})}{z_i}\,,\ i \in \{1, 2\}\,. \tag{2.2}$$

To apply thermodynamic integration we introduce the concept of a path between $q_1(\boldsymbol{\theta})$ and $q_2(\boldsymbol{\theta})$, linking the two densities via a series of intermediate ones. This family of densities, parameterised by the path coordinate λ, is denoted by $q(\lambda; \boldsymbol{\theta})$. An example path in λ is shown in Fig. 2.1.

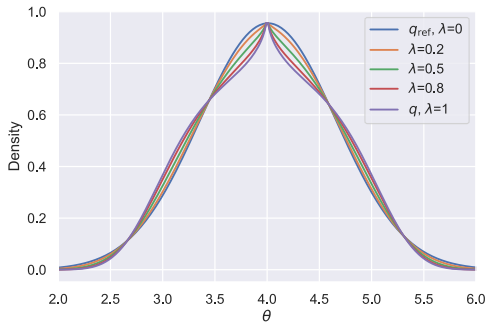


Figure 2.1: Illustration of $q^{\lambda} q_{\text{ref}}^{(1-\lambda)}$ for the 1d example density in parameter θ at selected λ values along the path. The example comes from the pedagogical exercise given later in Section 2.3.1, where more details are given.

The density $q(\lambda; \boldsymbol{\theta})$, linking q_1 to q_2 and defining the intermediate densities, can be chosen to have an optimal or in some way convenient path. A common choice based on convenience is the geometric one

$$q(\lambda; \boldsymbol{\theta}) = q_2^{\lambda}(\boldsymbol{\theta}) q_1^{1-\lambda}(\boldsymbol{\theta})\,,\ \lambda \in [0, 1]\,. \tag{2.3}$$

The important point to note is that for $\lambda = 0$, $q(\lambda; \boldsymbol{\theta})$ returns the first density $q(0; \boldsymbol{\theta}) = q_1(\boldsymbol{\theta})$, for $\lambda = 1$ it gives $q(1; \boldsymbol{\theta}) = q_2(\boldsymbol{\theta})$, and for in-between λ values a log-linear mixture of the endpoint densities. Just as we have defined a family of densities, there is an associated normalising constant for any point along the path, that for any value of λ is given by

$$z(\lambda) = \int_{\Omega(\lambda)} q(\lambda; \boldsymbol{\theta}) d\boldsymbol{\theta}\,. \tag{2.4}$$

A further small but important point to avoid complications is to have densities with common support, for example, $\Omega(\lambda = 1) = \Omega(\lambda = 0)$. Hereafter support is denoted by Ω.

Having set up the definitions of $q(\lambda;\boldsymbol{\theta})$ and $z(\lambda)$, the TI expression can be derived, to compute the log-ratio of $z_1 = z(\lambda = 0)$ and $z_2 = z(\lambda = 1)$, while avoiding explicit integrals over the models' parameters $\boldsymbol{\theta}$. This is laid out as follows:

$$\begin{aligned}
\log \frac{z_2}{z_1} &= \int_0^1 \partial_\lambda \log z(\lambda)\, d\lambda \\
&= \int_0^1 \frac{1}{z(\lambda)} \partial_\lambda z(\lambda)\, d\lambda \\
&= \int_0^1 \frac{1}{z(\lambda)} \partial_\lambda \int_\Omega q(\lambda;\boldsymbol{\theta})\, d\boldsymbol{\theta}\, d\lambda \\
&= \int_0^1 \frac{1}{z(\lambda)} \int_\Omega \left(\partial_\lambda \log q(\lambda;\boldsymbol{\theta})\right) q(\lambda;\boldsymbol{\theta})\, d\boldsymbol{\theta}\, d\lambda \\
&= \int_0^1 \mathbb{E}_{p(\lambda;\boldsymbol{\theta})} \left[\partial_\lambda \log q(\lambda;\boldsymbol{\theta})\right]\, d\lambda \\
&= \int_0^1 \mathbb{E}_{p(\lambda;\boldsymbol{\theta})} \left[\log \frac{q_2(\boldsymbol{\theta})}{q_1(\boldsymbol{\theta})}\right]\, d\lambda \\
&= \int_0^1 \mathbb{E}_{q(\lambda;\boldsymbol{\theta})} \left[\log \frac{q_2(\boldsymbol{\theta})}{q_1(\boldsymbol{\theta})}\right]\, d\lambda\,, \qquad (2.5)
\end{aligned}$$

Here we started with the fundamental theorem of calculus (first step), rules of differentiating logs (second step), definition of $z(\lambda)$ (third step), assumed exchangeability of ∂_λ and $\int d\boldsymbol{\theta}$ and log differentiation rules again (fourth step), identifying the expectation $\mathbb{E}_{q(\lambda;\theta)}$ from sampling distribution $q(\lambda;\theta)$ (fifth step), differentiation of the geometric path for $q(\lambda)$ (sixth step), and finally equivalence of sampling from q and p. The final line in the expression summarises the usefulness of TI. Instead of having to work with complicated high-dimensional integrals of Eq. 2.1 to find the log-Bayes factor $\log \frac{z_2}{z_1}$, which measures the relative probability of getting the data from one hypothesis compared to another, we only need to consider a 1-dimensional integral of an expectation, and that expectation can be readily produced by MCMC.

2.1.2 Our contributions

In this chapter, we examine the details of a *referenced TI* approach, which is a variation on the TI theme that we find useful to enable fast and accurate normalising constant calculations. Our main contributions are as follows:

- We show how to generate a reference normalising constant from an exactly-integratable reference density, through sampling or gradients, and with parameter constraints. We present how to use

this reference in the TI method to efficiently estimate a normalising constant of an arbitrary high-dimensional density.

- We discuss performance benchmarks for a well-known problem in the statistical literature Williams, 1959, which shows the method performs favourably in terms of accuracy and the number of iterations to convergence.

- Finally the technique is applied to a hierarchical Bayesian time-series model describing the COVID-19 epidemic in South Korea.

In relation to other work, we recognise using a reference for thermodynamic integration is a topic that has been raised, especially in early theoretically-oriented literature Diciccio et al., 1997, Grosse et al., 2013, Neal, 1993. Our additional contribution is to bridge the gap from theory and simple examples to application, which includes choosing the reference using MCMC samples or gradients, examination of reference support, comparisons of convergence, and illustration of the approach for a non-trivial real-world problem.

The remainder of this chapter is set out as follows: in Section 2.2 we provide a comprehensive explanation of the referenced thermodynamic integration and go over some practical recommendations. In Section 2.3, we present a number of pedagogical examples, followed by a more complex and realistic example of applying the methodology to COVID-19 modelling. Finally in Section 2.4, we summarise the findings and discuss the limitations of the proposed method.

2.2 Methods

In this section, we explain in detail the referenced thermodynamic integration and suggest some practical approaches for generating the reference density.

2.2.1 Referenced TI

Introducing a reference density and associated normalising constant as

$$\begin{aligned} z &= z_{\text{ref}} \frac{z}{z_{\text{ref}}} \\ &= z_{\text{ref}} \exp \int_0^1 \mathbb{E}_{q(\lambda;\boldsymbol{\theta})} \left[\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})} \right] d\lambda \,, \end{aligned} \tag{2.6}$$

yields an efficient approach to compute Bayes factors, or more generally to marginalise an arbitrary density for any application. To clarify notation, here z is the normalising constant of interest with density q, z_{ref} is a reference normalising constant with associated density q_{ref}. In the second line, the ratio z/z_{ref} is straightforwardly given by the thermodynamic integral identity in Eq. 2.5.

While the Eq. 2.5 can be directly applied to conduct a pairwise model comparison between two hypotheses, by introducing a reference we can naturally marginalise the density of a single model [Diciccio et al., 1997, Neal, 1993]. This is useful when comparing multiple models, as $n > \binom{n}{2}$ for $n > 3$. Another motivation to reference the TI is the MCMC computational efficiency of converging the TI expectation. In Eq. 2.6, with judicious choice of q_{ref}, the reference normalising constant z_{ref} can be evaluated analytically and account for most of z. In this case $\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})}$ tends to have a small expectation and variance and converges quickly.

This idea of using an exactly solvable reference, to aid in the solution of an otherwise intractable problem, has been a recurrent theme in the computational and mathematical sciences in general [Dirac and Bohr, 1927, Duff et al., 2015, Gell-Mann and Brueckner, 1957, Vočadlo and Alfè, 2002], and variations on this approach have been used to compute normalising constants in various guises in the statistical literature [Baele et al., 2016, Cameron and Pettitt, 2014, Fan et al., 2011, Friel and Pettitt, 2008, Friel and Wyse, 2012, Friel et al., 2017, Lartillot and Philippe, 2006, Lefebvre et al., 2010, Neal, 1993, Xie et al., 2011]. For example, in the generalised stepping stone method, a reference is introduced to speed up the convergence of the importance sampling at each temperature rung [Baele et al., 2016, Fan et al., 2011]. In the work of Lefebvre et al. [2010], a theoretical discussion has been presented that shows the error budget of thermodynamic integration depends on the J-divergence of the densities being marginalised. Noting this, Cameron and Pettitt [2014] illustrates a 2-dimensional example in their work on recursive pathways to marginal likelihood estimation. And in the power posteriors method, a reference is used but the reference is a prior density and thus $z_{\text{ref}} = 1$ [Friel and Pettitt, 2008]. This approach is elegant as the reference need

not be chosen — it is simply the prior — however, the downside is that for poorly chosen or uninformative priors, the thermodynamic integral will be slow to converge and susceptible to instability. In particular for complex hierarchical models with weakly informative priors, this is found to be an issue.

For referenced TI as presented here, the reference density in Eq. 2.6 can be chosen at convenience, but the main desirable features are that it should be easily formed without special consideration or adjustments and that z_{ref} should be analytically integratable and account for as much of z as possible. Such a choice of z_{ref} ensures the part with expensive sampling is small and converges quickly. An obvious choice in this regard is the Laplace-type reference, where the log-density is approximated with a second-order one, for example, a multivariate Gaussian. For densities with a single concentration, Laplace-type approximations are ubiquitous, and an excellent natural choice for many problems. In the following section, we consider approaches that can be used to formulate a reference normalising constant z_{ref} from a second-order log-density (though more generally other tractable references are possible). In each referenced TI scenario, we note that even if the reference approximation is poor, the estimate of the normalising constant based on Eq. 2.6 remains asymptotically exact—only the speed of convergence is affected (subject to the assumptions of matching support for end-point densities).

2.2.2 Taylor Expansion at the Mode Laplace Reference

The most straightforward way to generate a reference density is to Taylor expand the log-density to second order about a mode. Noting no linear term is present, we see the reference density is

$$q_{\text{ref}}(\boldsymbol{\theta}) \quad = \quad \exp\left(\log q(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^{\mathrm{T}}\mathbf{H}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right), \tag{2.7}$$

where $\mathbf{H}$ is the Hessian matrix and $\boldsymbol{\theta}_0$ is the vector of mode parameters. The associated normalising constant is

$$\begin{aligned} z_{\text{ref}} &= \int_{-\infty}^{\infty} q_{\text{ref}}(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \int_{-\infty}^{\infty} \exp\left(\log q(\boldsymbol{\theta}_0) + \frac{1}{2}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^{\mathrm{T}}\,\mathbf{H}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right) d\boldsymbol{\theta} \\ &= \mathrm{q}(\boldsymbol{\theta}_0)\int_{-\infty}^{\infty} \exp\left(\frac{1}{2}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^{\mathrm{T}}\,\mathbf{H}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right) d\boldsymbol{\theta} \\ &= q(\boldsymbol{\theta}_0)\sqrt{\det(2\pi\mathbf{H}^{-1})}\,. \end{aligned} \tag{2.8}$$

This approach to yield a reference density, using either analytic or finite difference gradients at mode, tends to produce a density close to the true one in the neighbourhood of $\boldsymbol{\theta}_0$. But this is far from guaranteed, particularly if the density is asymmetric, has non-negligible high-order moments, or is discontinuous for example exhibiting cusps. In many instances, a more reliable choice of reference can be found by using MCMC samples from the whole posterior density.

2.2.3 Sampled Covariance Laplace Reference

A second straightforward approach to forming a reference density, which is often more robust, is by drawing samples from the true density $q(\boldsymbol{\theta})$ to estimate the mean parameters $\hat{\boldsymbol{\theta}}$ and covariance matrix $\hat{\boldsymbol{\Sigma}}$, such that

$$q_{\text{ref}}(\boldsymbol{\theta}) \quad = \quad q(\hat{\boldsymbol{\theta}}) \exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\right) \tag{2.9}$$

Then the reference normalising constant is

$$z_{\text{ref}} = q(\hat{\boldsymbol{\theta}})\sqrt{\det(2\pi\hat{\boldsymbol{\Sigma}})}\,. \tag{2.10}$$

This method of generating a reference is simple and reliable. It requires sampling from the posterior $q(\boldsymbol{\theta})$ so is more expensive than gradient-based methods, but the cost associated with drawing enough samples to generate a sufficiently good reference tend to be quite low. In the primary application discussed later, regarding structured high-dimensional Bayesian hierarchical models, we use this approach to generate a reference density and normalising constant.

Though the sampled covariance reference is typically a good approach, it is not in general optimal within the Laplace-type family of approaches — typically another Gaussian reference exists with different parameters that can generate a normalising constant closer to the true one, thus potentially leading to overall faster convergence of the thermodynamic integral to the exact value. Such an optimal reference can be identified variationally (see Hawryluk et al. 2023).

2.2.4 Reference Support

If a model involves a bounded parameter space, for example, $\theta_1 \in [0, \infty)$, $\theta_2 \in (-1, \infty)$ etc., as commonly arise in structured Bayesian models, in referenced TI the exact analytic integration for the reference density should be equivalently limited. This is necessary not only so the reference is closer to the true density to speed up convergence, but also so MCMC samples from both densities can be drawn on the same parameter space, as is required for the thermodynamic integrand in Eq. 2.6 to be well-defined. However, the calculation of arbitrary probability density function orthants (an orthant is a specific region in a multi-dimensional space or more precisely a bounded n-dimensional space), even for well-known analytic functions such as the multivariate Gaussian, is in general a difficult problem. High-dimensional orthant computations usually require advanced techniques, the use of approximations, or sampling methods [Azzimonti and Ginsbourger, 2016, Curnow and Dunnett, 1962, Miwa et al., 2003, Owen, 2014, Ridgway, 2016, Ruben, 1964]. Fortunately, we can simplify our reference density to create a reference with tractable analytic integration for limits by using a diagonal approximation to the sampled covariance or Hessian matrix. For example the orthant of a diagonal multivariate Gaussian can be given in terms of the error function [Brown, 1963],

$$z_{\text{ref}} \quad = \quad q(\hat{\boldsymbol{\theta}}) \sqrt{\det(2\pi \boldsymbol{\Sigma}^{\text{diag}})} \prod_{i \in K} \left(1 + \text{erf} \left(\frac{\hat{\theta}_i - a_i}{\sqrt{2 \Sigma_i^{\text{diag}}}} \right) \right) , \tag{2.11}$$

where K denotes the set of indices of the parameters with lower limits a_i. Σ^{diag} is a diagonal covariance matrix, that is one containing only the variance of each of the parameters, without the covariance terms and Σ_i^{diag} denotes the i^{th} element of the diagonal. Restricting our density to a diagonal one is a poorer approximation than using the full covariance matrix. In practice, however, this has not been a substantial drawback to the convergence of the thermodynamic integral—and again we state that the quality of the reference affects only convergence rather than the eventual accuracy of the normalising constant. This behaviour is observed in the practical examples later considered, though we do recognise the distinction between accuracy and convergence and matters of asymptotic consistency using an MCMC estimator with finite iterations are not clear-cut.

2.2.5 Technical Implementation

Referenced TI was implemented in Python, using PyStan for sampling and inference. Using Stan enables fast MCMC simulations with Hamiltonian Monte Carlo and No-U-Turn algorithm [Carpenter et al., 2017, Hoffman and Gelman, 2014], and portability between other statistical languages. In the online repository, we also provide an example of carrying out referenced-TI in NumPyro [Phan et al., 2019]. The code for all examples shown in this chapter is available at `https://github.com/mrc-ide/referenced-TI`. In the examples shown in Section 2.3, we used 4 chains with 20,000 iterations per chain for the pedagogical examples, and 4 chains with 2,000 iterations for the other applications. In all cases, half of the iterations were used for the burn-in (alternatively warm-up, [Carpenter et al., 2017]). Mixing of the chains and the sampling convergence was checked in each case, by ensuring that the $\hat{R}$ value was ≤ 1.05 and investigating the trace plots. The $\hat{R}$ is a standard MCMC diagnostic that assesses the convergence of multiple parallel chains running in an MCMC simulation. It is computed by comparing the variance between chains to the variance within each chain. If chains have not fully converged, their variances will be larger compared to when they have reached convergence [Gelman and Rubin, 1992].

In all examples, the integral given in Eq. 2.5 was discretised to allow computer simulations. Each expectation $\mathbb{E}_{q(\lambda;\theta)}\left[\log \frac{q_1(\theta)}{q_0(\theta)}\right]$ was evaluated at $\lambda = 0.0, 0.1, 0.2, ..., 0.9, 1.0$, unless stated otherwise. To obtain the value of the integral in Eq. 2.5, we interpolated a curve linking the expectations using a cubic spline, which was then integrated numerically. The pseudo-code of the algorithm with sampled covariance Laplace reference is shown in Algorithm 1.

2.3 Applications

In this section, we present applications of the referenced TI approach. In subsections 2.3.1 and 2.3.2 we give 1- and 2-dimensional pedagogical introductions to the approach. In subsection 2.3.3 we select a linear regression model for a well-known problem in the statistical literature, and finally in subsection 2.3.4 we consider a challenging model selection task for a structured Bayesian model of the COVID-19 epidemic in South Korea.

Algorithm 1 Referenced thermodynamic integration algorithm

Input q - un-normalised density, q_{ref} - un-normalised reference density, Λ - set of coupling parameters λ, N - number of MCMC iterations

Output z - normalising constant of the density q

Define un-normalised density q and the reference density q_{ref}
Calculate z_{ref} analytically by using the determinant of the covariance matrix, as per Eqs 2.8 or 2.10.
for $\lambda \in \Lambda$ **do**
    Sample N values θ_n from $q^{\lambda} q_{\text{ref}}^{1-\lambda}$
    for $n = 1, 2, \ldots, N$ **do**
        Calculate $\log \frac{q(\theta_n)}{q_{\text{ref}}(\theta_n)}$
    end for
    Compute the mean, $\mathbb{E}_{\lambda} = \frac{1}{N} \Sigma_{n=1}^{N} \log \frac{q(\theta_n)}{q_{\text{ref}}(\theta_n)}$
end for
Interpolate between the consecutive $\mathbb{E}_{\lambda}$ values to obtain a curve $\partial_{\lambda} \log(z(\lambda))$
Integrate $\partial_{\lambda} \log(z(\lambda))$ over $\lambda \in [0, 1]$ to get $\log \frac{z}{z_{\text{ref}}}$
Calculate $z = z_{\text{ref}} \cdot \exp\left(\log \frac{z}{z_{\text{ref}}}\right)$

2.3.1 1D Pedagogical Example

To illustrate the technique we consider the 1-dimensional density

$$q(\theta) = \exp\left(-\frac{1}{2}\sqrt{|\theta - 4|} - \frac{1}{2}(\theta - 4)^4\right), \ \theta \in \mathbb{R}, \tag{2.12}$$

with normalising constant $z = \int_{-\infty}^{\infty} q(\theta) d\theta$. This density has a cusp and it does not have an analytical integral that easily generalises to multiple dimensions.

In this instance the Laplace approximation based on the second-order Taylor expansion at the mode will fail due to the cusp, so we use the more robust covariance sampling method. Sampling from the 1D density $q(\theta)$ we find a variance of $\hat{\sigma}^2 = 0.424$, giving a Gaussian reference density $q_{\text{ref}}(\theta)$ with normalising constant of $z_{\text{ref}} = 1.559$. The full normalising constant, $z = z_{\text{ref}} \frac{z}{z_{\text{ref}}}$, is evaluated by Eq. 2.6, by setting up a thermodynamic integration along the sampling path $q^{\lambda} q_{\text{ref}}^{(1-\lambda)}$. The expectation, $\mathbb{E}_{q(\lambda;\boldsymbol{\theta})}\left[\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})}\right]$, is evaluated at 5 points along the coupling parameter path $\lambda = 0.0, 0.2, 0.5, 0.8, 1.0$, shown in Fig. 2.2. In this simple example, the integral can be easily evaluated to high accuracy using quadrature (Piessens et al., 1983; Virtanen et al., 2020), giving a value of 1.523. Referenced TI reproduces this value, with the convergence of z shown in Fig. 2.2, converging to 1% of z with 500 iterations and 0.1% within 17,000 iterations.

This example illustrates notable characteristic features of the referenced TI. Here the reference $q_{\text{ref}}(\theta)$ is a

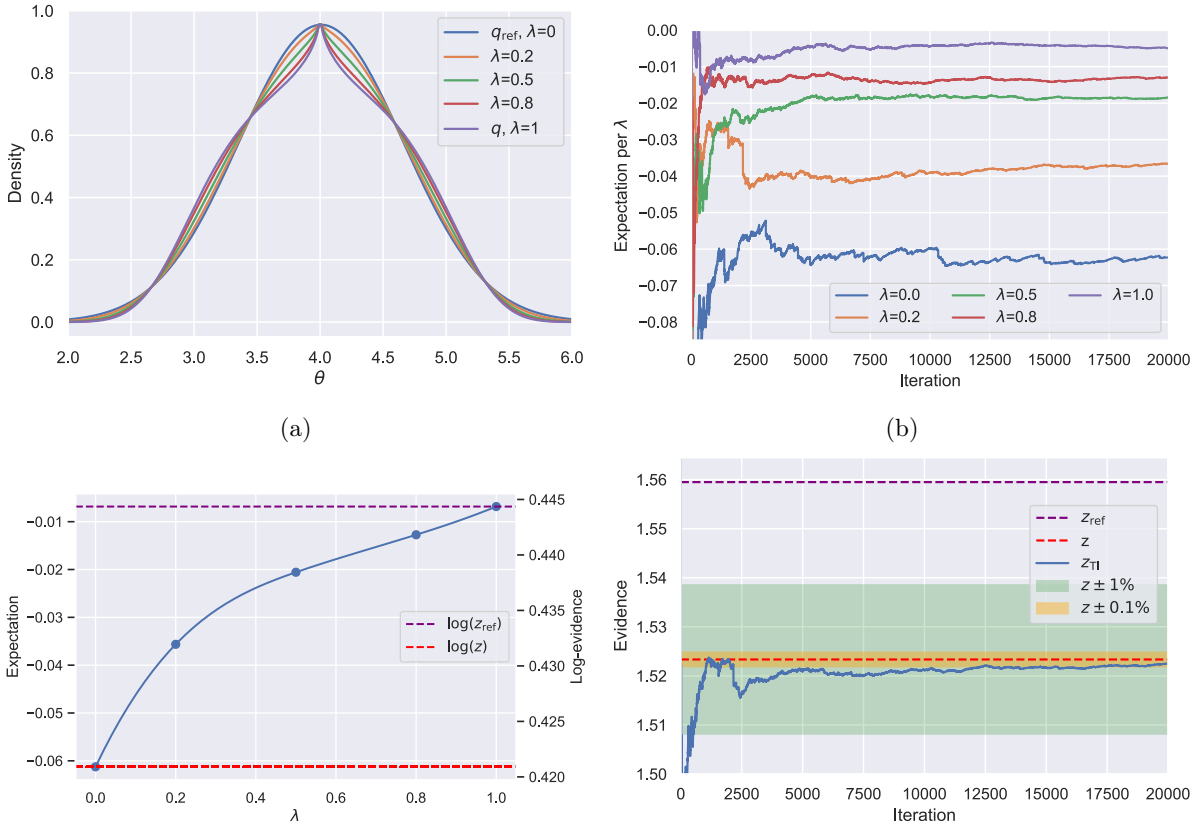


Figure 2.2: Illustration of steps from the 1D pedagogical example. (a) $q^{\lambda}q_{\text{ref}}^{(1-\lambda)}$ for the 1d example density in parameter θ (Eq. 2.12) at selected λ values along the path. (b) Expectation $\mathbb{E}_{q(\lambda;\boldsymbol{\theta})}\left[\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})}\right]$ vs MCMC iteration, shown at each value of λ sampled. (c) λ-dependence of $\mathbb{E}_{q(\lambda;\boldsymbol{\theta})}\left[\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})}\right]$, the TI contribution to the log-evidence. (d) Convergence of the evidence z, with 1% convergence after 500 iterations and 0.1% after $17,000$ iterations per λ.

good approximation to $q(\theta)$, with z_{ref} accounting for most of z ($z_{\text{ref}} = 1.02z$). Consequently $\frac{z}{z_{\text{ref}}}$ is close to 1, and the expectations, $\mathbb{E}_{q(\lambda;\boldsymbol{\theta})}\left[\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})}\right]$, evaluated by MCMC for the remaining part of the integral are small. For the same reasons, the variance at each λ is small, leading to favourable convergence within a small number of iterations. And finally $\mathbb{E}_{q(\lambda;\boldsymbol{\theta})}\left[\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})}\right]$ weakly depends on λ, so there is no need to use a very fine grid of λ values or consider optimal paths—satisfactory convergence is easily achieved using a simple geometric path with 4 λ-intervals.

2.3.2 2D Pedagogical Example with Constrained Parameters

As a second example, consider a 2-dimensional unnormalised density function with a constrained parameter space:

$$
\begin{aligned}
q(\theta_1, \theta_2) &= \exp(-\Theta)\,, \\
\Theta &= \frac{1}{4} \sum_{i,j \in \{1,2\}} \left(\theta_i + \frac{1}{2}\right)^{2j} + \frac{1}{8}\theta_1 \theta_2^2\,, \\
\theta_1 &\in [0, +\infty) \text{ and } \theta_2 \in (-\infty, +\infty)\,.
\end{aligned} \tag{2.13}
$$

A reference density $q_{\text{ref}}(\boldsymbol{\theta})$ can be constructed from the Hessian at the mode of $q(\boldsymbol{\theta})$. To marginalise it on a parameter space with restricted support, we use a reference density $q_{\text{ref}}^{\text{diag}}(\boldsymbol{\theta})$ based on a diagonal Hessian, that has an exact and easy to calculate orthant. All densities are shown in Fig. 2.3.

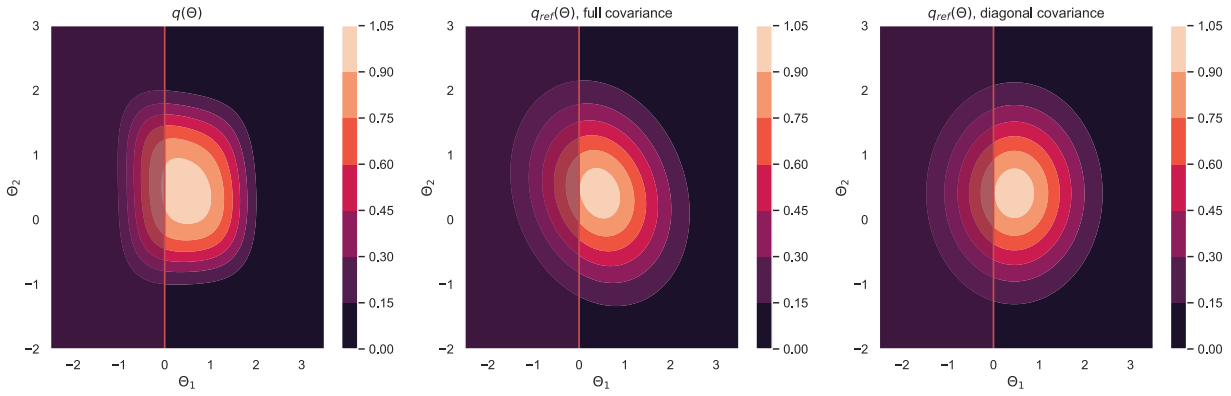


Figure 2.3: Contour plots of the un-normalised density q (left) and its two reference densities q_{ref}, one using a full covariance matrix (middle) and another using a diagonal covariance matrix (right) that can be easily marginalised. The red line shows the lower boundary $\theta_1 = 0$ and the shaded $\theta_1 < 0$ region to the left of the line is outside of the support of the density q.

To obtain the log-evidence of the model, we calculated the exact value numerically [Piessens et al., 1983, Virtanen et al., 2020], and using a sampled diagonal covariance matrix, as per Eq. 2.11, to account for the lower bound of the parameter θ_1. Without this restriction the final normalising constant is overestimated – if the support of the parameters in the MCMC is not the same as for the analytic z_{ref} calculation, z_{ref} as shown in Eq 2.6 does not cancel with the TI reference. A numerical comparison of the referenced TI to quadrature is presented in Table 2.1.

Table 2.1: Evidence calculated with different methods. Constraint correction refers to imposing the integration limits on the reference as per Eq. 2.11. Diagonal covariance means a covariance matrix where only the diagonal (variance) terms are non-zero. Full covariance is a covariance matrix in which all terms can be non-zero. * obtained numerically [Piessens et al., 1983, Virtanen et al., 2020].

Method	Evidence
Exact*	3.31
Laplace with full covariance	5.55
Laplace with diagonal covariance + constraint correction	3.81
Ref TI with full covariance	4.79
Ref TI with diagonal covariance + constraint correction	3.33

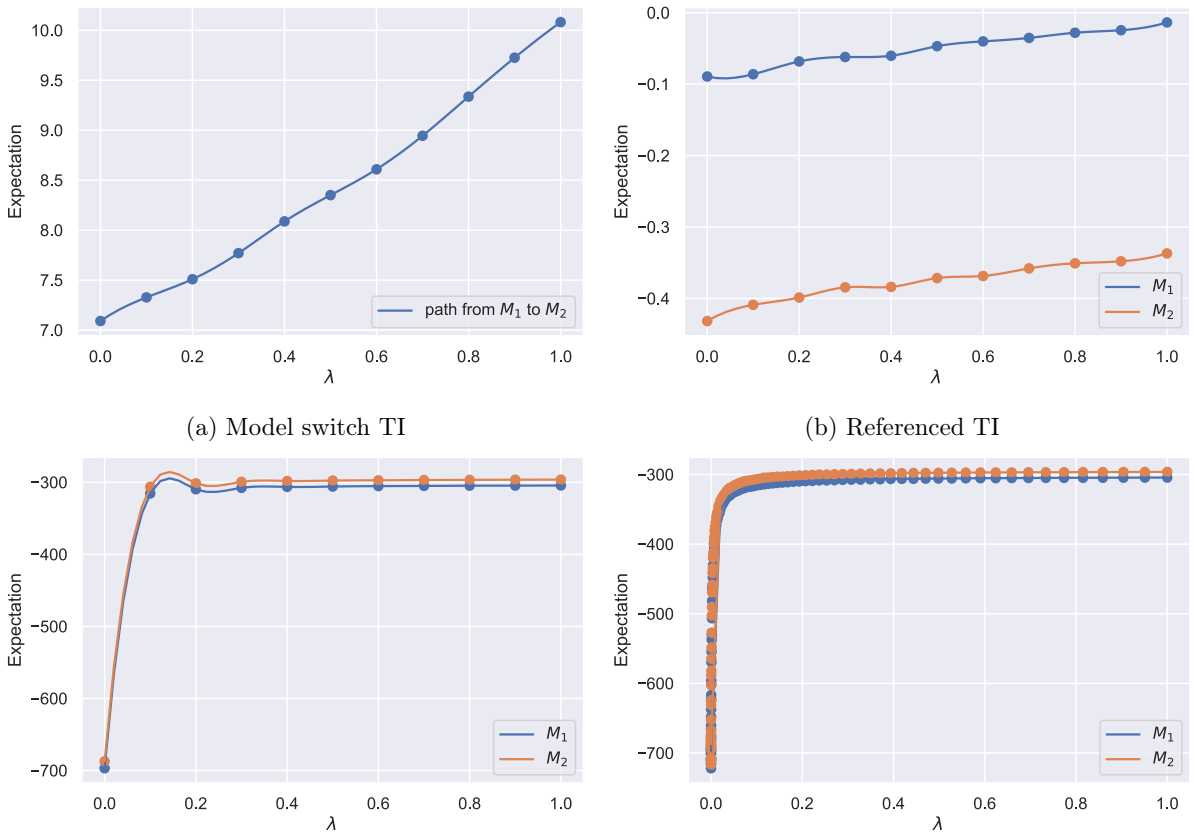
2.3.3 Benchmarks—*Radiata Pine*

To benchmark the application of the referenced TI in the model selection task, two non-nested linear regression models are compared for the *radiata pine* data set [Williams, 1959]. This example has been widely used for testing normalising constant calculating methods since in this instance the exact value of the model evidence can be computed. The data consists of 42 3-dimensional data points, expressed as y_i - maximum compression strength, x_i - density and z_i - density adjusted for resin content. In this example, we follow the approach of Friel and Wyse [2012], using the priors from therein, and test which of the two models M_1 and M_2 provides better predictions for the compression strength:

$$\begin{aligned} M_1 &: y_i = \alpha + \beta(x_i - \bar{x}) + \epsilon_i, \epsilon_i \sim N(0, \tau^{-1}), i = 1, ..., n\,, \\ M_2 &: y_i = \gamma + \delta(z_i - \bar{z}) + \eta_i, \eta_i \sim N(0, \rho^{-1}), i = 1, ..., n\,. \end{aligned} \tag{2.14}$$

In this example, we tested five methods of estimating the model evidence: Laplace approximation using a sampled covariance matrix, model switch TI along a path directly connecting the models [Lartillot and Philippe, 2006, Vitoratou and Ntzoufras, 2017] (by model switch TI we mean a classic thermodynamic integration method of estimating a ratio of two evidences, that is without the use of the reference density and without calculating the two evidences separately), referenced TI, power posteriors with equidistant 11 λ-placements (labelled here as PP_{11}) and power posteriors with 100 λ-s (PP_{100}), following the example from Friel and Wyse [2012]. For the model switch TI, referenced TI and PP_{11} we used $\lambda \in \{0.0, 0.1, ..., 1.0\}$.

The expectation from MCMC sampling per each λ for model switch TI, referenced TI, PP_{11} and PP_{100} and fitted cubic splines between the expectations are shown in Fig. 2.4. Both reference and model switch TI methods eliminate the problem of divergence of expectation for $\lambda = 0$, which is observed with the power posteriors, where samples for $\lambda = 0$ come from the prior density function. And both reference and model switch have smaller residuals for splines in λ fitted to $\mathbb{E}_{q(\lambda;\boldsymbol{\theta})}\left[\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})}\right]$ than power posteriors.



(a) Model switch TI

(b) Referenced TI

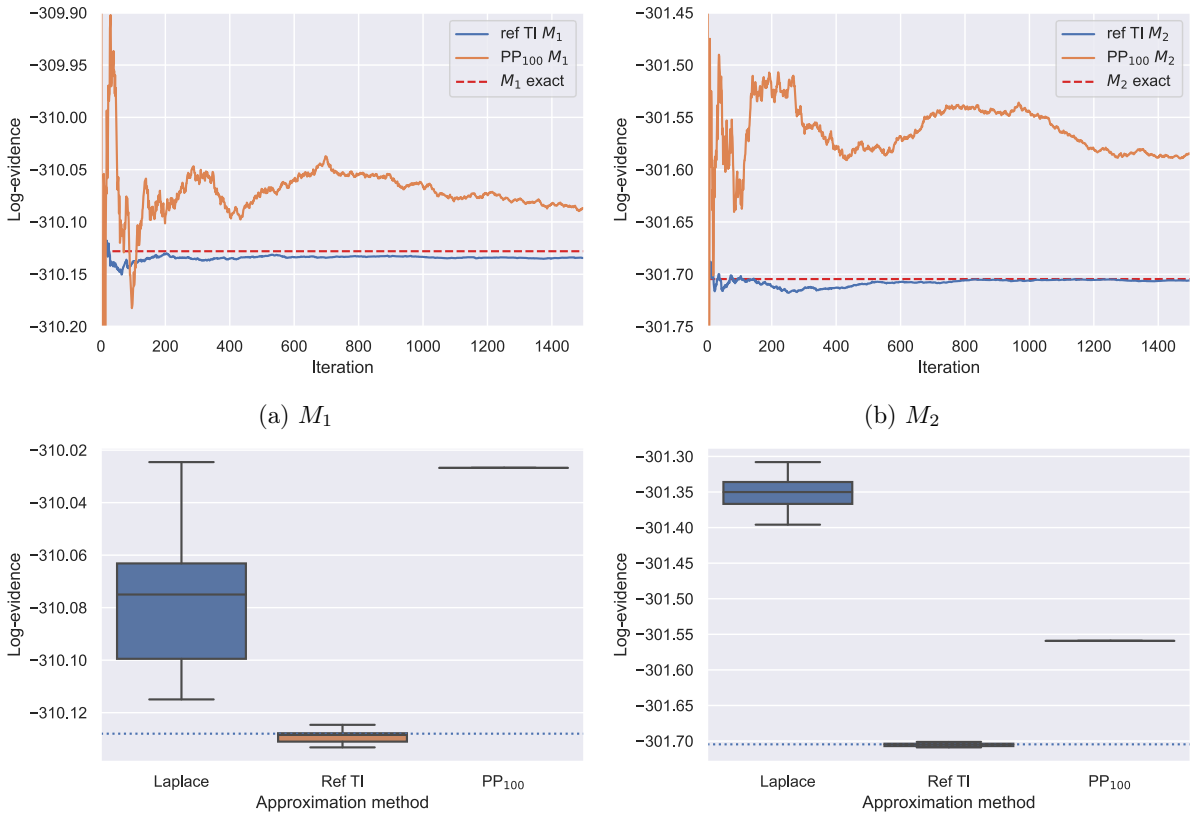
(c) PP_{11}

(d) PP_{100}

Figure 2.4: The HMC-evaluated expectation of $\mathbb{E}_{q(\lambda;\boldsymbol{\theta})}\left[\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})}\right]$ vs coupling parameter λ is given for models M_1 and M_2 for four methods of calculating the model evidence: (a) model switch TI, (b) referenced TI, (c) power posteriors with 11 λ-placements and (d) power posteriors with 100 λ placements. Model switch TI (a) creates the path directly between two competing densities, therefore only one line is shown (see Eq. 2.5). In each of the plots, the evaluated expectation for a given λ is shown with the dot, and the lines connecting the dots were obtained through interpolation.

For each approach, the splines fitted to $\mathbb{E}_{q(\lambda;\boldsymbol{\theta})}\left[\log \frac{q(\boldsymbol{\theta})}{q_{\text{ref}}(\boldsymbol{\theta})}\right]$ were integrated to obtain the log-evidence for models M_1 and M_2, and the log-ratio of the two models' evidences for the model switch TI. The rolling means of the integral over 1500 iterations for referenced TI and PP_{100} for M_1 and M_2 are shown in Fig. 2.5a and b. We can see from the plots, that referenced TI presents favourable convergence to the exact value,

whereas PP_{100} oscillates around it. Figs 2.5c and d show the distribution of log-evidence for the same model generated by 15 runs of the three algorithms: Laplace approximation with a sampled covariance matrix, referenced TI and PP_{100}. Although all three methods resulted in a log-evidence satisfactorily close to the exact solution, the referenced TI was the most accurate and importantly, converged fastest — 308 MCMC draws compared to 55,000 draws needed for the power posterior method to achieve a standard error of 0.5%, excluding burn-in (see Table 2.2).



(a) M_1 (b) M_2

(c) M_1 (d) M_2

Figure 2.5: Log-evidence of M_1 and M_2 for the three algorithms. (a) and (b) show the rolling mean of log-evidence of M_1 and M_2 over 1500 iterations per λ obtained by referenced TI (blue line) and PP_{100} (orange line) methods. The exact value is shown with a red dashed line. (c) and (d) show the mean log-evidence of the two models evaluated over 15 runs of the three algorithms. The exact value of the log-evidence is shown with the dotted line.

2.3.4 Model Selection for the COVID-19 Epidemic in South Korea

The final example of using referenced TI for calculating model evidence is fitting a renewal model to COVID-19 case data from South Korea during the first 200 days of the pandemic. The model is based on

Table 2.2: Comparison of Bayes factors for *radiata pine* models for each method. Here we show $BF_{21} = \frac{M_2}{M_1}$ to determine whether model M_2 is better than model M_1. Both TI and referenced TI methods used 11 equidistant λ-s. Power posteriors method was used with 11 (PP_{11}) and 100 (PP_{100}) λ-s. The third column shows the total number of MCMC steps required to achieve a standard error of 0.5%, excluding the warm-up steps. * - using sampled covariance matrix.

Method	BF_{21}	MCMC steps
Exact	4552.35	-
Laplace approximation*	6309.10	-
Model switch TI	4557.63	2,365
Referenced TI	4558.71	308
PP_{11}	4463.71	41,514
PP_{100}	4757.82	55,000

a statistical representation of a stochastic branching process whose expectation mechanistically follows a renewal-type equation. Its derivation and details are explained in Pakkanen et al. [2023] and Mishra et al. [2020] and a short explanation of the model is provided below.

COVID-19 Model

The COVID-19 model shown is based on the renewal equation derived from the Bellman-Harris process [Mishra et al., 2020, Pakkanen et al. 2023]. Here, we give a short overview of the $AR(2)$ model (see 1.5.2). The model has a Bayesian hierarchical structure and is fitted to the time-series data containing numbers of new confirmed COVID-19 cases per day in South Korea from 31st Dec 2019 to 18th July 2020, obtained from `https://opendata.ecdc.europa.eu/covid19/casedistribution/csv`. The incidence $y(t)$ is modelled by a negative binomial distribution, with a mean parameter taking the form of a renewal equation. That is, the number of new confirmed cases $y(t)$ is modelled as:

$$y \sim \text{NegBin}(f(t), \phi)\,, \tag{2.15}$$

where ϕ is an overdispersion or variance parameter and the mean of the negative binomial distribution is denoted as $f(t)$ and represents the daily case data through a renewal part:

$$f(t) = R_0 \int_{\tau=0}^{t} f(t-\tau)g(\tau)d\tau\,. \tag{2.16}$$

As the case data is not continuous but is reported per day, $f(t)$ can be represented in a discretised, binned form as:

$$f(t) = R_t \sum_{\tau<t} f(t-\tau)g(\tau)\,. \tag{2.17}$$

Here, $g(\tau)$ is a Raleigh-distributed serial interval with mean GI, which is discretised as

$$g_s = \int_{s-0.5}^{s+0.5} g(\tau)d\tau \text{ for } s = 2, 3, ... \text{ and } g_1 = \int_{0}^{1.5} g(\tau)d\tau\,. \tag{2.18}$$

R_t, the effective reproduction number, is parametrised as $R_t = \exp(\epsilon_t)$, with exponent ensuring positivity. ϵ_t is an autoregressive process with two-days lag, that is AR(2), with $\epsilon_1 \sim N(-1, 0.1)$, $\epsilon_2 \sim N(-1, \sigma)$ and

$$\epsilon_t \sim N(\rho_1\epsilon_{t-1} + \rho_2\epsilon_{t-2}, \sigma_t) \text{ for } t = \{3, 4, 5, ...\}. \tag{2.19}$$

The model's priors are:

$$\begin{aligned}
\sigma &\sim N^{+}(0, 0.2)\,,\\
\rho_1 &\sim N^{+}(0.8, 0.05)\,,\\
\rho_2 &\sim N^{+}(0.1, 0.05)\,,\\
\phi &\sim N^{+}(0, 5)\,,\\
GI &\sim N^{+}(0.01, 001)\,.
\end{aligned} \tag{2.20}$$

The model is fitted to the time-series case data and estimates a number of parameters, including serial interval and the effective reproduction number, R_t.

Three modifications of the original model are tested here:

- variation of the infection generation interval for values $GI = 5, 6, 7, 8, 9, 10, 20$ — where GI denotes the mean of Rayleigh-distributed generation interval,
- changing the order of the autoregressive model for the reproduction number, for $AR(k)$ with $k = 2, 3, 4$-days lags,
- varying the length of the sliding window for estimating the reproduction number for values in $W = 1, 2, 3, 4, 7$ days.

Within each group of models, GI, AR and W, we want to select the best model through the highest evidence method. The dimension of each model was dependent on the modifications applied, but in all the cases the normalising constant was a 40- to 200-dimensional integral. The log-evidence of each model was calculated using the Laplace approximation with a sampled covariance matrix, and then correction to the estimate was obtained using the referenced TI method.

The first group of models analysed was the $AR(2)$ model described above but with the GI parameter fixed to a certain value instead of inferring that parameter from the data. $AR(3)$ and $AR(4)$ models had additional parameters ρ_3 and ρ_4, which allowed us to model the autoregressive process with a longer lag (3- and 4- days respectively). Finally, models $W = k$, $k = 1, .., 7$ were similar to the $AR(2)$ model, but the underlying assumption of these models is that the R_t stays constant for the duration of the length of the sliding window $W = k$.

Results

Values of the log-evidence for each model calculated by both Laplace and referenced TI methods are given in Table 2.3. Interestingly, the favoured model in each group, that is the model with the highest log-evidence, was different when the evidence was evaluated using the Laplace approximation than when it was evaluated with referenced TI. For example, using the Laplace method, a sliding window of length 7 was incorrectly identified as the best model, whereas with referenced TI window of length 2 was chosen to be the best among the tested sliding windows models, which agrees with the previous studies of the window-length selection in H1N1 influenza and SARS outbreaks [Parag and Donnelly, 2020]. This exposes how essential it is to accurately determine the evidence, even good approximations can result in misleading results. Bayes factors for all model pairs are shown in Fig. 2.6.

Table 2.3: Log-evidence estimated by Laplace and referenced TI approximations. In each section, a model with the highest log-evidence estimated by Laplace or referenced TI method is indicated in bold. The credible intervals for log-evidence come from calculating the quantiles of the integral from Eq. 2.5, where the integral values were obtained from the spline interpolated using running means of the expectations per λ over all iterations.

Model	Log-evidence Laplace	Log-evidence ref TI [95% CrI]
GI=5	-1274	-716 [-715.6, -715.2]
GI=6	-1274	-703 [-703.3, -702.7]
GI=7	-1255	-732 [-732.4, -731.8]
GI=8	-1245	**-685** [-685.5, -684.7]
GI=9	-1310	-803 [-802.8, -802.3]
GI=10	-1313	-805 [-805.1, -805.3]
GI=20	**-1170**	-796 [-796.3, -795.5]
AR(2)	**-1207**	-711 [-711.2, -710.6]
AR(3)	-1293	**-704** [-704.7, -703.7]
AR(4)	-2166	-821 [-820.6, -819.2.]
W=1	-1260	-802 [-802.1, -801.6]
W=2	-1069	**-791** [-791.2, -790.7]
W=3	-1003	-807 [-807.5, -807.2]
W=4	-940	-811 [-811.1, -810.7]
W=7	**-875**	-814 [-813.7, -813.5]

Interpretation of the COVID-19 model selection

The importance of performing model selection rigorously is clear from Fig. 2.7, where the generated R_t time-series are plotted for the models favoured by Laplace and referenced TI methods. The differences in the R_t time-series show the pitfalls of selecting an incorrect model. The differences between the two favoured models were most extreme for the $GI = 8$ and $GI = 20$ models. While a $GI = 8$ is plausible, even likely for COVID-19, $GI = 20$ is implausible given observed data [Bi et al., 2020]. This is further supported by observing that for $GI = 20$, favoured by the Laplace method, R_t reached the value of over 125 in the first peak—around 100 more than for the $GI = 8$. The second peak was also largely overestimated, where R_t reached a value of 75. It is important here to note, that we chose the South Korea COVID-19 cases dataset for demonstrating purposes rather than for the conservative epidemiological analysis. This time series contains only cases which were reported, but many have been undetected due to lower diagnostics capacity at the time of the study and high numbers of symptomless infections, as discussed in 1.1.2. For a more reliable estimate of the epidemiological quantities from that time, we would refer to recorded deaths,

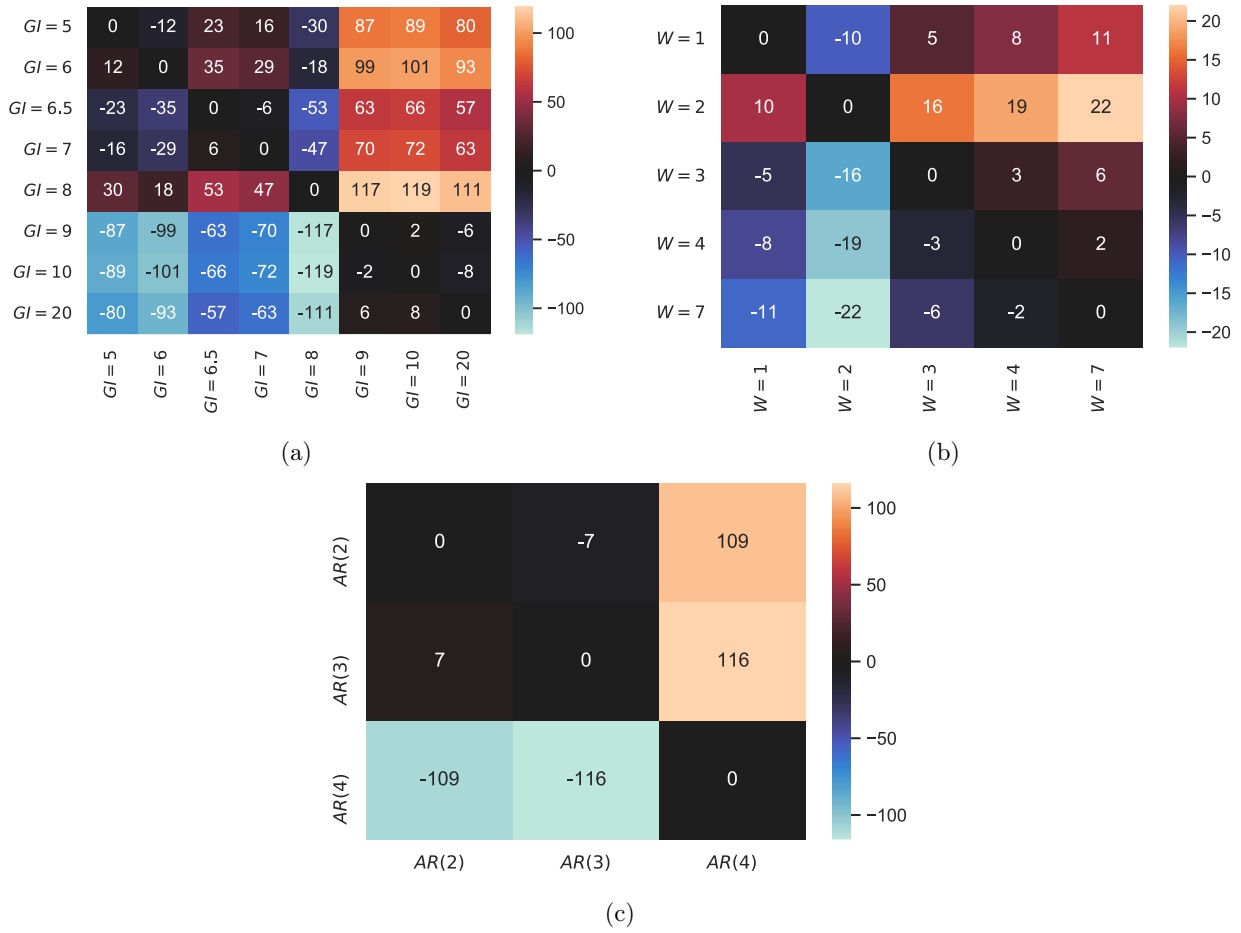
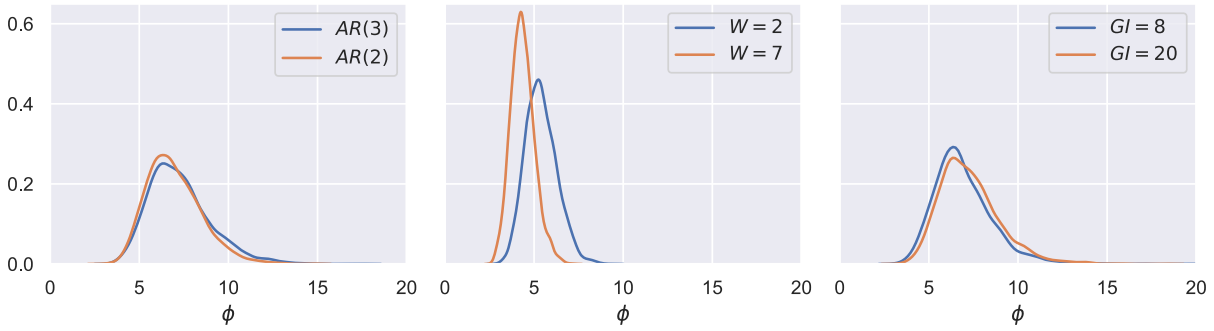
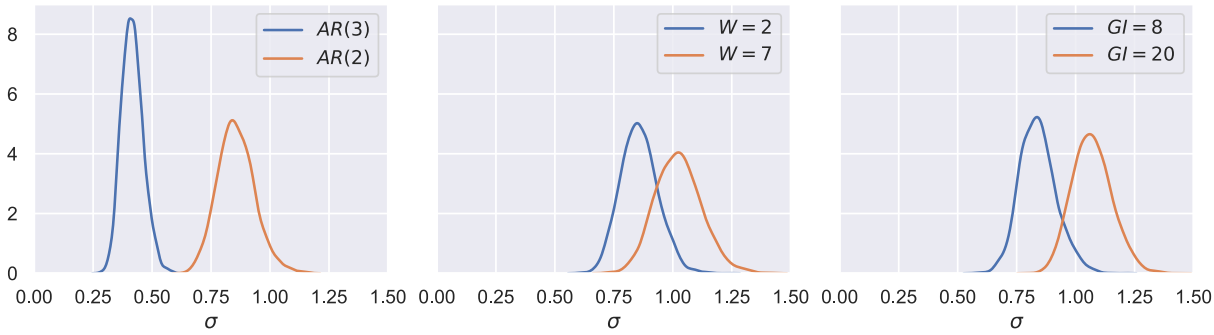


Figure 2.6: Logarithms of Bayes factors for the analysed COVID-19 renewal models, evaluated using the normalising constants ratios obtained by referenced TI. In each cell, the colour indicates the value of the $BF_{1,2}$ for models M_1 (row) and M_2 (column). Higher values (brighter orange colour) suggest that M_1 is better than M_2, and values below 0 (blue palette) indicate that M_1 is worse than M_2. $GI = 8$ performed best out of fixed GI models (a), $W = 2$ best out of sliding window models (b), and $AR(3)$ performed better than $AR(2)$ and $AR(4)$ when the autoregressive lag was varied (c). For the interpretation of the BF values see [Kass and Raftery, 1995].

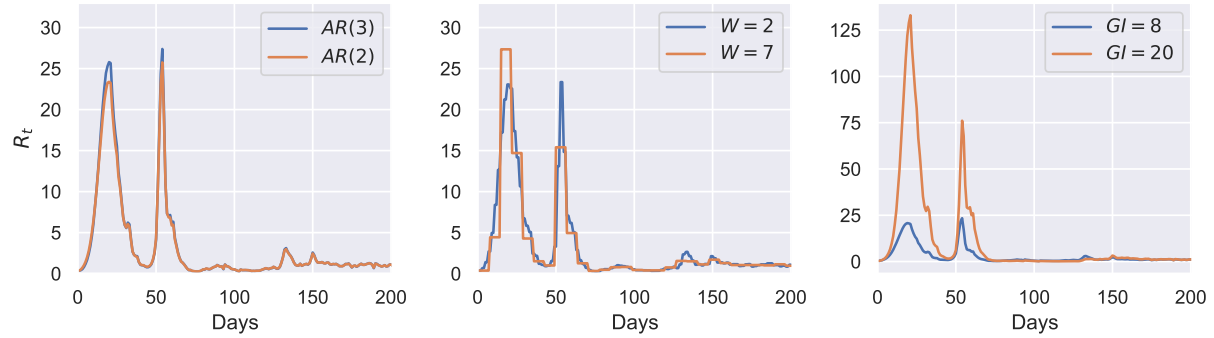
as they would be much more trustworthy. Additionally, the South Korean dataset potentially contains super-spreading events, which would require special treatment in the analysis in order to avoid R_t spikes for the general population.



(a) Posterior distributions for overdispersion parameter ϕ



(b) Posterior distributions for σ parameter



(c) R_t generated by the favoured models

Figure 2.7: Posterior distributions for models' parameters for models favoured by BFs using the Laplace approximation (orange lines) and referenced TI (blue lines). Note, that the fitting data in this example contains superspreading events and is not reliable at the time numbers of cases of COVID-19 rather than deaths (which leads to very high values of R_t on certain days) so is not representative of SARS-CoV-2 transmission generally.

We find it interesting to note that all models present a similar fit to the confirmed COVID-19 cases data (see Fig. 2.8). This makes it impossible to select the best model through visual inspection and comparison of the model fits, or by using model selection methods that do not take the full posterior distributions into account. Although the models might fit the data well, other quantities generated, which are often of interest to the modeller, might be completely incorrect. Moreover, it emphasises the need to test multiple models before any conclusions are drawn, especially with complex, hierarchical models.

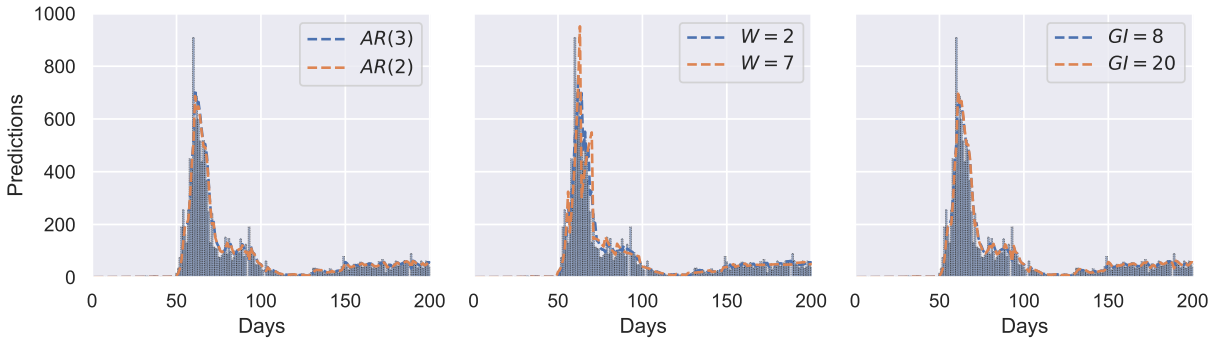


Figure 2.8: Cases of SARS-CoV-2 infections in South Korea from the data (shown with bars) and the cases predicted by different models. On each graph, predictions made by the model favoured by the Laplace approximation are shown with a blue dashed line, and predictions made by the referenced TI-favoured models are shown with an orange dashed line. The lines in all three subplots are largely overlapping, revealing that all models fit the case data similarly well.

The Laplace approximation can often be sufficient to pick the best model, but not in this epidemiological model selection problem. We can see in Table 2.3 that the evidence was the highest for the "boundary" models when the Laplace approximation was applied. For example, for the sliding window length models, when the Gaussian approximation was applied, the log-evidence was monotonically increasing with the value of W within the range of values that seem reasonable ($W = 1$ to 7). In contrast, with referenced TI, the log-evidence geometry is concave within the range of *a priori* reasonable parameters.

2.4 Discussion

Summary of the findings

The examples shown in Section 2.3 illustrate the applicability of the referenced TI approach for calculating model evidence. In the *radiata pine* example, referenced TI performed better than the other tested methods

in terms of accuracy and speed. When using referenced TI, at $\lambda = 0$ values are sampled from the reference density rather than the prior, as is the case for the power posterior method, which should be closer to the original density (in the sense of Kullback-Leibler or Jensen-Shannon divergence, see 1.3.5). This leads not only to a more accurate estimate of the normalising constant but also to much faster convergence of the MCMC samples. A detailed theoretical characterisation of rates of convergence is beyond the scope of this chapter, nonetheless, the empirical tests presented consistently show faster convergence than with comparative approaches. This is useful, especially for evaluating model evidence in complex hierarchical models where each MCMC iteration is computationally demanding.

In the application to the renewal model for the COVID-19 epidemic in South Korea, we showed that for a complex structured model, hypothesis selection by Laplace approximation of the normalising constant can give misleading results. Using the referenced TI, we calculated model evidence for 16 models, which enabled a quick comparison between chosen pairs of competing models. Importantly, the evidence given by the referenced TI was not monotonic with the increase of one of the parameters, which was the case for the Laplace approximation. The referenced TI method presented here will be similarly useful in other situations, particularly where the high-dimensional posterior distribution is uni-modal but non-Gaussian.

Model selection methods

In epidemiology, as well as other fields, the best model is often picked using simple methods such as AIC or BIC [Grassly and Fraser, 2008]. Their main advantage is that they can be computed with virtually no additional computational cost and are often provided by off-the-shelf statistical software. It is important to note, that neither AIC nor BIC incorporates the uncertainty of the parameters and the model's predictions [Kass and Raftery, 1995]. As a result, AIC often favours the models with a larger number of parameters, and although BIC penalises more complex models, it might give misleading answers when multiple models with similar scores are compared [Kass and Raftery, 1995]. Additionally, none of these methods takes into account prior probabilities of the parameters [Xie et al., 2011] and does not force the researcher to consider prior predictive checks on model appropriateness [Yang, 2005]. Fully Bayesian model selection methods which compare the models' evidences are often more appropriate, as they take into account the full posterior densities, instead of just maximum a posteriori probability estimates [Pooley and Marion, 2018].

Another model selection method, popular due to its accuracy and computational efficiency is the Leave-one-out cross-validation (LOO-CV) algorithm using Pareto-smoothed importance sampling (LOO-CV PSIS) and Widely Applicable Information Criterion (WAIC) [Vehtari et al., 2017]. The main benefit of these methods is that they use the MCMC samples in order to assess the goodness of fit, instead of re-fitting the model as is normally the case with cross-validation (CV). LOO-CV generally fails when applied to the temporal data, such as the time-series cases data for the COVID-19 model from subsection 2.3.4, because the assumption of conditional exchangeability between observations is not satisfied for these types of data [Vehtari et al., 2019]. However, these shortcomings of the LOO-CV methods are now partially resolved [Bürkner et al., 2020].

Using model evidence and BFs as a method of model selection has been disputed in scientific literature [Llorente et al., 2023, Navarro, 2019, Vehtari and Ojanen, 2012]. BFs have been shown to be sensitive to the choice of priors of the model, which some consider to be a disadvantage. However, we show in Section 2.3.4 that sometimes it proves useful when the competing models generate outputs that cannot be compared to the empirical data. It is also possible that all competing models make similar predictions under sensible priors, but their marginal likelihood might differ substantially once integrated over the parameter space [Vehtari and Ojanen, 2012]. Due to this sensitivity, evaluating the BFs over a range of possible prior choices is recommended, which may become computationally expensive [Kass and Raftery, 1995]. In general, however, if a sample size is sufficiently large, the effect of the priors is small [Kass and Raftery, 1995, Vehtari and Ojanen, 2012]. Worth noting as well, that BFs indicate which one of the two models is relatively better, and do not give an absolute score of an individual model's accuracy. For a comprehensive discussion of the applicability of BFs, we refer the reader to Llorente et al. [2023], Vehtari and Ojanen [2012].

Limitations and future work

Although referenced thermodynamic integration and other methods using path-sampling have theoretically asymptotically exact Monte Carlo estimator limits, in practice several considerations affect accuracy. For example, biases will be introduced to the referenced TI estimate if one endpoint density substantially differs from another. An example of this and explanation is provided in Hawryluk et al. [2023].

Furthermore, the discretisation of the coupling parameter path in λ can introduce a discretisation bias.

For the power posteriors method, Friel et al. [2017] propose an iterative way of selecting the λ-placements to reduce the discretisation error. Calderhead and Girolami [2009] test multiple λ-placements for 2- and 20D regression models and report relative bias for each tested scenario. In the referenced TI algorithm discretisation bias is negligible — the use of the reference density results in TI expectations that are both small and have low variance, and therefore curvature with respect to λ. In our framework, we use geometric paths with equidistant coupling parameters λ between the un-normalised posterior densities. There exist other possible choices of the path constructions, for example, a harmonic [Gelman and Meng, 1998] or hypergeometric path [Vitoratou and Ntzoufras, 2017]. This optimisation might be worth exploring, however, as illustrated in Fig. 2.4, the expectations evaluated vs λ are typically near-linear with referenced TI suggesting limited gains, although the extent of this will differ from problem to problem.

We presented two possible theoretical methods of building the reference density, but we acknowledge that there might be other ways of doing that, e.g. using variational approximation or multi-step Taylor expansion. These ways are further developed in Hawryluk et al. [2023]. Additionally, we have shown that the referenced TI converges faster than the model switch TI and power posteriors method. It would be worthwhile to further test the speed of the algorithm against other available methods of simulating normalising constants.

From the perspective of modeling emerging pathogens, where the results of repeated, real-time analyses are required immediately, the computational load of calculating Bayes factors presents a significant disadvantage. The time required for the referenced thermodynamic integration algorithm to converge highly depends on the dimensionality of the model and the appropriate choice of the reference density. As such, we do not advocate that this method is superior to other widely used methods readily available through statistical packages in R or Python, such as cross-validation or DIC, especially for modeling the ongoing pandemic. However, we believe this method should be known to practitioners from various domains, including epidemiology, and should be easy to implement for their models, particularly in cases where cross-validation gives ambiguous outputs, or when they want the priors of the model to have a greater influence on the selected best model.

2.5 Conclusions

Normalising constants are fundamental in Bayesian statistics. In this chapter, I gave an account of referenced thermodynamic integration (TI), in terms of theoretical consideration regarding the choice of reference, and showed how it can be applied to realistic practical problems. I showed how referenced TI allows efficient calculation of a single model's evidence by sampling from geometric paths between the unnormalised density of the model and a judiciously chosen reference density — here, a sampled multivariate normal that can be generated and integrated with ease. The referenced TI method has practical utility for substantially challenging problems of model selection in epidemiology and we suggest it has applicability in other fields of applied machine learning that rely on high-dimensional Bayesian models.

Building on this work, in the next chapter I further apply the Bayes factors and referenced thermodynamic integration in an epidemiological setting. Specifically, I use the method within a Bayesian hierarchical model to select the best-fitting densities (likelihoods) for various hospitalisation distributions, such as onset-to-death, which play a crucial role in modelling COVID-19 infections.

Chapter 3

Inference of COVID-19 epidemiological distributions from Brazilian hospital data

Knowing COVID-19 epidemiological distributions, such as the time from patient admission to death, is directly relevant to effective primary and secondary care planning, and the mathematical modelling of the pandemic generally. We determine epidemiological distributions for patients hospitalised with COVID-19 using a large dataset ($N = 21{,}000 - 157{,}000$) from the Brazilian Sistema de Informação de Vigilância Epidemiológica da Gripe database. A joint Bayesian subnational model with partial pooling is used to simultaneously describe the 26 states and one federal district of Brazil and shows significant variation in the mean of the symptom-onset-to-death time, with ranges between $11.2 - 17.8$ days across the different states, and a mean of 15.2 days for Brazil. We find strong evidence in favour of specific probability density function choices: for example, the gamma distribution gives the best fit for onset-to-death and the generalised log-normal for onset-to-hospital-admission. Our results show that epidemiological distributions have considerable geographical variation, and provide the first estimates of these distributions in a low and middle-income setting.

3.1 Introduction

Surveillance of COVID-19 has progressed from initial reports on 31st December 2019 of pneumonia with unknown etiology in Wuhan, China [World Health Organisation, 2020a], to the confirmation of 9,826 cases of SARS-CoV-2 across 20 countries one month later [World Health Organisation, 2020b], to the current pandemic of greater than 28 million confirmed cases and 900,000 deaths globally to date[1] [World Health Organisation, 2020c]. Early estimates of epidemiological distributions provided critical input that enabled modelling to identify the severity and infectiousness of the disease. The onset-to-death distribution [Donnelly et al., 2003, Garske et al., 2009], characterising the range of times observed between the onset of first symptoms in a patient and their death, proved crucial in early estimates of the Infection Fatality Ratio (IFR) where it was used to estimate the cumulative number of deaths at the beginning of the epidemic in Wuhan [Verity et al., 2020]. Similarly, the onset-to-death distribution was used in recent approaches to modelling the transmission dynamics of SARS-CoV-2 to estimate the reproduction number R_t and other important epidemiological quantities such as serial interval distribution [Dana et al., 2020, Flaxman et al., 2020, Jombart et al., 2020, Linton et al., 2020, Wu et al., 2020a-c].

Initial estimates of COVID-19 epidemiological distributions necessarily relied on relatively few data points, with the events comprising these distributions occurring in a period of time that was short compared to the temporal pathologies of the disease progression, resulting in wide confidence or credible intervals and a sensitivity to time-series censoring effects [Verity et al., 2020]. Global surveillance of the disease over the past 197 days has provided more data to re-evaluate the time-delay distributions of the disease. In particular, public availability of a large number of patient-level hospital records – over 390,000 in total at the time of writing – from the SIVEP-Gripe (*Sistema de Informação de Vigilância Epidemiológica da Gripe*) database published by Brazil's Ministry of Health [Ministério da Saúde, 2020], provides an opportunity to make robust statistical estimates of the onset-to-death and other time-delay distributions such as onset-to-diagnosis, length of ICU stay, onset-to-hospital-admission, onset-to-hospital-discharge, onset-to-ICU-admission, and hospital-admission-to-death.

[1] Referring to the date of the article publication 25th Nov 2020.

3.1.1 Our contributions

In this work, we fit and present an analysis of these epidemiological distributions, with the chapter set out as follows. Section 3.2 describes the SIVEP-Gripe database [Ministério da Saúde, 2020], from which the data were obtained, and the methodological approach applied to fit distributions using a hierarchical Bayesian model with partial pooling. Section 3.3 describes the results from fitting epidemiological distributions at the national- and state-levels to a range of probability density functions (PDFs). The results are discussed in Section 3.4, including associations with socioeconomic factors, such as education, segregation, and poverty, and conclusions are given in Section 3.5.

3.2 Methods

3.2.1 Data

The SIVEP-Gripe database provides detailed patient-level records for all individuals hospitalised with severe acute respiratory illness, including all suspected or confirmed cases of severe COVID-19 reported by both private and public sector healthcare institutions, from small rural hospitals to large metropolitan academic centres [Baqui et al., 2020, Bastos et al., 2020, de Souza et al., 2020, Ministério da Saúde, 2020, Niquini et al., 2020]. The records include the dates of admission, onset of symptoms and outcome (death or discharge) and state where the patient lives and where they are being treated, among other diagnosis-related variables. We extracted the data for confirmed COVID-19 records starting on 25$^{\text{th}}$ February 2020 and considered records in our analysis ending on 7$^{\text{th}}$ July 2020. The dataset was filtered to obtain rows for onset-to-death, hospital-admission-to-death, length of ICU stay, onset-to-hospital-admission, onset-to-hospital-discharge, onset-to-ICU-admission and onset-to-diagnosis. Onset-to-diagnosis data were split into the diagnoses confirmed by PCR and those confirmed by other methods, such as rapid antibody and antigen tests, called non-PCR throughout this chapter. Entries resulting in distribution times greater than 133 days were considered a typing error and removed, as the first recorded COVID-19 case in Brazil was on 25$^{\text{th}}$ February 2020 [The Lancet, 2020].

Additional filtering of the data was applied for onset-to-ICU-admission, onset-to-hospital-admission and onset-to-death in order to eliminate bias introduced by potentially erroneous entries identified in the data

for these distributions. We removed the rows where admission to the hospital or ICU or death happened on the same day as the onset of symptoms, assuming that these were incorrectly inputted entries. The decision to test removing the first day is motivated firstly by the observation of several conspicuous data entry errors in the database, and secondly by anomalous spikes corresponding to same-day events observed in these distributions. An example of anomalous spikes in the onset-to-death distribution is shown in Fig. 3.1 for selected states.

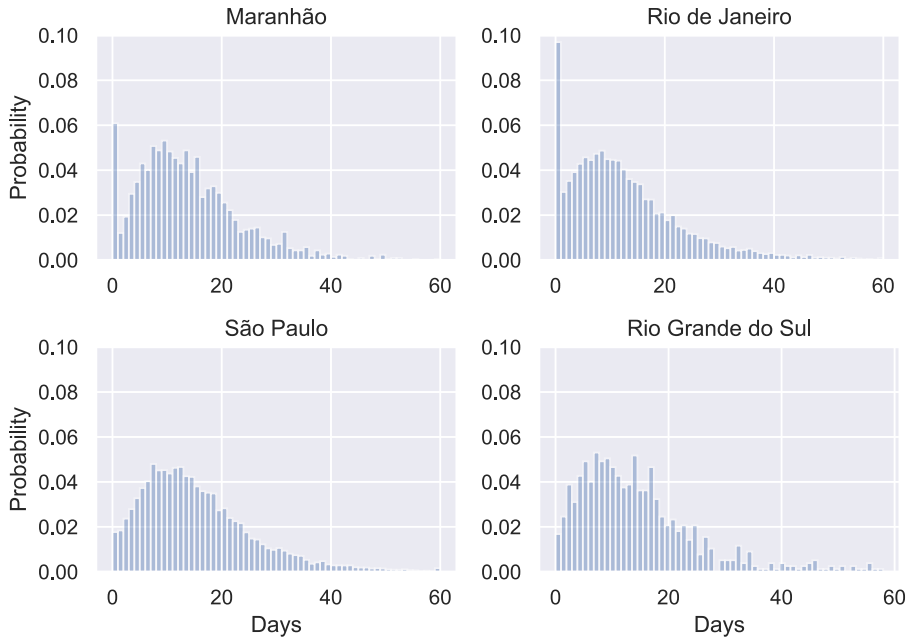


Figure 3.1: Distribution of onset-to-death for Maranhão, Rio de Janeiro, São Paulo and Rio Grande do Sul. Anomalous spikes for the first day can be observed for Maranhão and Rio de Janeiro, indicating they might be a reporting error.

Sensitivity analyses on data inclusion, regarding the removal of anomalous spikes in first-day data indicative of reporting errors (e.g. onset to hospital admission), and regarding the sensitivity of the dataset to time-series censoring effects, are set out in the Results Section 3.3.3.

A summary of the data, including the number and range of samples per variable from the SIVEP-Gripe dataset is given in Table 3.1. The age-sex structure of hospitalised patients in the database with confirmed COVID-19 diagnoses is presented in Fig. 3.2. A breakdown of the number of data samples per state is provided in Table 3.2.

Table 3.1: Summary of the distribution data extracted from SIVEP-Gripe database [Ministério da Saúde, 2020]. Numbers of samples (N_{samples}) are given for the whole country.

Distribution	N_{samples}	Range (days)
Onset-to-death	59,271	1-114
Hospital-admission-to-death	52,821	0-99
ICU-stay	21,709	0-89
Onset-to-hospital-admission	141,618	1-129
Onset-to-hospital-discharge	69,478	0-120
Onset-to-ICU-admission	46,617	0-101
Onset-to-diagnosis (PCR)	156,558	0-129
Onset-to-diagnosis (non-PCR)	19,438	0-102

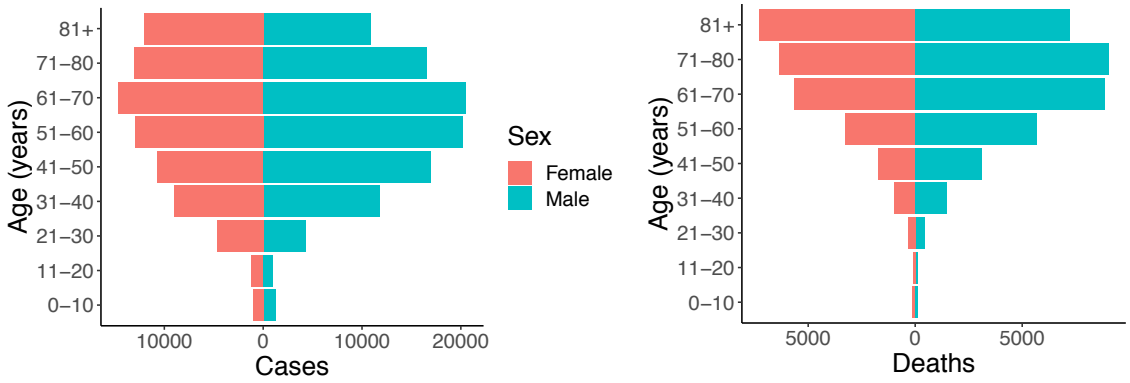


Figure 3.2: Demography of COVID-19 patients in Brazil. The left plot shows the number of confirmed COVID-19 cases and the right shows the number of confirmed COVID-19 deaths. data were extracted from the SIVEP-Gripe database from 25$^{\text{th}}$ Feb 2020 up to 7$^{\text{th}}$ July 2020. [Ministério da Saúde, 2020]

Table 3.2: Number of data points per state for each of the datasets analysed in the study. Abbreviations are explained in Table 3.8.

	Onset-death	Admission-Death	ICU-stay	Onset-hospital admission	Onset-hospital discharge	Onset-ICU admission	Onset-diagnosis (PCR)	Onset-diagnosis (non-PCR)
AC	239	115	2	225	4	9	345	1
AL	1040	894	680	1600	629	859	1344	416
AM	2736	2403	1010	5971	2573	1323	4502	1604
AP	181	175	68	299	136	80	183	153
BA	2241	2013	982	4563	1338	2300	5266	352
CE	5801	4905	1534	9685	4536	2768	8286	1749
DF	662	655	499	2687	1415	1198	2864	311
ES	1292	1023	589	1409	507	778	1774	321
GO	698	637	375	1813	783	819	2018	122
MA	1950	1097	197	1485	247	341	1562	821
MG	1223	1176	603	4782	2210	1521	4910	604
MS	131	124	46	723	417	171	764	126
MT	286	248	83	1347	2191	384	4695	2175
PA	4727	3934	1270	8226	3034	1993	6921	1351
PB	1136	1037	349	1992	508	740	1584	644
PE	4408	3284	311	6574	1888	1566	9745	190
PI	515	497	139	2161	341	490	2314	240
PR	793	773	898	3174	1952	1168	3490	124
RJ	9750	9068	1490	18019	7438	7165	21159	1446
RN	876	821	337	1878	664	693	1517	544
RO	254	238	180	554	180	284	488	293
RR	270	265	53	98	51	56	200	92
RS	790	770	971	3565	2328	1277	4144	477
SC	408	389	291	1600	777	599	1634	343
SE	303	295	193	938	181	306	1116	117
SP	16348	15808	8515	55735	32937	17642	63184	4769
TO	213	177	44	515	213	87	549	53

3.2.2 Model fitting

Gamma, Weibull, log-normal, generalised log-normal [Singh et al., 2012], and generalised gamma [Stacy, 1962] PDFs are fitted to several epidemiological distributions, with the specific parameterisations provided in Tables 3.3 and their means and variances in Table 3.4.

Table 3.3: Probability density functions. y denotes the data, $\Gamma(\cdot)$ is a gamma function. GG – generalised gamma [Stacy, 1962], GLN – generalised log-normal [Singh et al., 2012].

PDF
$\text{Gamma}(y\|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1}\exp(-\beta y)$
$\text{Log-normal}(y\|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{y}\exp\left(-\frac{1}{2}\left(\frac{\log y-\mu}{\sigma}\right)^2\right)$
$\text{GG}(y\|a,d,p) = \frac{1}{\Gamma\left(\frac{d}{p}\right)}\left(\frac{p}{a}\right)^d x^{d-1}\exp\left(-\left(\frac{y}{a}\right)^p\right)$
$\text{GLN}(y\|\mu,\sigma,s) = \frac{1}{y}\frac{s}{2^{\frac{s+1}{s}}\sigma\Gamma\left(\frac{1}{s}\right)}\exp\left(-\frac{1}{2}\lvert\frac{\log y-\mu}{\sigma}\rvert^s\right)$

Table 3.4: Probability density functions with analytical formulae for mean and variance. y denotes the data, $\Gamma(\cdot)$ is a gamma function. GG – generalised gamma [Stacy, 1962], GLN – generalised log-normal [Singh et al., 2012].

PDF	Mean	Variance
$\text{Gamma}(y\|\alpha,\beta)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
$\text{Weibull}(y\|\alpha,\sigma)$	$\sigma\Gamma\left(1+\frac{1}{\alpha}\right)$	$\sigma^2\left(\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right)$
$\text{Log-normal}(y\|\mu,\sigma)$	$\exp\left(\mu+\frac{\sigma^2}{2}\right)$	$\left(\exp\left(\sigma^2\right)-1\right)\exp\left(2\mu+\sigma^2\right)$
$\text{GG}(y\|a,d,p)$	$a\frac{\Gamma((d+1)/p)}{\Gamma(d/p)}$	$a^2\left[\frac{\Gamma((d+2)/p)}{\Gamma(d/p)}-\left(\frac{\Gamma((d+2)/p)}{\Gamma(d/p)}\right)^2\right]$
$\text{GLN}(y\|\mu,\sigma,s)$	$\exp(\mu)\left[1+\frac{S}{2\Gamma(1/s)}\right]$, $S=\sum_{j=1}^{\infty}\sigma^j\left(1+(-1)^j\right)2^{j/s}\frac{\Gamma\left(\frac{j+1}{s}\right)}{\Gamma(j+1)}$	$\exp(2\mu)\left[1+\frac{S}{2\Gamma(1/s)}\right]$-$[\text{Mean}]^2$, $S=\sum_{j=1}^{\infty}2\sigma^j\left(1+(-1)^j\right)2^{j/s}\frac{\Gamma\left(\frac{j+1}{s}\right)}{\Gamma(j+1)}$

The parameters of each distribution are fitted in a joint Bayesian hierarchical model with partial pooling, using data from the 26 states and one federal district of Brazil, extracted and filtered to identify specific epidemiological distributions such as onset-to-death, ICU-stay, and so on.

As an example consider fitting a gamma PDF for the onset-to-death distribution. The gamma distribution for the i^{th} state is given by

$$\text{Gamma}(\alpha_i,\beta_i)\,, \tag{3.1}$$

where shape and scale parameters are assumed to be positively constrained, normally distributed random

variables

$$\alpha_i \sim N(\alpha_{\text{Brazil}}, \sigma_1) \tag{3.2}$$

and

$$\beta_i \sim N(\beta_{\text{Brazil}}, \sigma_2)\,. \tag{3.3}$$

The parameters α_{Brazil} and β_{Brazil} denote the national level estimates, and

$$\sigma_1 \sim N^+(0,1)\,,\ \sigma_2 \sim N^+(0,1)\ , \tag{3.4}$$

where $N^+(\cdot)$ is a truncated normal distribution. In this case, parameters α_{Brazil} and β_{Brazil} are estimated by fitting a gamma PDF to the fully pooled data, that is including the observations for all states. Prior probabilities for the national level parameters for each of the considered PDFs are chosen to be $N^+(0,1)$. The only exception was for the more complex generalised gamma distribution which used more informed priors to speed up fitting. The priors for the generalised gamma distribution were chosen based on the previous fits to be: $\mu_{\text{Brazil}} \sim N^+(2, 0.5)$, $\sigma_{\text{Brazil}} \sim N^+(0.5, 0.5)$ and $s_{\text{Brazil}} \sim N^+(1.5, 0.5)$. Additionally, for all fitted densities, the mean and variance parameters were constrained to be positive.

Posterior samples of the parameters in the model are generated using Hamiltonian Monte Carlo (HMC) with Stan [Carpenter et al., 2017, Hoffman and Gelman, 2014]. For each fit, we use 4 chains and $2,000$ iterations, with half of the iterations dedicated to warm-up.

The preference for one fitted model over another is characterised in terms of the Bayesian support, with the model evidence calculated to see how well a given model fits the data and comparison between two models using Bayes factors (BFs). BFs provide a principled fully Bayesian approach to select between models, incorporating the full posterior densities and thus also the uncertainty of each of the parameters instead of point estimates [Jefferys and Berger, 1992, Jeffreys, 1965, Sivia and Skilling, 2006, Smith and Spiegelhalter, 1980]. Moreover, BFs naturally balance the complexity and accuracy of the compared models, ensuring that the excessively complex models are not automatically favoured. Historically simpler methods have been favoured, as BFs can be costly to compute for complex models, however using recent efficient methods this is not an issue (see Hawryluk et al., 2023, or Chapter 2). The details of how to estimate the model evidence and calculate the Bayes factors for each pair of models are given in the previous Chapter 2.

Data cleaning and the analysis of the results were conducted using Python (version 3.7.7) programming

language [Python Software Foundation]. PyStan (version 2.19.0.0) interface was used for running model fitting with Stan [Stan Development Team]. The code and data for this analysis are available online at `https://github.com/mrc-ide/Brazil_COVID19_distributions`.

3.3 Results

3.3.1 Brazilian epidemiological distributions

Five trial PDFs – gamma, Weibull, log-normal, generalised log-normal and generalised gamma – were fitted to the epidemiological data shown in Fig. 3.3.

All of the models' fits were tested by using the Bayes factors based on the Laplace approximation and corrected using referenced thermodynamic integration [Gelman and Meng, 1998, Hawryluk et al., 2023, Meng and Wong, 1996a], as described in Chapter 2. The thermodynamic integration contribution was negligible suggesting the posterior distributions are satisfactorily approximated as multivariate normal. The conclusions on the preferred PDF were not sensitive to the choice of prior distributions, that is the preferred model was still the favoured one even when more informative prior distributions were applied for all PDFs. The Bayes factors used for model selection are shown in Table 3.5. The discrepancy between the best models selected by the Bayes factors shown in Table 3.5 and the visual assessment of the fits in Fig. 3.3 is due to the BFs calculated for the full hierarchical models including the subnational partial-pooling, whereas the figure shows fits only on the national level.

Table 3.5: Bayes factors (BFs) for the analysed distributions and models. For each distribution (rows), the values represent BF for the best-fitting model against other models. A value of 0 indicates the model that fits the data the best. Value > 10 indicates very strong evidence against the given model compared to the best one. GLN - generalised log-normal, GG - generalised gamma. NA - not analysed. The BF values are reported here as $2\log(B_{ij})$ following the notation from Kass and Raftery [1995].

Distribution	Gamma	Weibull	Log-normal	GLN	GG
Onset-to-death	0	2156	2208	198	301
Admission-death	0	195	4349	3096	188
ICU stay	0	231	588	607	352
Onset-to-hospital-admission	4000	17073	494	0	NA
Onset-to-hospital-discharge	2819	8346	6079	0	3087
Onset-to-ICU-admission	798	4359	142	0	1244
Onset-to-diagnosis (PCR)	1111	10400	13882	0	1257
Onset-to-diagnosis (non-PCR)	578	793	4340	0	461

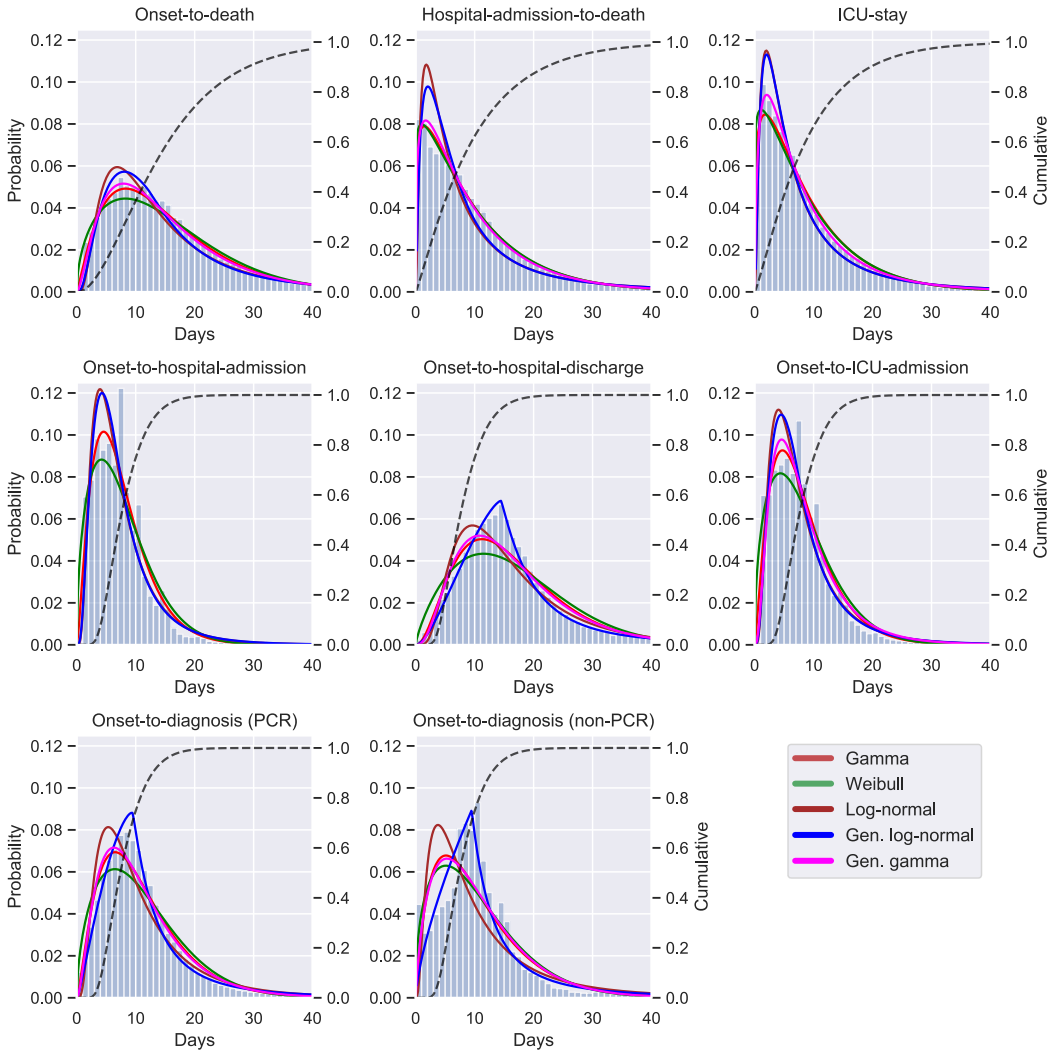


Figure 3.3: Histograms for onset-to-death, hospital-admission-to-death, ICU-stay, onset-to-hospital-admission, onset-to-hospital-discharge, onset-to-ICU-admission onset-to-diagnosis (PCR) and onset-to-diagnosis (non-PCR) distributions show data for Brazil extracted from the SIVEP-Gripe database Ministério da Saúde, 2020. For each distribution, solid lines are for fitted PDFs and the dashed line shows the cumulative distribution function of the best-fitting PDF. The left-hand side y-axis gives the probability value for the PDFs and the right-hand side y-axis shows the value for the cumulative distribution function. All values on the x-axes are given in days. State-level fits are shown in Figs. 3.4, 3.6 and 3.7

The gamma PDF provided the best fit for the onset-to-death, hospital-admission-to-death and ICU-stay data. For the remaining distributions – onset-to-diagnosis (non-PCR), onset-to-diagnosis (PCR), onset-to-hospital-discharge, onset-to-hospital-admission and onset-to-ICU-admission – the generalised log-normal distribution was the preferred model. The list of preferred PDFs for each distribution, the estimated mean, variance and PDFs' parameter values for the national fits are given in Table 3.6. Due to the large volume of data, the 95% credible intervals (CrI) for parameters of each of the preferred PDFs was less than 0.1

wide, therefore in Table 3.6 we show only point estimates.

Table 3.6: For each COVID-19 distribution the preferred PDF with the largest Bayesian support is listed, along with the estimated mean, variance and other parameters of the PDF. 95% credible intervals are given in brackets for mean and variance. The parameters p_1, p_2 and p_3 for the preferred PDFs gamma and generalised log-normal (GLN) are given in the form $Gamma(x|p_1, p_2) = Gamma(\alpha, \beta)$ and $GLN(x|p_1, p_2, p_3) = GLN(\mu, \sigma, s)$, with the formulae of the PDFs given in Table 3.3. The credible intervals for parameters p_1, p_2 and p_3 are less than 0.1 wide, so only the point estimates are shown. † The variance diverges for the onset-to-diagnosis (non-PCR) PDF.

Distribution	Preferred PDF	Mean (days)	Variance (days^2)	p_1	p_2	p_3
Onset-to-death	Gamma	15.2 (15.1, 15.3)	105.3 (103.7, 106.9)	2.2	0.1	-
Hospital-admission-to-death	Gamma	10.0 (9.9, 10.0)	84.8 (83.2, 86.4)	1.2	0.1	-
ICU-stay	Gamma	9.0 (8.9, 9.1)	64.9 (63.1, 66.8)	1.2	0.1	-
Onset-to-hospital-admission	Gen. log-normal	7.8 (7.7, 7.8)	35.7 (35.0, 36.5)	1.8	0.6	1.8
Onset-to-hospital-discharge	Gen. log-normal	17.6 (17.6, 17.7)	248.7 (233.7, 265.6)	2.7	0.3	1.2
Onset-to-ICU-admission	Gen. log-normal	8.5 (8.4, 8.5)	48.0 (46.1, 50.0)	1.9	0.6	1.8
Onset-to-diagnosis (PCR)	Gen. log-normal	12.5 (12.5, 12.6)	252.3 (236.4, 269.6)	2.3	0.3	1.2
Onset-to-diagnosis (non-PCR)	Gen. log-normal	14.5 (14.3, 14.7)	†	2.3	0.3	1.0

Additionally, in Fig. 3.3, in each instance the cumulative probability distribution is given for the best model fit, revealing that out of patients for whom COVID-19 is terminal, almost 70% die within 20 days of symptom onset. Out of patients who die in the hospital, nearly 60% die within the first 10 days since admission.

The estimated mean number of days for each distribution for Brazil is compared in Table 3.7 with values found in the literature for China, the US and France. The majority of the data obtained through searching the literature pertained to the early stages of the epidemic in China, and no data were found for low- and middle-income countries. The mean onset-to-death time of 15.2 (95% CrI $15.1 - 15.3$) days, from a best-fitting gamma PDF, is shorter than the 17.8 (95% CrI $16.9 - 19.2$) days estimate from Verity et al. [2020] and 20.2 (95% CrI $15.1 - 29.5$) days estimate (14.5 days without truncation) from Linton et al. [2020]. In both cases, estimates were based on a small sample size from the beginning of the epidemic in China. The mean number of days for hospital-admission-to-death of 10.8 (95% CrI $10.7 - 10.9$) for Brazil closely matches the 10 days estimated by Salje et al. [2020b].

Table 3.7: Epidemiological distributions for COVID-19 for Brazil, China, France and US. PDF means for Brazil have been obtained using MCMC sampling, using the PDF with the maximum Bayesian support for each data distribution (see Table 3.5). For China, France and the US the values have been obtained from the literature. All values are given in days, and 95% CrI are given in brackets unless stated otherwise. * adjusted for censoring, † PCR confirmed, ‡ non-PCR confirmed, [a] median (interquartile range), [b] mean (standard deviation).

Distribution	Brazil	China	France	US
Onset-to-death	15.2 (15.1, 15.3) 16.0* (15.9, 16.1)	17.8 (16.9, 19.2) Verity et al. 2020 18.8* (15.7, 49.7) Verity et al. 2020 14.5 (12.5, 17.0) Linton et al. 2020 20.2* (15.1, 29.5) Linton et al. 2020		13.59[b] (7.85) Abdollahi et al. 2020
Hospital-admission-to-death	10.0 (9.9, 10.0) 10.8* (10.7, 10.9)	5.0[a] (3.0, 9.3) Chen et al. 2020 8.9 (7.3-10.4) Linton et al. 2020 13.0* (8.7-20.9) Linton et al. 2020	10.0 Salje et al. 2020b	
ICU-stay	9.0 (8.9, 9.1) 10.1* (9.9, 10.2)	8.0[a] (4.0, 12.0) Zhou et al. 2020	17.6 (17.0, 18.2) Salje et al. 2020b	
Onset-to-hospital-admission	7.8 (7.7, 7.8)	10.0[a] (7.0-12.0) Chen et al. 2020		
Onset-to-hospital-discharge	17.6 (17.6, 17.7)	22.0[a] (18.0, 25.0) Zhou et al. 2020		
Onset-to-ICU-admission	8.5 (8.4, 8.5)	9.5[a] (7.0, 12.5) Yang et al. 2020		
Onset-to-diagnosis	12.5†(12.5, 12.6) 14.5‡(14.3, 14.7)	5.5 (5.4, 5.7) Li et al. 2020a		

3.3.2 Subnational Brazilian epidemiological distributions

The onset-to-death distribution, and other time-delay distributions such as onset-to-diagnosis, length of ICU stay, onset-to-hospital-admission, onset-to-hospital-discharge, onset-to-ICU-admission, and hospital-admission-to-death, have been fitted in a joint model across the 26 states and one federal district of Brazil using partial pooling.

The explanation of the abbreviations for the states of Brazil used throughout the figures and tables in this section is given in Table 3.8.

Table 3.8: Brazilian States and Regions

Abbreviation	**State**	**Region**
AC	Acre	North
AP	Amapá	North
AM	Amazonas	North
PA	Pará	North
RO	Rondônia	North
RR	Roraima	North
TO	Tocantins	North
AL	Alagoas	Northeast
BA	Bahia	Northeast
CE	Ceará	Northeast
MA	Maranhão	Northeast
PB	Paraíba	Northeast
PE	Pernambuco	Northeast
PI	Piauí	Northeast
RN	Rio Grande do Norte	Northeast
SE	Sergipe	Northeast
DF	Distrito Federal	Central-West
GO	Goiás	Central-West
MT	Mato Grosso	Central-West
MS	Mato Grosso do Sul	Central-West
ES	Espírito Santo	Southeast
MG	Minas Gerais	Southeast
RJ	Rio de Janeiro	Southeast
SP	São Paulo	Southeast
PR	Paraná	South
RS	Rio Grande do Sul	South
SC	Santa Catarina	South

The mean number of days, plotted in Fig. 3.4, shows substantial subnational variability – for example the mean onset-to-hospital-admission for Amazonas state was estimated to be 9.9 days (95% CrI $9.7-10.1$), whereas for Mato Grosso do Sul the estimate was 6.7 (95% CrI $6.4-7.1$) days and Rio de Janeiro - 7.2 days

(95% CrI 7.1 − 7.3). Amazonas state had the longest average time from onset-to-hospital-admission and ICU-admission. The state with the shortest average onset-to-death time was Roraima. Santa Catarina state on the other hand had the longest average onset-to-death and hospital-admission-to-death time, as well as the longest average ICU-stay.

We also observe discrepancies between the five geographical regions of Brazil. For example, states belonging to the southern part of the country (Paraná, Rio Grande do Sul and Santa Catarina) had a longer average ICU-stay and hospital-admission-to-death time as compared to the states in the North region. Full results, including detailed estimates of mean, variance, and estimates for each of the distributions' parameters for Brazil and Brazilian states can be accessed at `https://github.com/mrc-ide/Brazil_COVID19_distributions/blob/master/results/results_full_table.csv`.

For a visualisation of the uncertainty in our mean estimates for each state, see the posterior density plots in Figs. 3.6 and 3.7. Additional national and state-level results for the onset-to-death gamma PDF, including the posterior plots for mean and variance, are shown in Fig. 3.5. The values of the posterior mean and variance are given in Table 3.9.

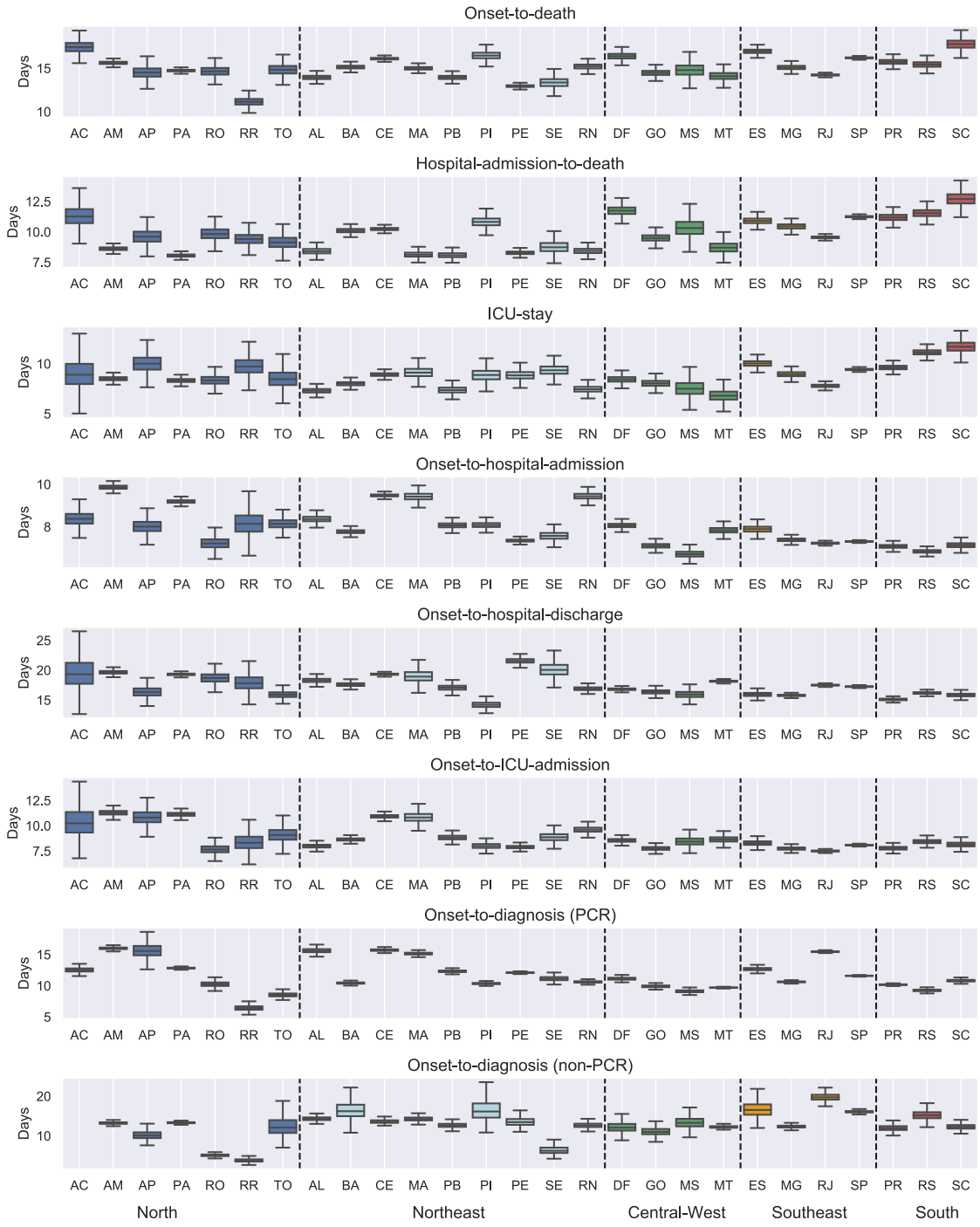


Figure 3.4: Estimates of the mean time in days for onset-to-death, hospital-admission-to-death and each of the other distributions fitted in the joint model of Brazil. Estimates are grouped by the five regions of Brazil, North (blue), Northeast (light blue), Central-West (green), Southeast (orange), and South (red). For state Acre (AC), the onset-to-diagnosis (non-PCR) mean diverged due to the small number of samples (n=1). Abbreviations are explained in Table 3.8.

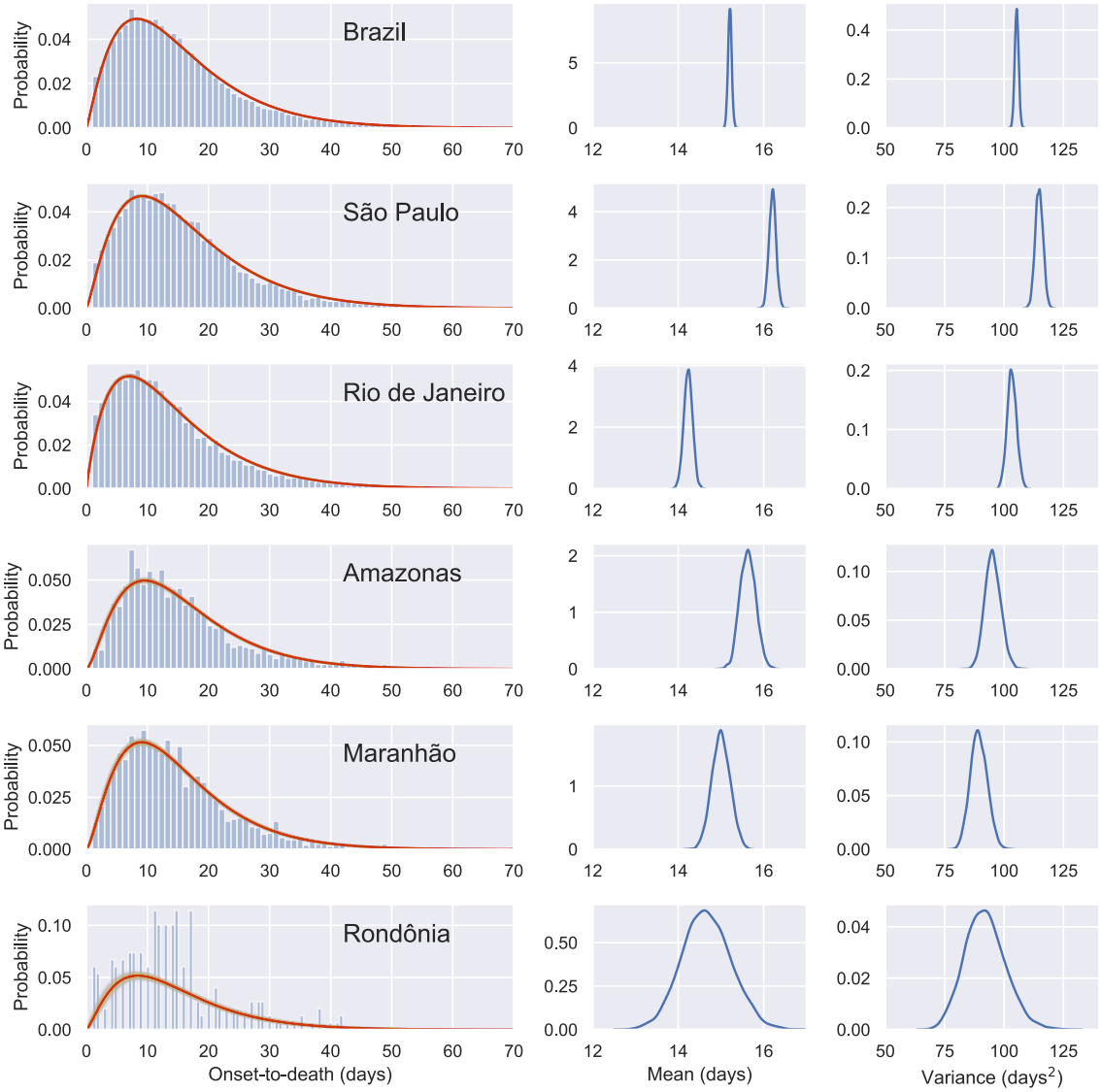


Figure 3.5: Gamma PDF $Gamma(\alpha, \beta)$ fitted to the onset-to-death data for Brazil and five states of Brazil. The PDFs were fitted with HMC partially pooling each state with the whole country. The red lines represent the model using the mean parameter estimates. Individual PDFs selected during MCMC sampling are shown in yellow. Posterior mean and variance distributions for each region are given in the middle and right-hand side columns.

Table 3.9: State-level onset-to-death estimates for gamma PDF: mean, variance, parameters values, with 95% confidence intervals. The parameters p_1 and p_2 are given in the form $Gamma(x|p_1, p_2) = Gamma(\alpha, \beta)$. The full PDFs for other distributions are available at https://github.com/mrc-ide/Brazil_COVID19_distributions/blob/master/results/results_full_table.csv. Abbreviations are explained in Table 3.8.

State	Mean (days)	Variance (days^2)	p_1	p_2
AC	17.4 (16.1, 18.8)	119.4 (98.8, 143.6)	2.6 (2.2, 2.9)	0.1 (0.1, 0.2)
AL	14.0 (13.4, 14.5)	82.5 (74.3, 91.9)	2.4 (2.2, 2.5)	0.2 (0.2, 0.2)
AM	15.6 (15.3, 16.0)	95.3 (89.1, 102.1)	2.6 (2.4, 2.7)	0.2 (0.2, 0.2)
AP	14.5 (13.2, 16.0)	99.1 (79.8, 122.7)	2.1 (1.9, 2.4)	0.1 (0.1, 0.2)
BA	15.1 (14.7, 15.6)	116.6 (107.9, 126.1)	2.0 (1.9, 2.1)	0.1 (0.1, 0.1)
CE	16.1 (15.8, 16.4)	116.4 (111.1, 122.0)	2.2 (2.2, 2.3)	0.1 (0.1, 0.1)
DF	16.4 (15.6, 17.2)	105.0 (92.7, 119.0)	2.6 (2.3, 2.8)	0.2 (0.1, 0.2)
ES	17.0 (16.4, 17.5)	107.8 (98.2, 118.1)	2.7 (2.5, 2.9)	0.2 (0.1, 0.2)
GO	14.5 (13.8, 15.2)	87.9 (77.9, 99.1)	2.4 (2.2, 2.6)	0.2 (0.2, 0.2)
MA	15.0 (14.6, 15.4)	89.4 (82.7, 96.5)	2.5 (2.4, 2.7)	0.2 (0.2, 0.2)
MG	15.1 (14.6, 15.7)	95.1 (86.3, 104.7)	2.4 (2.2, 2.6)	0.2 (0.1, 0.2)
MS	14.8 (13.3, 16.4)	93.9 (74.8, 116.8)	2.4 (2.0, 2.7)	0.2 (0.1, 0.2)
MT	14.1 (13.1, 15.1)	80.6 (67.2, 96.4)	2.5 (2.2, 2.8)	0.2 (0.2, 0.2)
PA	14.7 (14.5, 15.0)	90.2 (85.7, 94.9)	2.4 (2.3, 2.5)	0.2 (0.2, 0.2)
PB	14.0 (13.4, 14.5)	78.7 (71.2, 87.3)	2.5 (2.3, 2.7)	0.2 (0.2, 0.2)
PE	13.0 (12.7, 13.2)	89.7 (84.6, 95.1)	1.9 (1.8, 1.9)	0.1 (0.1, 0.2)
PI	16.5 (15.6, 17.4)	114.8 (99.4, 131.7)	2.4 (2.1, 2.6)	0.1 (0.1, 0.2)
PR	15.7 (15.1, 16.4)	91.9 (81.8, 102.7)	2.7 (2.5, 2.9)	0.2 (0.2, 0.2)
RJ	14.2 (14.0, 14.4)	103.3 (99.5, 107.3)	2.0 (1.9, 2.0)	0.1 (0.1, 0.1)
RN	15.2 (14.6, 15.9)	91.9 (81.8, 103.0)	2.5 (2.3, 2.7)	0.2 (0.2, 0.2)
RO	14.7 (13.6, 15.8)	92.1 (76.4, 110.0)	2.3 (2.1, 2.6)	0.2 (0.1, 0.2)
RR	11.2 (10.2, 12.1)	68.1 (55.9, 83.0)	1.8 (1.6, 2.1)	0.2 (0.1, 0.2)
RS	15.4 (14.7, 16.2)	116.0 (103.0, 130.8)	2.1 (1.9, 2.2)	0.1 (0.1, 0.1)
SC	17.8 (16.7, 19.0)	146.8 (125.1, 173.5)	2.2 (1.9, 2.4)	0.1 (0.1, 0.1)
SE	13.4 (12.2, 14.5)	112.5 (91.4, 138.6)	1.6 (1.4, 1.8)	0.1 (0.1, 0.1)
SP	16.2 (16.0, 16.4)	114.8 (111.6, 118.0)	2.3 (2.2, 2.3)	0.1 (0.1, 0.1)
TO	14.8 (13.5, 16.2)	97.3 (79.1, 119.7)	2.3 (2.0, 2.6)	0.2 (0.1, 0.2)
Brazil	15.2 (15.1, 15.3)	105.3 (103.7, 106.9)	2.2 (2.2, 2.2)	0.1 (0.1, 0.1)

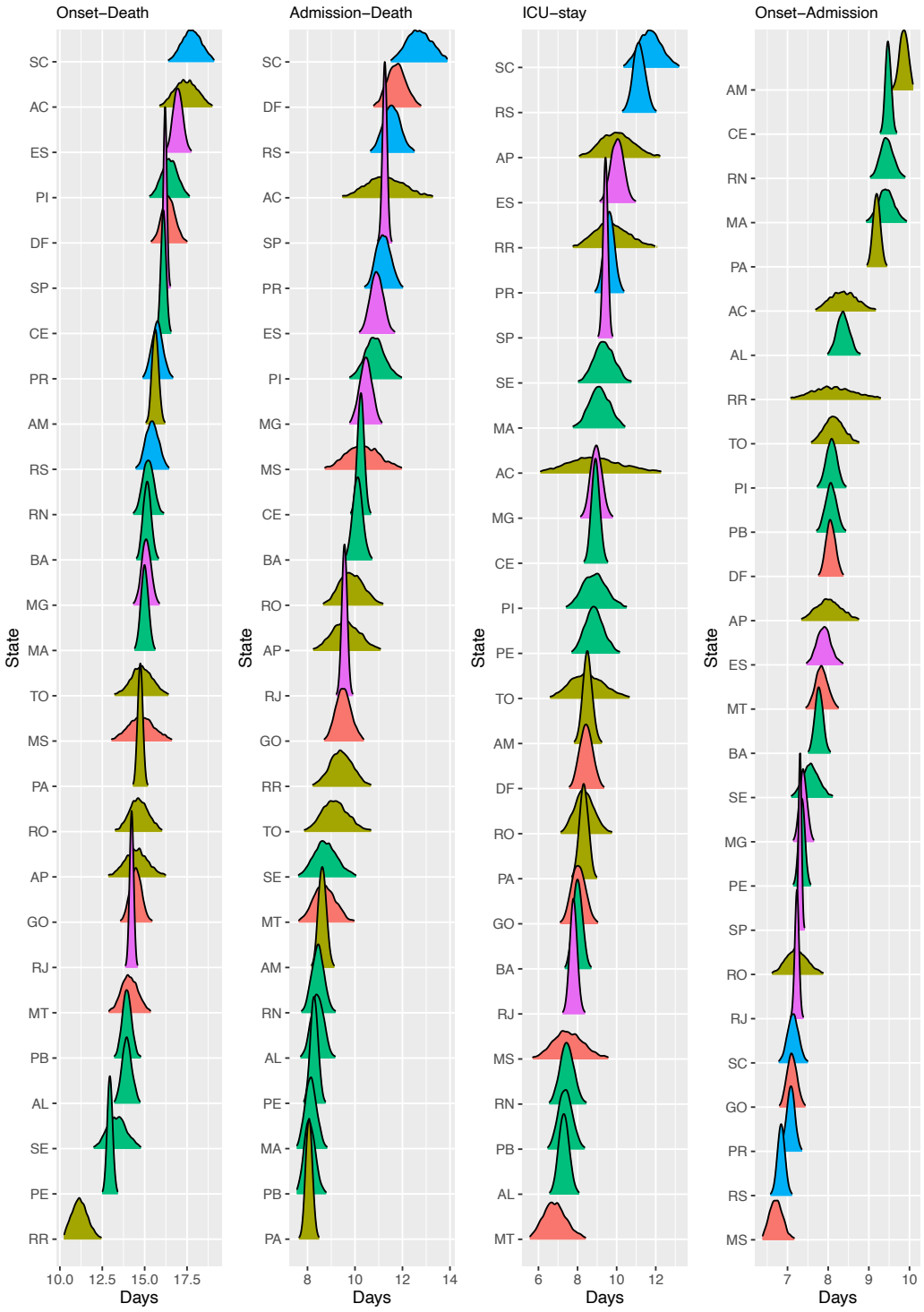


Figure 3.6: Posterior distribution of mean times (in days) for onset-to-death, hospital-admission-to-death, ICU stay and onset-to-hospital-admission, sorted by mean value. Plots are colour-coded by the geographical region to which the state belongs: North (yellow), Northeast (green), Central-West (orange), Southeast (purple), and South (blue). Abbreviations are explained in Table 3.8.

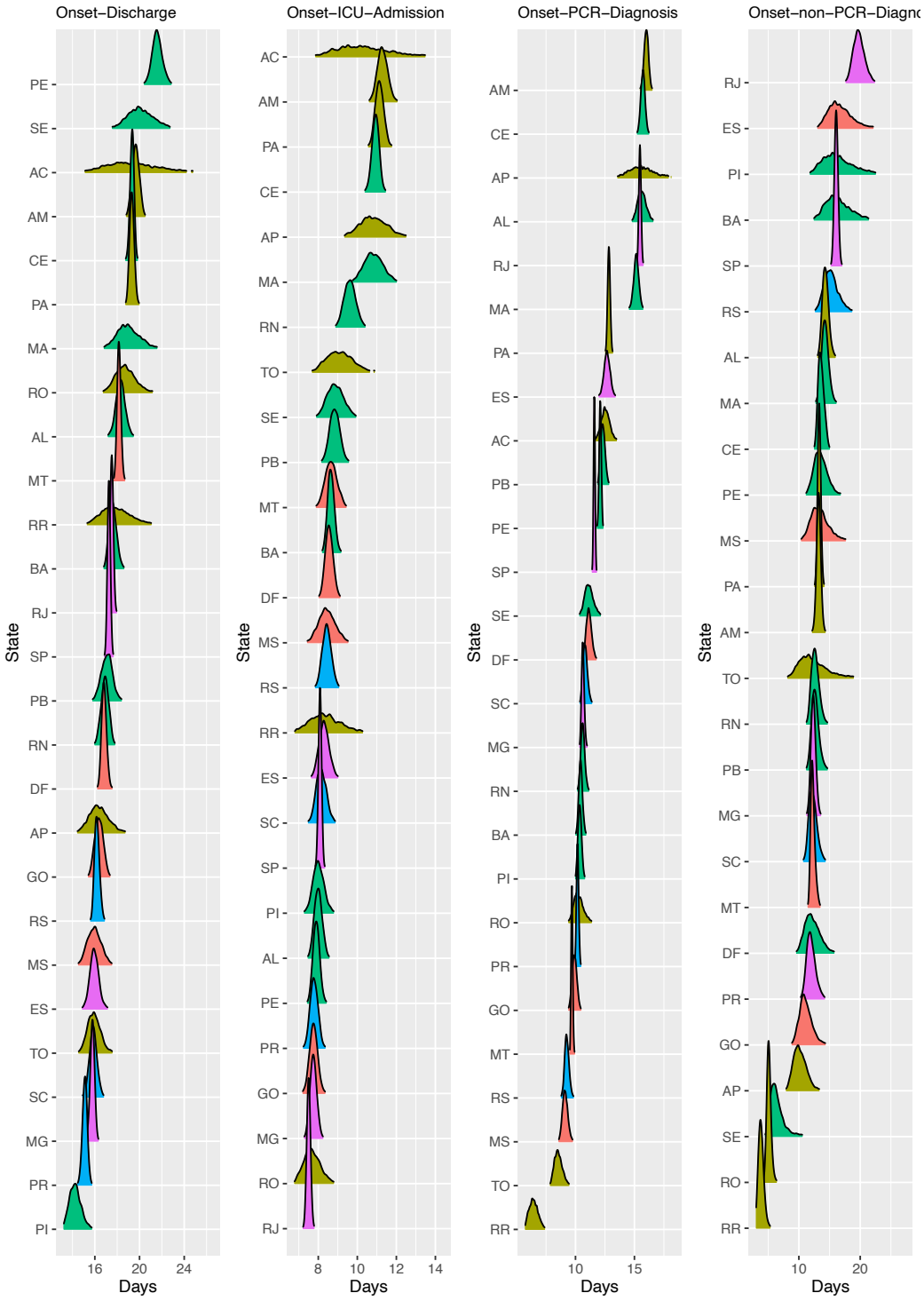


Figure 3.7: Posterior distribution of mean times (in days) for onset-to-hospital-discharge, onset-to-ICU-admission, onset-to-diagnosis (PCR) and onset-to diagnosis (non-PCR), sorted by mean value. Plots are colour-coded by the geographical region to which the state belongs: North (yellow), Northeast (green), Central-West (orange), Southeast (purple), and South (blue). Abbreviations are explained in Table 3.8

3.3.3 Sensitivity analyses

In order to remove the potential bias towards shorter outcomes from left- and right-censoring, we tested the scenario in which the data to fit the models was truncated. For example, based on a 95% quantile of 35 days for the hospital-admission-to-death distribution, entries with a starting date (hospital admission) after 2nd June 2020 and those with an end date (death) before 1st April 2020 were truncated, and the models were refitted. With censored parts of the data removed, the mean time from start to outcome increased for every distribution, e.g. for hospital-admission-to-death it increased from 10.0 days (95% CrI $9.9 - 10.0$) to 10.8 (95 % CrI $10.7 - 10.9$), and for onset-to-death it changed from 15.2 days (95% CrI $15.1 - 15.3$) to 16.0 (95% CrI $15.9 - 16.1$). The effect of truncation on censored data is given in Fig. 3.8

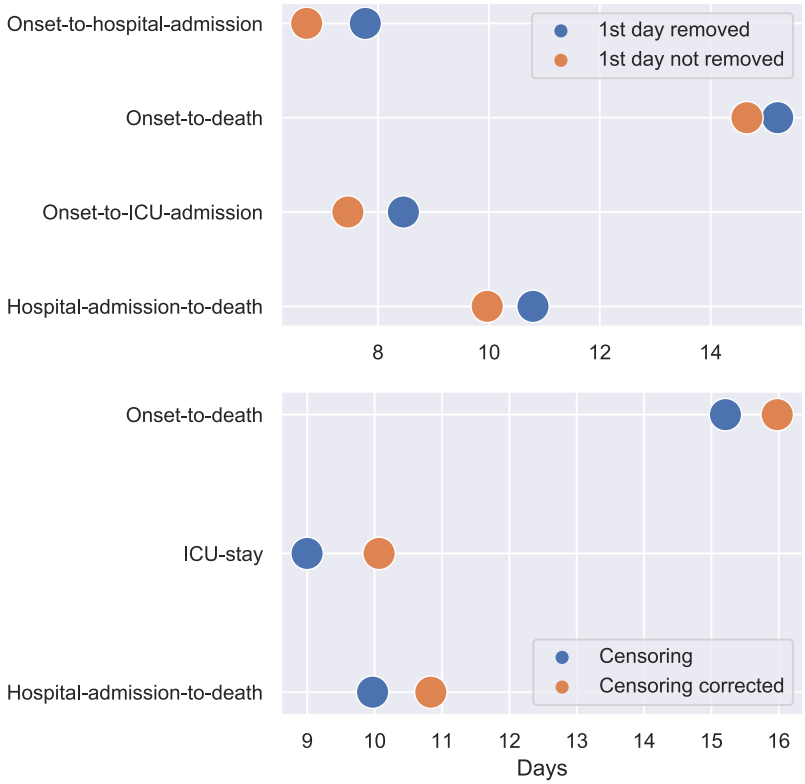


Figure 3.8: Estimated mean per distribution in different scenarios: excluding first-day data points (top) and censoring correcting (bottom). The credible intervals were not shown as due to the large amount of data available the intervals were negligible.

To test the impact of keeping or removing entries identified as potentially resulting from erroneous data transcription (see the Methods Section 3.2), we fitted the PDFs to some of the distributions on a national level with and without those entries. For onset-to-hospital-admission, onset-to-ICU and onset-to-death we

found that generalised gamma PDF was preferred when the first day of the distribution was included, and gamma (for onset-to-death) and generalised log-normal PDFs if the first day was removed. For hospital-admission-to-death, a gamma distribution fitted most accurately when the first day was included, and Weibull when it was excluded. Removing the first-day results in the mean values shifting to the right by approximately 1 day for both onset-to-hospital- and ICU-admission, and by 0.5 days for hospital-admission-to-death (see Fig. 3.8).

3.4 Discussion

Using Bayesian hierarchical models, we fitted multiple probability density functions to several epidemiological datasets, such as onset-to-death or onset-to-diagnosis, from the Brazilian SIVEP Gripe database [Ministério da Saúde, 2020].

Summary of the findings

Our findings provide the first reliable estimates of the various epidemiological distributions for the COVID-19 epidemic in Brazil and highlight a need to consider a wider set of specific parametric distributions. Instead of relying on the ubiquitous gamma or log-normal distributions, we show that often these PDFs do not best capture the behaviour of the data. For instance, the generalised log-normal is preferable for several epidemiological distributions in Table 3.6. These results can specifically inform modelling of the epidemic in Brazil [Mellan et al., 2020], and other low- and middle-income countries [Walker et al., 2020]. We expect they are also highly relevant to the epidemics unfolding in other countries.

Geographic heterogeneity

Across Brazil, the epidemic has strong geographic heterogeneity, with some states such as Amazonas and Maranhão reported to be at advanced stages [Buss et al., 2021, da Silva et al., 2020]. To accurately describe the observed differences at the subnational level, it is essential to account for variation in model parameters by state. By making use of the state-level custom-fitted onset-to-death distributions reported here, we have estimated the number of active infections on 23rd June 2020 across ten states spanning the five regions of Brazil using a Bayesian hierarchical renewal-type model [Flaxman et al., 2020, Mellan et al.,

2020, Mishra et al., 2020]. The relative change in the number of active infections from modelling the cases using heterogeneous state-specific onset-to-death distributions, compared to using a single common Brazil one, is shown in Fig. 3.9 to be substantial. we observe relative changes of up to 18% more active infections, suggesting common assumptions of onset-to-death spatial homogeneity are unreliable and closer attention is required when fitting models of SARS-CoV-2 transmission dynamics in large countries.

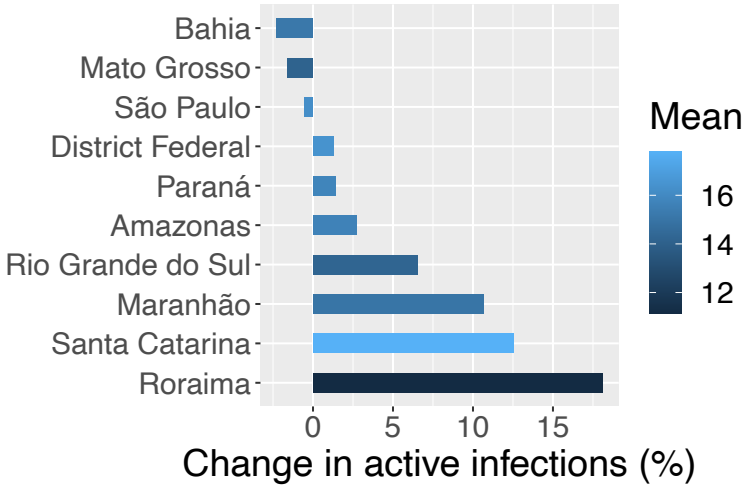


Figure 3.9: This figure shows the percentage change in active infections, estimated on the 23rd June 2020, that results from using state-specific onset-to-death distributions (see Table 3.9) compared to a single national-level one. The effect for each state is coloured according to the mean of the state's onset-to-death gamma distribution, given in days. The mean onset-to-death for Brazil is 15.2 days. Figure courtesy of Dr Thomas A. Mellan.

Notably, large subnational variability was observed for all fitted distributions, with the mean onset-to-death ranging between 11.2 days in Roraima to 17.8 in Santa Catarina. Hospital-admission-to-death time showed substantial variation between the regions of Brazil, ranging between 8.1 and 11.3 in the North, and between 9.6 and 12.8 in the South. A plausible hypothesis is that the observed differences in outcome timings could be explained by greater difficulty accessing hospitals in the North, or by a limited access to equipment such as ventilators.

Correlations with socioeconomic factors

In order to explain the origin of the geographic variation of average distribution times across states, shown in Fig. 3.4, we present a basic exploratory analysis based on relevant high-level features. We examined the correlation between socioeconomic factors, such as education, poverty, income, wealth, deprivation

and segregation, using several socioeconomic state-level indicators obtained from Barrozo et al. 2020 and additional datasets containing the mean age per state and percentage of people living in the urban areas (urbanicity) [Brazilian Institute of Geography and Statistics]. The Pearson correlation coefficients, shown in Table 3.10, suggest that poverty, income, segregation and deprivation elements were most strongly correlated with the analysed onset-time datasets. In particular, poverty was strongly negatively correlated with hospital-admission-to-death (-0.68), whereas income and segregation had a high positive correlation coefficient for the same distribution (+0.60, +0.62 respectively). The strongest correlation was observed for hospital-admission-to-death and deprivation indicator, which measures the access to sanitation, electricity and other material and non-material goods [Barrozo et al. 2020]. Interestingly, the indicators measuring economic situation were more correlated with average hospitalisation times than mean age per state. This suggests that although low- and middle-income countries typically have younger populations, their healthcare systems are more likely to struggle in response to the COVID-19 epidemic. Socioeconomic factors have also been shown to correlate with the accessibility of the COVID-19 diagnosis in the Metropolitan Region of São Paulo, which emphasises the impact of the spatial heterogeneity of the socioeconomic status on the various aspects of the epidemic, from capturing the active cases to providing treatment for the patients [de Souza et al., 2020]. More detailed analysis is necessary to fully appreciate the impact of the economic components on the COVID-19 epidemic response.

Table 3.10: Pearson correlation coefficients for mean distribution times and socioeconomic indicators. The sample size was equal to 27 (number of states).

	ICU-stay	Onset-death	Admission-death	Onset-discharge	Onset-hospital admission	Onset-ICU admission	Onset-diagnosis (PCR)
Education	-0.32	-0.25	-0.62	0.41	0.48	0.39	0.34
Poverty	-0.31	-0.31	-0.68	0.52	0.69	0.54	0.49
Deprivation	0.38	0.35	0.71	-0.49	-0.59	-0.49	-0.41
Wealth	-0.08	0.26	0.37	-0.24	-0.07	-0.21	-0.17
Income	0.21	0.28	0.60	-0.35	-0.40	-0.33	-0.35
Segregation	0.40	0.35	0.62	-0.43	-0.57	-0.47	-0.30
Mean age	0.13	0.25	0.43	-0.45	-0.57	-0.68	-0.25
Urbanicity	0.12	0.11	0.43	-0.34	-0.52	-0.40	-0.19

Spatial heterogeneity is not the only source of variability in hospitalisation times. Although in this study we did not stratify the population according to age or other demographic features, other recent studies have utilised the SIVEP-Gripe database to characterise the COVID-19 epidemic in Brazil. Namely, they looked at the regional and ethnic distribution of the hospitalised patients [Baqui et al. 2020, Niquini et al. 2020], age-sex structure and clinical characteristics such as co-morbidities and symptoms [de Souza et al.

Table 3.11: Pearson correlation coefficients for mean distribution times. The sample size was equal to 27 (number of states).

Distribution	Onset-death	Admission-death	Onset-discharge	Onset-hospital admission	Onset-ICU admission	Onset-diagnosis (PCR)
Onset-death	1	0.69	-0.35	0.06	0.24	0.15
Admission-death	0.69	1	-0.52	-0.48	-0.20	-0.36
Onset-discharge	-0.35	-0.52	1	0.39	0.43	0.40
Onset-hospital-admission	0.06	-0.48	0.39	1	0.72	0.53
Onset-ICU-admission	0.24	-0.20	0.43	0.72	1	0.50
Onset-diagnosis (PCR)	0.15	-0.36	0.40	0.53	0.50	1

2020, Niquini et al., 2020]. de Souza et al. [2020] show that 65.5% of cases are patients over 50 years old. Moreover, they also found that 84% of the patients reported having at least one underlying condition. It is clear, that both age and co-morbidities are highly correlated with adverse outcomes such as hospitalisation or death, and to calibrate the epidemiological models of COVID-19 the time-onset distributions presented in this study could be refined even further.

Limitations

In the work presented we acknowledge several limitations. The database from which distributions have been extracted, though extensive, contains transcription errors, and the degree to which these bias our estimates is largely unknown. Secondly, the fitted PDFs are based on observational hospital data and therefore should be cautiously interpreted for other settings. Thirdly, though we have fitted PDFs at subnational as well as national levels, this partition is largely arbitrary. Further work is required to understand the likely substantial effect of age, sex, ethnic variation, co-morbidities, and other factors.

Post study

The results of this study have been utilised and confirmed by studies conducted after the results presented in this chapter have been published in Hawryluk et al. [2020]. The spatial heterogeneity of the COVID-19 infections in Brazil has been further explored for example in Brizzi et al. [2022], where evidence is given for the spatial and temporal variation in the in-hospital mortality rates across Brazil. Those variations are hypothesised to be driven mainly by the shortages in healthcare capacity due to regional inequalities and increased healthcare pressure [Brizzi et al., 2022]. Additionally, in Castro et al. [2021], the authors

add political misalignment between the states of Brazil to the reason for the geographic heterogeneity and emphasise the differences in the surveillance systems in the country, which led to the epidemic spreading disproportionally and in an unmitigated way in the states with poorer surveillance.

The posteriors of the fitted densities provided input to other epidemiological models, including modelling the epidemic in the whole of Brazil [Mellan et al., 2020], investigating the epidemiology of a newly emerged SARS-COV-2 variant of concern in Amazonas [Faria et al., 2021] and COVID-19 spread elsewhere [Eales and Riley, 2024]. In [Inward et al., 2022], the authors use the PDFs that we have proposed as 'best' fitting ones and fit those PDFs for Argentina, Brazil, Colombia and Mexico using state-level data and the hierarchical model with partial pooling based on the method proposed in this chapter.

3.5 Conclusions

In this chapter, I provided the first (at the time of the study) estimates of the common epidemiological distributions for the COVID-19 epidemic in Brazil, based on the SIVEP-Gripe hospitalisation data [Ministério da Saúde, 2020]. I discovered an extensive heterogeneity between the different states. Quantification of the time-delay for COVID-19 onset and hospitalisation data could provide useful input parameters and priors for COVID-19 epidemiological models, especially those modelling the healthcare response in low- and middle-income countries.

The best-fitting densities approximating each of the hospitalisation distributions in this chapter have been selected using Bayes factors and the observed data extracted from the healthcare system database. This process assumed that all the observed data were correct. In reality, the data suffers from many limitations — specifically in the case of close to real-time reporting of hospitalisation during the peak pandemic times, the data tends to be lagged. In the next chapter, I dive deeper into this topic and propose a method of correcting the delays in the observed hospitalisation data, focusing on mortality reporting in Brazil.

Chapter 4

Gaussian Process Nowcasting: Application to COVID-19 Mortality Reporting

Updating observations of a signal due to delays in the measurement process is a common problem in signal processing, with prominent examples in a wide range of fields. An important example of this problem is the nowcasting of COVID-19 mortality: given a stream of reported counts of daily deaths, can we correct for the delays in reporting to paint an accurate picture of the present, with uncertainty? Without this correction, raw data will often mislead by suggesting an improving situation. We present a flexible approach using a latent Gaussian process capable of describing the changing auto-correlation structure present in the reporting time-delay surface. This approach also yields robust estimates of uncertainty for the estimated nowcasted numbers of deaths. We test assumptions in model specification such as the choice of kernel or hyperpriors, and evaluate model performance on a challenging real dataset from Brazil. Our experiments show that Gaussian process nowcasting performs favourably against both comparable methods and against a small sample of expert human predictions. Our approach has substantial practical utility in disease modelling — by applying our approach to COVID-19 mortality data from Brazil, where reporting delays are large, we can make informative predictions on important epidemiological quantities such as the current effective reproduction number.

4.1 Introduction

In many real-world settings, current observations from a noisy signal can be systematically biased, with these biases only being corrected after subsequent updates create more complete data. Often, these updates occur much later in the future due to data processing or reporting delays. Not accounting for these delays would result in biased predictions, while waiting for updates would result in a lack of timely estimates. The need for timely estimates to predict the present is colloquially known as *nowcasting* and its importance has been shown in a wide range of fields such as actuarial science, economics, and epidemiology [Bastos et al., 2019, Kaminsky, 1987, Lawless, 1994, McGough et al., 2020].

Nowcasting, as defined by Banbura et al. [2010] at the European Central Bank, is the process of predicting the present, the very recent past, and the very near future using time-series data known to be incomplete. An example from economics is using monthly data to nowcast the current state of important economy indicators such as GDP or income. More broadly, nowcasting is relevant for scenarios not only where the data are incomplete, but when the data are comprised of a biased subsample that will be updated in the future retrospectively, following lengthy delays.

In epidemiology, nowcasting is required due to delays in reporting arising from limitations in testing capacity, data curation, and the requirement for pseudonymisation of patient data [Bastos et al., 2019]. These delays are further compounded by the noise inherent in such data due to limited sampling (typically only a subset of the population is sampled). Throughout this chapter, we specifically focus on the delays in the reporting of deaths. An individual dies of a disease on a given day, but the delay between this event and the death being reported (and appearing in the dataset) can be substantial because of the reasons noted above. These reporting delays mask the true *current* state of the epidemic and have material consequences for our understanding of both the present and future evolution of the epidemic. For example, the estimation of key epidemiological quantities such as the effective reproduction number (R_t) would be systematically biased. Contemporary, real-time and unbiased estimates are necessary for effective public health planning and policy.

In this chapter, we propose a nowcasting framework based on latent Gaussian processes (GPs). This methodology is used to address the specific problem of delayed reporting of the true incidence of deaths due to COVID-19 in Brazil.

4.1.1 Related Methods

Previous methods for nowcasting exist in several different contexts. Bańbura and Modugno [2014] propose a maximum likelihood approach with a dynamic factor model to predict GDP. Shi et al. [2015] use a deep learning approach based on long short-term memory network (LSTM) to nowcast rainfall intensity. Codeco et al. [2018] provide a framework to gather epidemiological information and correct the reporting delays in Brazilian data. Bastos et al. [2019] present a Bayesian hierarchical model for nowcasting applied to data relating to dengue fever and severe acute respiratory infection cases. In McGough et al. [2020] a Bayesian nowcasting approach is proposed that produces accurate estimates that capture the time evolution of the epidemic curve. Specifically for COVID-19, Bayesian nowcasting approaches have been used to correct the reporting delays in Bavaria and Sweden [Altmejd et al., 2020, Günther et al., 2020][1]. Further discussion around the challenges in estimating reporting delays can be found in Seaman and De Angelis [2020]. The problem and background context for delays in reporting with corrected data in Brazil is further explained in Bastos et al. [2020], Villela [2020].

Our methods build upon and generalize the NobBS (Nowcasting by Bayesian Smoothing) method originally proposed by McGough et al. [2020]. NobBS is a Bayesian method that produces smooth and accurate nowcasted estimates in the presence of multiple diseases. NobBS allows for both uncertainty in the delay distribution and the evolution of the epidemic curve. While an effective method, NobBS has several limitations, such as the inability to pick up fast-occurring changes in the delay distribution, which we overcome in this chapter. The extensions we show result in comparable performance for COVID-19 mortality surveillance in Brazil, but present a better fit to the dynamic delays distribution.

4.1.2 Our contributions

The problem tackled in this chapter is conceptually illustrated in Fig. 4.1. The black points are the data available to us at a given time, and the red points are the ground truth that is only available much further in the future. It can be seen that the discrepancy between the presently available data and the underlying ground truth data grows markedly as we approach the present — a distinguishing characteristic of reporting delays. Alongside this, in Fig. 4.1, we also show 3 estimates of the effective

[1]The nowcasting for COVID-19 context has been further explored after the publication of this work, e.g. in Birrell et al. [2021], Wolffram et al. [2023].

reproduction number R_t (defined as the average number of infections an infected individual will go on to infect), obtained using a Bayesian hierarchical renewal-type model [Flaxman et al., 2020, Mellan et al., 2020, Mishra et al., 2020]. Understanding this epidemiological quantity is vital — $R_t > 1$ results in epidemics growing, while $R_t < 1$ results in epidemics declining. Fig. 4.1B shows estimates of R_t derived from the raw data, while Figs 4.1C and 4.1D show estimates of R_t derived from the ground truth data and our nowcasting approach respectively. These plots show that not correcting for delays can lead to a fundamentally different picture of the current epidemic state. Delays in death reporting lead to an underestimation of the true number of deaths in the observed data — the results suggest a declining epidemic, even though the epidemic is growing.

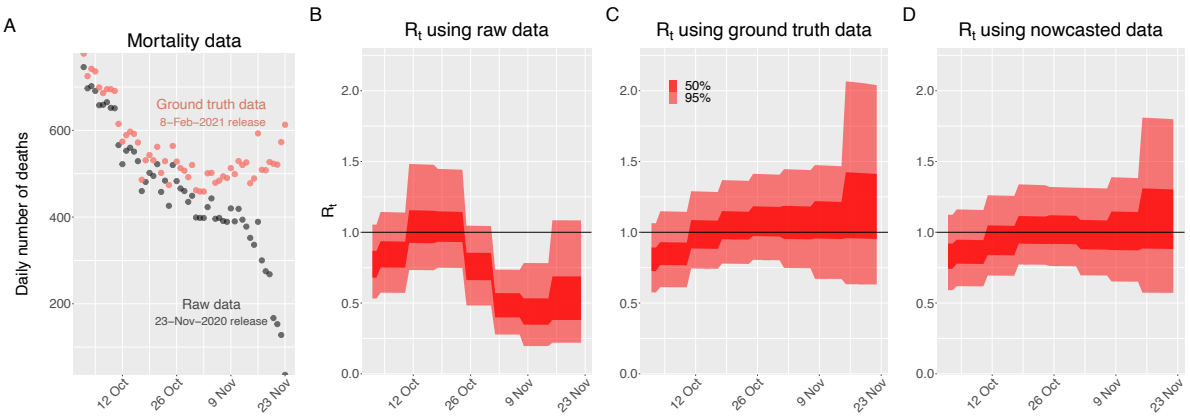


Figure 4.1: A) Reported daily hospital deaths are censored in recent times due to reporting delays. This can be seen by comparing the raw data with ground truth from two months in the future when the records have been backdated. B-C) The effective reproduction number R_t for SARS-CoV-2 infections in Brazil from 30$^{\text{th}}$ Jun 2020 to 23$^{\text{rd}}$ Nov 2020, estimated using deaths from the raw reported data released on the 23$^{\text{rd}}$ Nov 2020, and using a backdated ground truth based on data released on 8$^{\text{th}}$ Feb 2021. D) R_t estimates based on nowcasted mortality data. Whereas the raw data results in misleading estimates of R_t, with the estimated $R_t < 1$, by applying nowcasting to the death counts we achieve a picture of the epidemic closer to the truth. R_t estimation and the figure courtesy of Dr Thomas A. Mellan.

In this chapter, we focus on the Brazilian death data from the publicly available hospitalisation database with deaths from both confirmed and suspected COVID-19 diagnostic status [Ministério da Saúde, 2020]. Our central premise is that using these daily death data alone causes the policy decisions to be made based on false statistics and trends [Villela, 2020]. To facilitate well-informed policy-making based on unreliable data streams we propose and implement a nowcasting method using latent GPs. These GPs are capable of capturing the complex correlation structure in delayed data and present an effective means to correct the reporting delays. We use this corrected death data to calculate the effective reproduction number R_t using raw retrospective observed data, nowcasted data and the ground truth updated dataset (Fig. 4.1).

Our contributions are the following:

- We provide a new, flexible and accurate way to correct for delays in reporting. Our framework solves the nowcasting problem using latent GPs and provides realistic estimates for the deaths today given incomplete data. Our approach closely predicts the non-observed/missing values and simultaneously learns the underlying (latent) data-generating mechanisms of the delays.
- We compare our approach to an established alternative method (NobBS), and in a novel contribution, also provide a comparison to a small human expert panel of infectious disease epidemiologists. Domain knowledge is of primary importance for such applications and is frequently the primary approach taken to interpret data. In generating estimates that are improved over both existing computational methods as well as human experts, we demonstrate the utility of our approach.
- An important contribution of this work is the results and estimates provided. Implementing our approach enables generating more accurate estimates of the reproduction number; and in turn, a better understanding of the evolution of the COVID-19 epidemic in Brazil. Our framework is implemented in PyStan and the code is available online at `https://github.com/ihawryluk/GP_nowcasting`.

The structure of the chapter is as follows: in Section 4.2 we briefly introduce Gaussian Processes and describe the latent GP nowcasting models with several variants. In Section 4.3 we describe the data and perform retrospective tests to evaluate the accuracy of the new models and compare them with a sample of human experts' predictions. Finally, we discuss the advantages and limitations of the GP nowcasting framework in Section 4.5.

4.2 Gaussian Process nowcasting

This section delves into a comprehensive explanation of the methodologies applied throughout the study.

4.2.1 Nowcasting

Let n_t denote the response variable of interest that needs to be nowcasted at time t. In this chapter n_t represents the reported COVID-19 mortality in a week t. The mortality observations, in general, consist of measurements from an online data source, subject to distributed observation delays. The central task of nowcasting approaches is to identify a regular time-delay structure and use this to estimate n_t, at a time when it has only been partially observed. The structure that nowcasting identifies is the additive decomposition of the observable over the reporting delay d. That is, the true signal at a given time t is the sum over all the delayed partial observations for that time:

$$n_t = \sum_d n_{t,d} \,. \tag{4.1}$$

The intuition behind this formulation is that the "true" deaths that occurred at time t are distributed over various delays d due to the delays in reporting them.

A visual example of partial observation in recent times is the right-censored epidemiological data shown in Fig. 4.2A. For all data releases, we observe steep declines in contemporary data, which is then revised upwards as the data becomes more complete. In the COVID-19 context this occurs due to time lags in registering and reporting death certificates [Villela, 2020]. Fig. 4.2B shows that most deaths are reported to near completeness after around 5 weeks, and 90% are reported within 10 weeks. Splitting the data by delay index we form a 2D array in time and delay, $n_{t,d}$. The filled-in part of $n_{t,d}$, called the *reporting triangle*, is shown in Fig. 4.2C. The lower triangle part of this 2D array is missing, since at any time T only the number of deaths reported with delay $d \leq T - t$ are known for each epidemiological week t.

The representation of the data by time and delay, rather than time and reporting date, induces a regular structure — one that is auto-correlated and approximately monotonically decreasing in delay (Fig. 4.2C). This relatively simple structure makes this problem amenable to statistical modelling. The lower triangle of the $n_{t,d}$ matrix can be predicted with the model, and therefore an estimate of the true signal is available for any time up to the current time by Eq. 4.1, by summing over the delays. This is a common theme from which variations of nowcasting branch out.

To model the discrete positive values of $n_{t,d}$, we can use a Poisson or a negative-binomial likelihood for

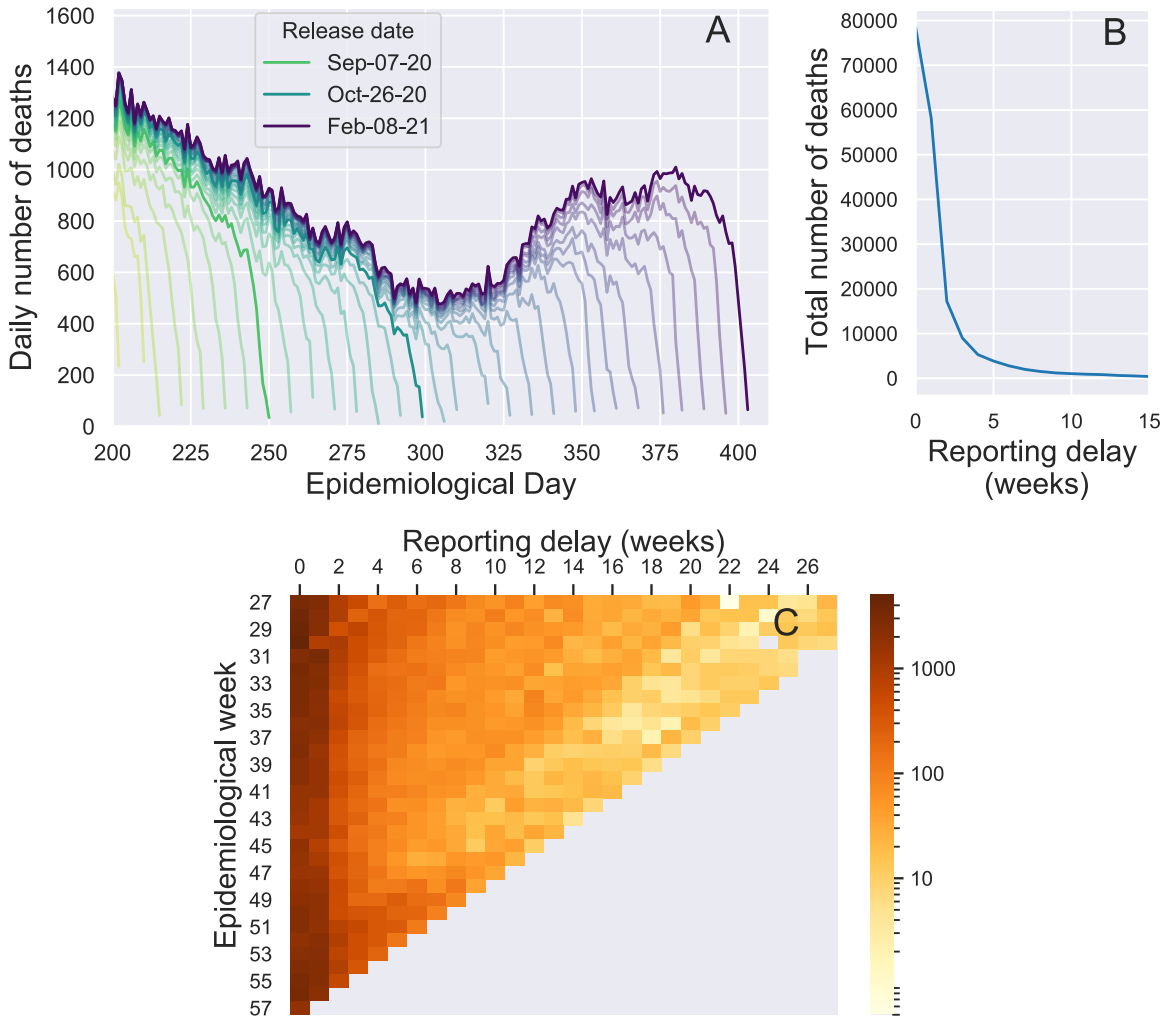


Figure 4.2: A) Daily COVID-19 deaths in Brazil as reported in data releases between July 2020 and Feb 2021. Each line represents a single release. B) Total number of deaths reported per reporting delay in weeks. Most deaths are reported with a delay of ≤ 5 weeks. C) The reporting triangle, shows the number of COVID-19 deaths reported each week with specific reporting delay. Mortality data were obtained from the SIVEP-Gripe hospitalisation database [Ministério da Saúde, 2020].

overdispersed data:

$$n_{t,d} \sim \text{NB}(\lambda_{t,d}, r)\,. \tag{4.2}$$

In the negative-binomial case, the dispersion parameter r is a hyperparameter that can be learnt or given an informative prior based on the problem. The latter approach is common among the established Bayesian nowcasting methods [Bastos et al., 2019, Günther et al., 2020, McGough et al., 2020]. The mean of the negative binomial, λ_{t_d}, is often modelled as a random walk [Bastos et al., 2019] or as an auto-regressive process [McGough et al., 2020] along the time dimension, that is joint independent with a learnt vector of

delays. The second approach is taken in the NobBS model [McGough et al., 2020], used in this chapter as a benchmark. Specifically, the NobBS model describes λ_{t_d} as:

$$\log(\lambda_{t,d}) = \alpha_t + \log(\beta_d) \tag{4.3}$$

where α_t is a latent signal for week t and β_d is the probability of reporting with delay d. It is worth noting, that with this approach, the distribution of delays β_d is fixed throughout the window of analysis.

This approach has been successful for dengue and influenza surveillance [Bastos et al., 2019, Codeco et al., 2018], but has limitations in terms of the generality of the time-delay covariance structure that can become apparent in more unpredictable nowcasting scenarios, such as an evolving epidemic with changing delay distributions, as we show in Fig. 4.3. Such issues can be minimised by tuning the window over which the static delay vector is estimated, or by manually adding cross-term covariates. Here we employ Gaussian Processes as a generic flexible alternative to model arbitrarily structured λ_{t_d}. The details of this are set out in the following section.

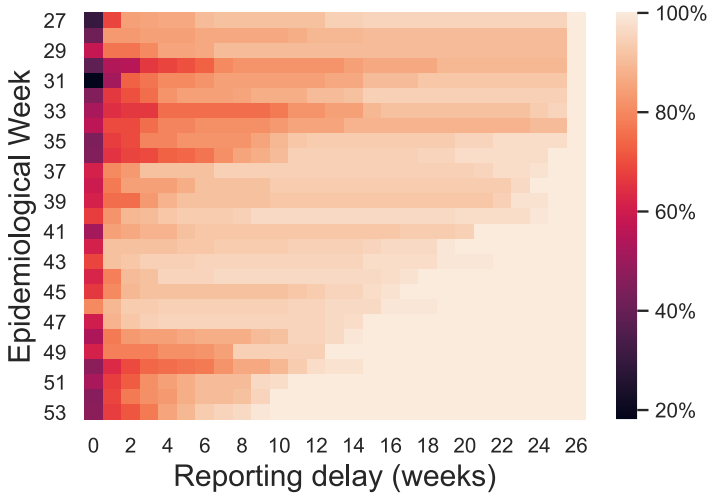


Figure 4.3: The change in the distribution of delays for Manaus (state Amazonas). Each cell in this heatmap shows the per cent of all deaths which occurred at week t and were reported with given delay d. Here we assume, that up till epidemiological week 53, all deaths have been recorded in the database. We use all available SIVEP data up to the release on 5^{th} April 2021.

4.2.2 Latent GP

The introductory model we consider consists of a latent GP with a 1D kernel. In general terms, GPs are a class of Bayesian non-parametric models that define prior over functions (see 1.5.3). They are a powerful tool in machine learning for learning complex functions with applications in regression and classification problems [Rasmussen and Williams, 2006, Wilson and Adams, 2013]. In recent years GPs have gained popularity in statistics and machine learning, due to their flexibility and excellent performance for many spatial and spatiotemporal problems [Flaxman et al., 2015, Wilson and Adams, 2013], including COVID-19 modelling [Qian et al., 2020]. The covariance function, or kernel, together with the mean function completely define a GP. The mean function is the base function around which all of the realisations of the GP are distributed. The *covariance kernel* is a crucial component of the Gaussian Process, as it describes the covariance of the Gaussian Process random variables, i.e. how similar two points are. Therefore, the kernel defines the shape of the distribution and which type of functions are more probable.

One of the most popular choices of covariance kernel, and the one we chose to introduce the model with, is the *squared exponential* kernel, k_{SE}, with entries defined by a covariance function $k_{\text{SE}}(\cdot,\cdot)$ such that

$$k_{\text{SE}}(t_i,t_j) = \alpha^2 \text{exp}\left(-\frac{||t_i - t_j||_2^2}{2\rho^2}\right). \tag{4.4}$$

The parameter α defines the kernel's variance scale, and ρ is a lengthscale parameter that specifies how nearsighted the correlation between pairs of time points (t_i) is. The kernel results in a prior over a set of functions to describe, $\lambda_{t,d}$, the mean of the statistical model in Eq. 4.2. This is modelled as a zero mean log-space latent Gaussian Process

$$\log(\lambda_{t,d}) \sim \text{GP}(0, k_{\text{SE}}). \tag{4.5}$$

Due to weak identifiability [Rasmussen and Williams, 2006], a strategy to identify the hyperparameters ρ and α is to fix the lengthscale ρ to the maximum delay time considered in the nowcasting problem, and learn only the scale parameter α. Markov Chain Monte Carlo (MCMC) is used to generate posterior summaries for arbitrary (non-normal) latent Gaussian Processes.

4.2.3 Generalised model

Additive Kernel Model

The basic model introduced above can be extended to provide a more expressive description of the data. The purpose of this is to be able to describe the complex structure in $n_{t,d}$.

Using the compositional kernel approach [Duvenaud et al., 2013, Wilson and Adams, 2013, Wilson et al., 2016], we can create a new additive kernel over multiple lengthscales, indexed s, as

$$k_{\text{add}} = \sum_s k_s \,,$$
$$\log\left(\lambda_{t,d}\right) \sim \text{GP}(0, k_{\text{add}}) \,. \tag{4.6}$$

The lengthscale hyperparameters are fixed or given strongly informative priors, ρ_s, while each α_s is learnt. In the simplest case, we consider a kernel with two lengthscale contributions, short- and long-range correlation structure:

$$k_{\text{add}}(t_i, t_j) = k_{\text{long}}(t_i, t_j) + k_{\text{short}}(t_i, t_j) + \sigma^2 \delta_{ij} \,, \tag{4.7}$$

plus a regularising term with a Kronecker delta function ensuring σ^2 Gaussian noise is only added when $i = j$. The choice of kernel confers bias that can result in a better generalisation. The logic of this kernel is to split the covariance into two components: (a) a smooth long-range component, used to extrapolate the trend into the unknown part of the reporting triangle where large distances from the observed points exist and (b) a part for describing variation in $n_{t,d}$ over shorter lengthscales. Additionally, the separation of kernels provides a generic method to describe more complex data-generating processes – for example, the long-range kernel can be squared-exponential, while the short-range can be a less smooth type with a different power spectrum such as Matérn (1/2). This can create a general statistical model for $n_{t,d}$. Furthermore, in this regard the δ contribution provides a source of regularisation which may be useful if there is reason to believe $n_{t,d}$ values are subject to variation beyond the scope of the basic nowcasting framework. For example, if a death can switch category from a COVID-19 suspected death to a cause other than COVID-19 in later data releases, this could result in a negative $n_{t,d}$ count, which can be modelled as an error to be regularised.

A further modification that can be applied if the time-delay surface $n_{t,d}$ has a complex structure, is to split the data into two components and model each with separate kernels. For example, if delays of 0 or 1 weeks account for a large fraction of total counts, they can be considered separately to delays > 1. This approach is considered later in Section 4.3.2. But a more generic formulation is to consider a 2D kernel to fully account for the time-delay correlation structure, which is introduced below.

2D Kernel Model

As a further expansion of the approach described before, we introduce a separable two-dimensional kernel over time and delay, $k((t_i, t_j), (d_i, d_j)) = k_t(t_i, t_j)k_d(d_i, d_j)$. Separable kernels can be efficiently implemented using Kronecker product algebra as described in Flaxman et al. [2015]. Specifically, individual Gram matrices for time and delay are combined using the Kronecker product such that

$$K_{t,d} = K_t \otimes K_d \,. \tag{4.8}$$

As before, the kernel can be given an additive structure over multiple lengthscales. For example,

$$\begin{aligned} k_{\text{long}}(t, d) &= k^t_{\text{long}} k^d_{\text{long}} \,, \\ k_{\text{short}}(t, d) &= k^t_{\text{short}} k^d_{\text{short}} \,, \\ \log(\lambda_{t,d}) &\sim \text{GP}(0, k_{\text{long}}(t, d) + k_{\text{short}}(t, d)) \,. \end{aligned} \tag{4.9}$$

This approach captures the relationship between t and d. In both 1D and 2D kernel approaches it is possible to perform partial pooling of the model parameters by combining two or more spatial locations with similar features, for example neighbouring states, if limited data is available. In practice, however, we found a limited gain in doing so as our approach works well with relatively few observations.

4.3 Data and Model properties

4.3.1 Data

The numbers of deaths per date have been extracted from the Brazilian Ministry of Health's Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe) database [Ministério da Saúde, 2020]. SIVEP-Gripe is a large publicly available database providing anonymised patient-level records of all individuals who died or were hospitalised with suspected or confirmed COVID-19 in Brazil [Bastos et al., 2020, de Souza et al., 2020, Niquini et al., 2020]. New data have been released regularly online, every week, in the second half of 2020 considered here. In this study, we extracted all SIVEP-Gripe data releases from 7$^{\text{th}}$ July 2020 to 31$^{\text{st}}$ May 2021[2]. We consider all cases of suspected or confirmed COVID-19 (classes 4 and 5).

There are several potential sources of error in the reported SIVEP data. One is under ascertainment — systematic biases which are beyond the scope of correction by this nowcasting methodology. Another source of error is delayed classification. After the initial input of patient data into the database (usually at the time of hospitalisation), the entry might be later updated with clinical and laboratory data, including confirmatory COVID-19 testing. Further updates will include the outcome and its date (i.e. date of death or date of hospital discharge) and cases receive a final classification. Cases can be classified as confirmed (class 5) or suspected COVID-19 (class 4), or other causes (classes 1-3). Despite being described as a "final classification", reclassification does occur and it is especially common for unknown cases to be reclassified as COVID-19 once results from confirmatory tests are informed to the health authorities. On the other hand, some deaths attributed to suspected SARS-CoV-2 infection are later "removed" from the SIVEP database, due to duplicate filtering or because they are eventually attributed to other diseases. That can cause the number of deaths on certain days to decrease in consecutive data releases, as shown in Fig. 4.4.

The number of deaths per day as reported by each release is presented in Fig. 4.2, together with a reporting triangle, showing the distribution of the reporting delays across time. According to the SIVEP-Gripe dataset, over 90% of all deaths have been reported with a delay of less than 10 weeks (Fig. 4.2B). We therefore choose the maximum reporting delay D for our data to be $D = 10$, and sum up all deaths reported with a delay longer than 10 weeks. Finally, to create the reporting triangle appropriate for our

[2]The paper Hawryluk et al. [2021] was submitted in May 2021.

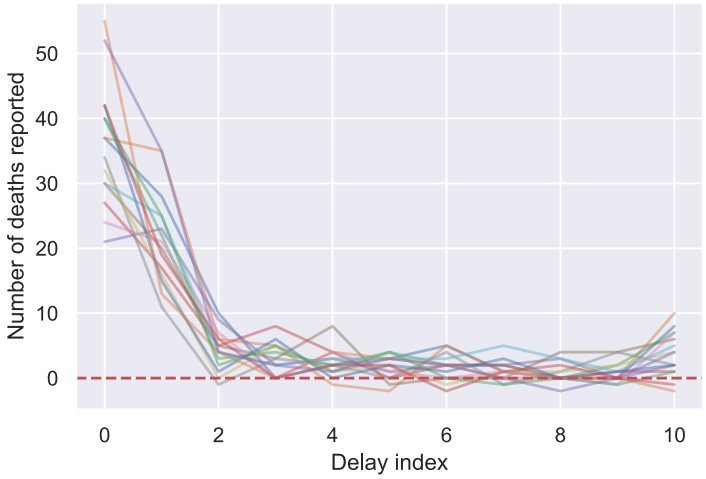


Figure 4.4: Example of weekly reporting delay with negative signal for Amazonas, epidemiological weeks 27 to 42. Each week's mortality data are plotted as a single line. Some lines fall under the $y = 0$ line (red dashed line), indicating that some deaths were incorrectly assigned to a given week, which was corrected by the following data releases.

model, we aggregate the data into weeks.

4.3.2 Model Fit

We fit and present 7 models. For 1D kernel GPs, we consider a single SE kernel (1D SE), and additive long- and short-range component kernels (1D SE+SE and 1D SE+Mat). The additive long- and short-range component kernels are also considered coupled with splitting the data across delays greater and less than one (1D SE+SE data-split). Finally, we consider a 2D kernel GP model with additive long and short-range components (2D additive). The NobBS model of McGough et al. [2020] is fitted and presented for reference of the current state-of-the-art. All models are fitted to the SIVEP-Gripe weekly COVID-19 deaths reported in Brazil, at the time of the study available until 31st May 2021.

Posterior samples of the parameters in the models were generated using Hamiltonian Monte Carlo with Stan [Carpenter et al., 2017, Hoffman and Gelman, 2014], using the PyStan interface (version 2.19.0.0). For each fit, we used 4 chains and 1000 iterations, with 400 iterations dedicated to warm-up. The convergence of each model fit was evaluated by ensuring that $\hat{R} < 1.01$ for each parameter. Traceplots and other MCMC diagnostic measures were also investigated (see Section A.2 in the Appendix).

Each of the models, characterised by the likelihood given in Eq. (4.2), and a latent GP part for modelling

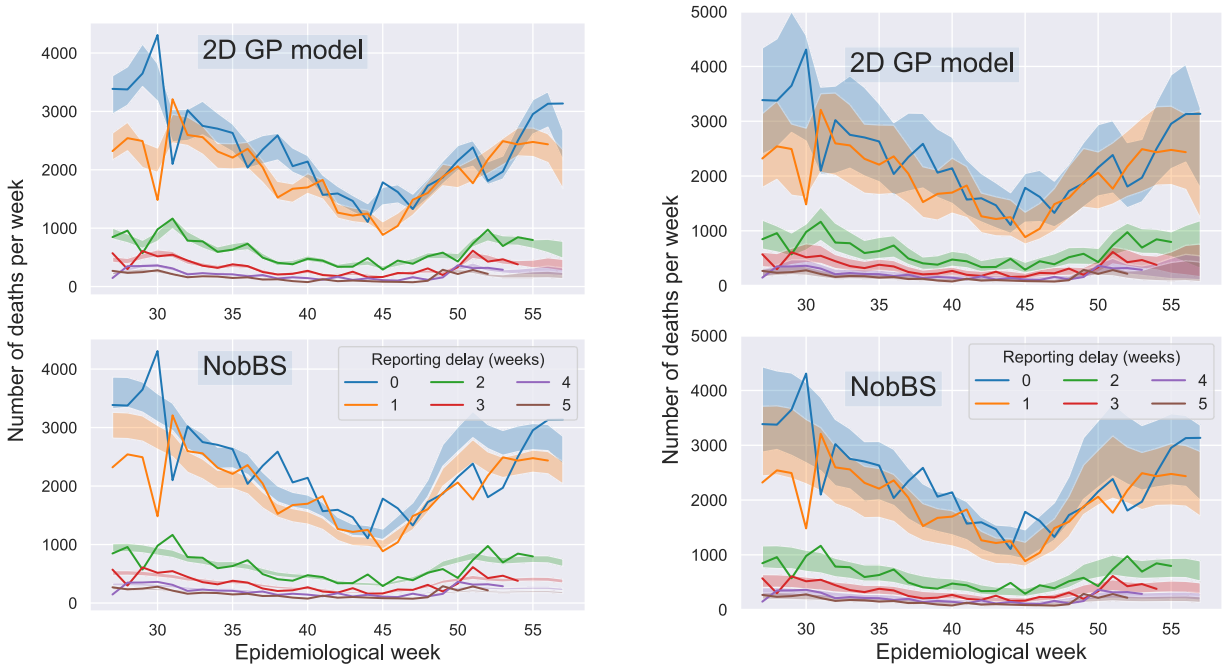
the $\lambda_{t,d}$ (Section 4.2.2) is trained by supplying the reporting triangle $n_{t,d}$ filled with data available up to the point of the nowcast. Each of the parameters governing the model, such as overdispersion r or lengthscales and variances of the GPs are learnt during the model fit. The best-performing hyperparameters of the prior distribution were selected conditioned on the observed results. The parameters and their prior densities are given in Table 4.1. The model training and nowcasting through sampling each element of the $n_{t,d}$ matrix is done simultaneously. Specifically, at each iteration parameter values are sampled and immediately used to sample from the negative-binomial distribution to obtain all elements of the $n_{t,d}$ matrix.

Table 4.1: Priors for the analysed models. Here T is the maximum number of weeks for which the data is available, that is rows in the reporting triangle, and D is the maximum reporting delay, that is columns in the reporting triangle. Γ denotes a Gamma distribution. For the overdispersion parameter r, the priors were informed by the data but tested in the sensitivity analysis (see 4.4.3). * the same hyperparameters were used for the models with Matérn(1/2) and Matérn(3/2) kernels.

	1D SE	1D SE+SE	1D SE+Mat*	1D SE+SE data-split	2D additive	NobBS
r	$\Gamma(500, 2)$	$\Gamma(500, 2)$	$\Gamma(500, 2)$	$\Gamma(500, 2)$	$\Gamma(200, 2)$	$\Gamma(500, 2)$
$\alpha_{1,\text{long}}$	$\mathcal{N}(1, 1)$	$\mathcal{N}(15, 2)$	$\mathcal{N}(15, 2)$	$\mathcal{N}(15, 2)$	-	-
$\alpha_{2,\text{long}}$	-	-	-	$\mathcal{N}(20, 2)$	-	-
$\alpha_{1,\text{short}}$	-	$\mathcal{N}(5, 2)$	$\mathcal{N}(5, 1)$	$\mathcal{N}(5, 1)$	-	-
$\alpha_{1,\text{short}}$	-	-	-	$\mathcal{N}(3, 1)$	-	-
$\alpha_{1,t}$	-	-	-	-	$\mathcal{N}(T, 1)$	-
$\alpha_{2,t}$	-	-	-	-	$\mathcal{N}(D, 1)$	-
$\alpha_{1,d}$	-	-	-	-	$\mathcal{N}(0, 1)$	-
$\alpha_{2,d}$	-	-	-	-	$\mathcal{N}(0, 1)$	-
$\rho_{1,\text{long}}$	$\mathcal{N}(3, 1)$	$\mathcal{N}(T, 0.1)$	$\mathcal{N}(T, 0.1)$	$\mathcal{N}(T, 0.1)$	-	-
$\rho_{2,\text{long}}$	-	-	-	$\mathcal{N}(D, 0.1)$	-	-
$\rho_{1,\text{short}}$	-	$\mathcal{N}(1, 0.01)$	$\mathcal{N}(1, 0.01)$	$\mathcal{N}(1, 0.01)$	-	-
$\rho_{2,\text{short}}$	-	-	-	$\mathcal{N}(1, 0.01)$	-	-
$\rho_{1,t}$	-	-	-	-	T	-
$\rho_{2,t}$	-	-	-	-	1	-
$\rho_{1,d}$	-	-	-	-	D	-
$\rho_{2,d}$	-	-	-	-	1	-
σ_1	$\mathcal{N}(0, 1e-6)$	$\mathcal{N}(0, 1e-6)$	$\mathcal{N}(0, 1e-6)$	$\mathcal{N}(0, 1e-6)$	$\mathcal{N}(0, 1e-7)$	-
σ_2	-	-	-	$\mathcal{N}(0, 1e-6)$	$\mathcal{N}(0, 1e-7)$	-
τ	-	-	-	-	-	$\Gamma(0.01, 0.01)$
$\alpha[1]$	-	-	-	-	-	$\mathcal{N}(0, \sqrt{\frac{1}{0.001}})$
$\alpha[t]$	-	-	-	-	-	$\mathcal{N}(\alpha[t-1], \sqrt{\frac{1}{\tau}})$
β	-	-	-	-	-	Dirichlet(0.1)

Other nowcasting methods, including NobBS, focus primarily on estimating only the 'missing' part of the $n_{t,d}$ array and comparing the total numbers n_t, that is sums of each row of the array. Here, we aim to obtain a statistical model explaining all elements of the $n_{t,d}$ matrix. The reason for that is twofold: firstly,

having a model that describes the whole $n_{t,d}$ surface well increases the reliability of the model, which is vital in any healthcare setting. Secondly, the SIVEP-Gripe database contains hard-to-identify errors discussed in Section 4.3.1, therefore it is preferable to treat the reported data with additional statistical uncertainty. The fit of the 2D GP and the NobBS models to the $n_{t,d}$ matrix is presented in Fig. 4.5 and shows that the GP-based nowcasting method fits the time-delay structure much closer than NobBS. For comparison, similar plot for the 1D GP nowcasts is given in Fig. 4.6, from which it is clear that the 2D GP fits to the delay structure better.



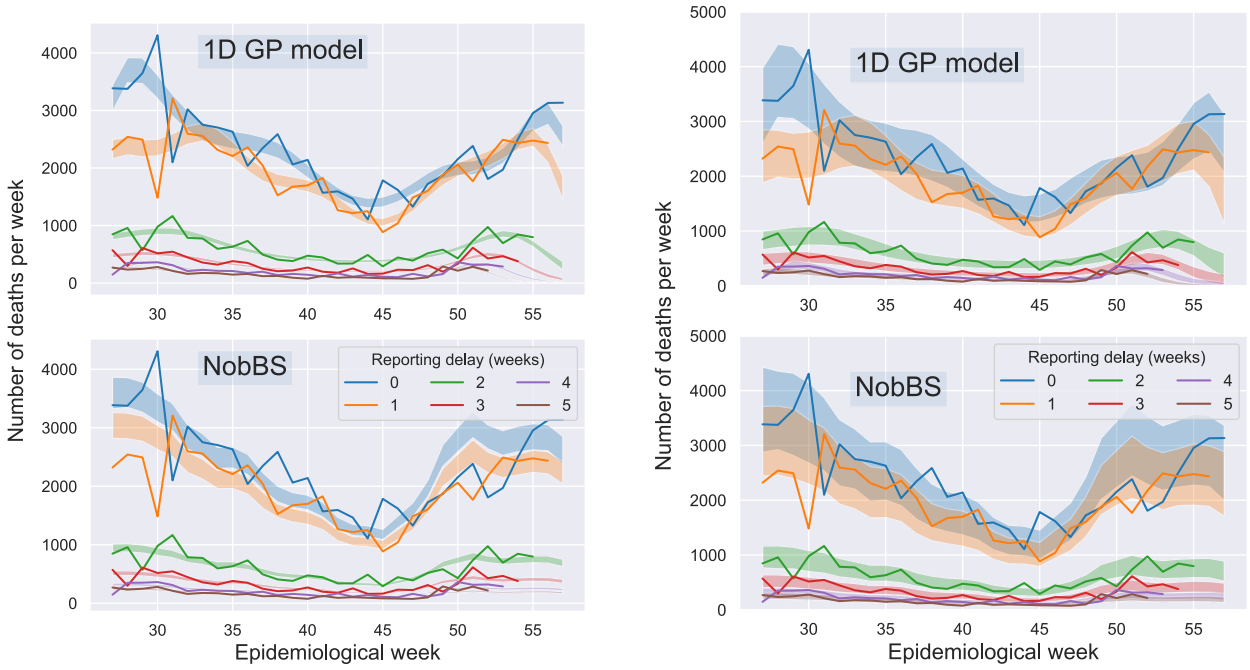
(a) 50% CrI

(b) 95% CrI

Figure 4.5: Reported and nowcasted numbers of deaths with reporting delay 0 to 5 weeks generated by the 2D additive GP and NobBS models. These plots show the columns of the reporting triangle $n_{t,d}$. The reported data are shown with solid lines, and the 50% CrI for the nowcasts with the ribbons in the first column, and the 95% CrI in the second column.

4.4 Results

In this section, we present the analysis of various tests of the proposed GP nowcasting model — retrospective testing and benchmarking against the NobBS method, comparison with the human epidemiology experts' predictions, and sensitivity analysis.



(a) 50% CrI

(b) 95% CrI

Figure 4.6: Reported and nowcasted numbers of deaths with reporting delay 0 to 5 weeks generated by the 1D GP and NobBS models. These plots show the columns of the reporting triangle $n_{t,d}$. The reported data are shown with solid lines.

4.4.1 Retrospective testing

To evaluate the accuracy of the competing nowcasting models, we fit all of the models to the retrospective data sets available each week between 5$^{\text{th}}$ Oct and 30$^{\text{th}}$ Nov 2020, using the numbers of deaths recorded for the whole of Brazil. This way we obtain 63 different sets of nowcasts, which we compare to the deaths reported by the most recent SIVEP-Gripe data release from 31$^{\text{st}}$ May 2021. The start date gives us at least 15 weeks of training data for each nowcast. The end date of 30$^{\text{th}}$ Nov is 26 weeks before the most recent release[3], so the number of deaths reported in the most recent release can be confidently taken as a true value. The comparison is done by calculating the weighted and unweighted rooted mean squared error (RMSE) and the continuous ranked probability score (CRPS) between the "true" values from the most recent release and the nowcasted values. The differences between the ground truth and raw data, as shown in Fig. 4.1A, are used as weights. For each RMSE and CRPS evaluation, we use the mortality data from the 10 weeks leading up to the date of the nowcast.

[3] At the time of writing, that is May 2021

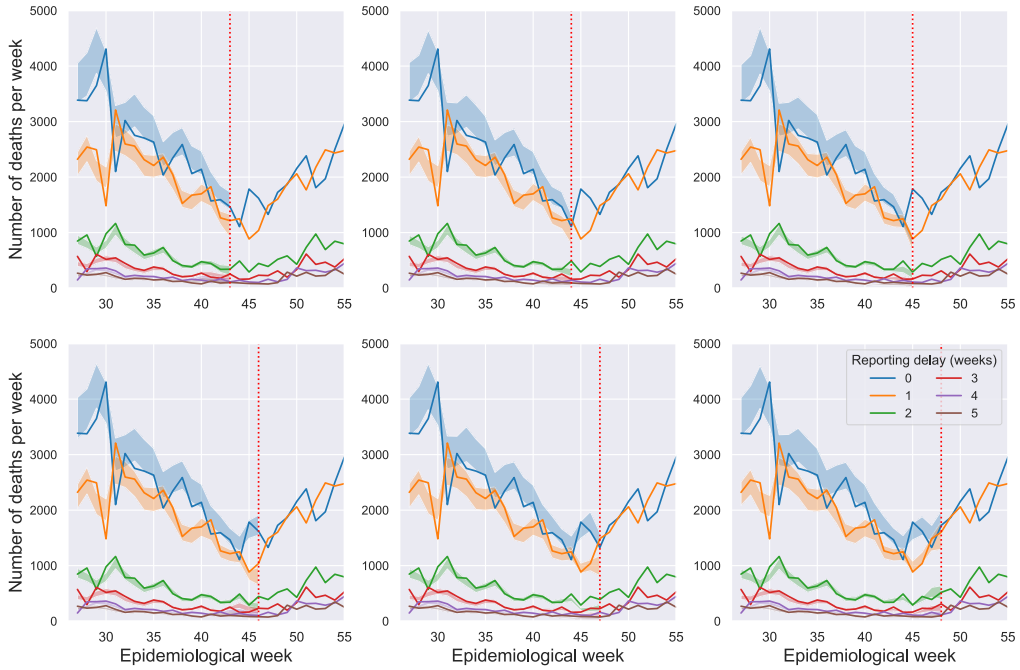


Figure 4.7: Distribution of reporting delays for each of the retrospective tests. The moment of the "nowcast" is shown with the red dotted line in each plot. Solid line presents the data extracted from the release of SIVEP data from 31st May 2021, and the ribbons show 50% CrI of the model fit obtained using the 2D additive kernel GP model.

Model fits for the 2D additive GP model, including the 95% credible intervals (CrI), are shown in Figs 4.7 and 4.8. Specifically, Fig. 4.7 shows the fits to each of the delay distributions in the reporting triangle, and Fig. 4.8 gives the fits to the total numbers of deaths per week. The fits for the remaining models are shown in the Appendix in Figs. A.2 to A.8. GP models with 1D kernels with 2 components (SE+SE and SE+Mat) performed worse than the benchmark. The predictive accuracy of the other models was comparable to that of the benchmark, as shown in Fig. 4.9, while also simultaneously giving an appropriate statistical description of the data (Fig. 4.5). These results provide empirical evidence that our proposed method, under correct specification, gives a complete and accurate approach for nowcasting COVID-19 death data.

4.4.2 Comparing against human experts

In addition to the forecasting metrics, as a novelty, we also evaluate how the GP nowcasting model performs when compared to human experts' predictions. We asked a group of infectious disease epidemiology

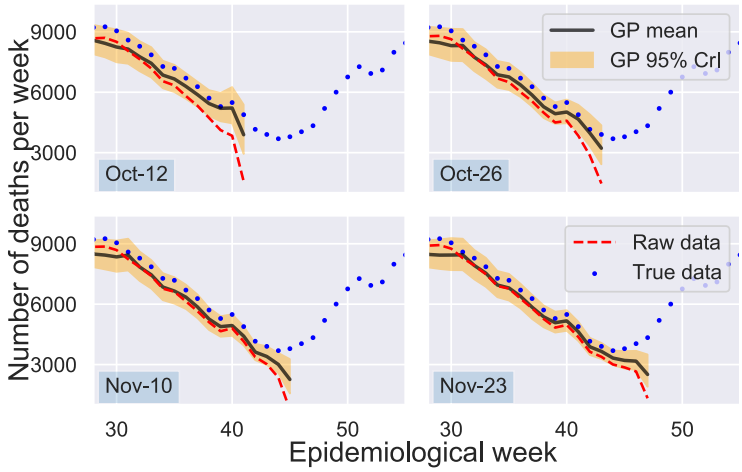


Figure 4.8: Retrospective testing for the whole of Brazil using the 2D additive GP model.

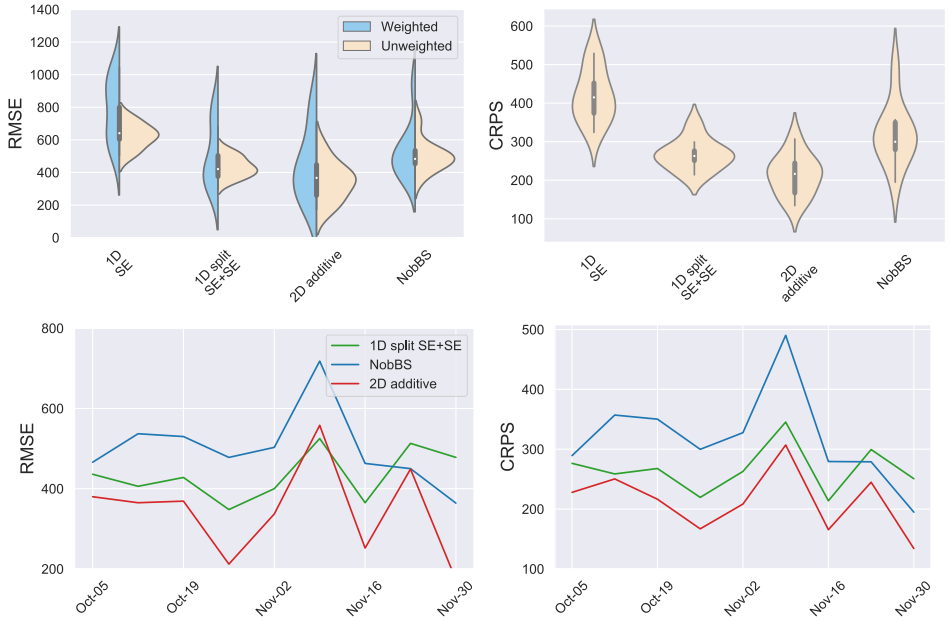


Figure 4.9: RMSE and CRPS evaluated for n_t for tested nowcasting methods over the weeks 5$^{\text{th}}$ Oct to 30$^{\text{th}}$ Nov 2020. For the weighted RMSE, the weights were calculated as a difference between the ground truth and the raw data available up to the nowcasting date.

experts[4] to provide a series of nowcasts when presented with time-series data up to 12$^{\text{th}}$ Oct and 23$^{\text{rd}}$ Nov 2020 (solid lines in Fig. 4.10). They were asked for their estimates of the true numbers of deaths due to COVID-19 in Brazil on the 8$^{\text{th}}$ Oct and 19$^{\text{th}}$ Nov 2020. The dates were specifically chosen to represent different scenarios. In the first one, 8$^{\text{th}}$ Oct 2020, both the raw data and the updated numbers of daily

[4]Through a survey sent out in January 2021 to staff and students of the Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London. The answers were anonymous.

deaths were declining. Whereas in the second date, 19$^{\text{th}}$ Nov 2020, the raw data were declining while the updated release revealed that the true numbers of daily deaths were actually increasing.

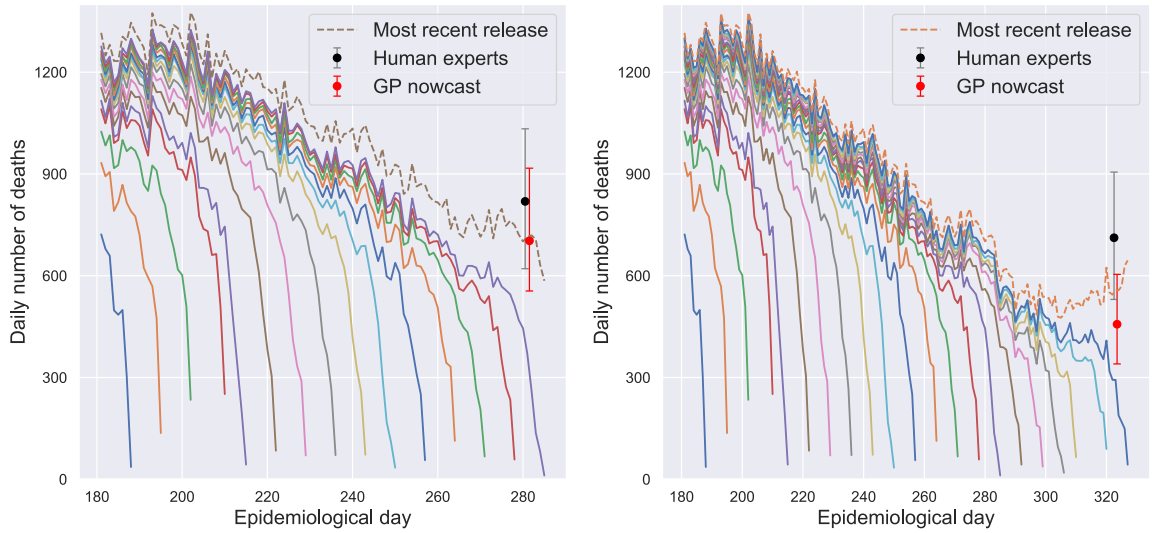


Figure 4.10: Human epidemiology experts and 1D SE+SE data-split GP model estimates of a true number of deaths on 8$^{\text{th}}$ Oct 2020 (left) and 19$^{\text{th}}$ Nov 2020 (right) plotted together with the reported data.

For this experiment, 36 anonymous experts provided their point estimates and confidence intervals, presented in Fig. 4.11. To extract daily deaths from the model's weekly estimates, we performed a simple interpolation, by setting the nowcasted number of deaths per week divided by 7 to the middle day of the given week and interpolating the remaining values using splines (example shown in Fig. 4.12).

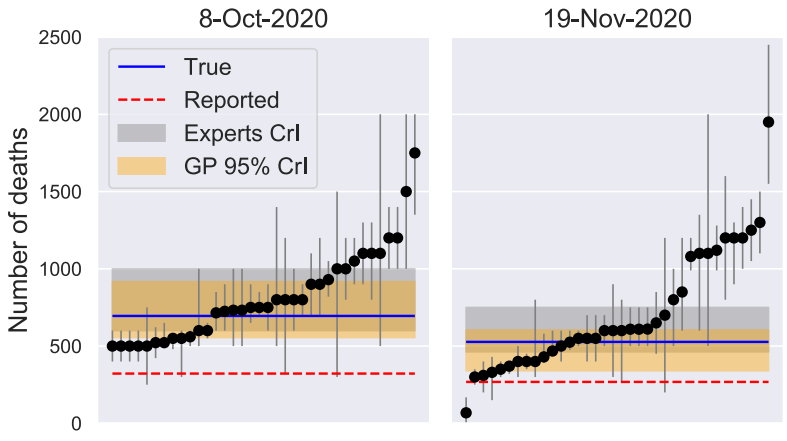


Figure 4.11: Human experts' estimates of the true number of deaths are shown with the black points and error bars.

Notably in both cases, the median value guessed by the human experts was not far from the "true" value (difference of 55 and 73 deaths/day respectively for the human median estimate and 13 and 72 for

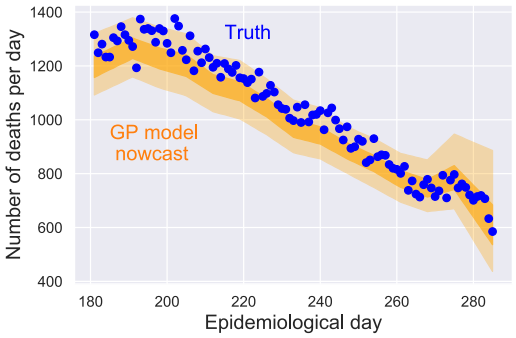


Figure 4.12: Example of the interpolation of numbers of deaths per day based on the weekly nowcasted values. The 50% and 95% CrI for the nowcast are shown with the ribbon.

the nowcasting model mean), however, only 36% and 50% of all answers included the true value within the provided credible intervals, respectively. The confidence intervals given by the human experts were comparable with the 95% CrI of the model, with the model confidence narrower by 19 deaths/day for the first date and 27 deaths/day for the second date.

4.4.3 Sensitivity analysis

We performed a basic sensitivity analysis for the 1D SE+SE data-split GP model. We first varied the prior for the overdispersion parameter r. This parameter is often unidentifiable by the models and has to be chosen based on the data [McGough et al., 2020]. This is confirmed by our sensitivity analysis, where changing the prior for r changed the width of the confidence interval, but did not impact the mean predictions, as shown in Figs 4.13 and 4.14.

Changing the priors for the scale parameter α to less informative, that is increasing the variance of the priors does not significantly affect the mean predictions (Figs. A.11 and A.12 in the Appendix). Changing the mean of the priors does however have an impact on the predictions and results in bi-modality of the posterior distribution if the model is misspecified, e.g. when $\mathrm{N}(0, 1)$ priors are used (Figs. A.13 and A.14 in the Appendix).

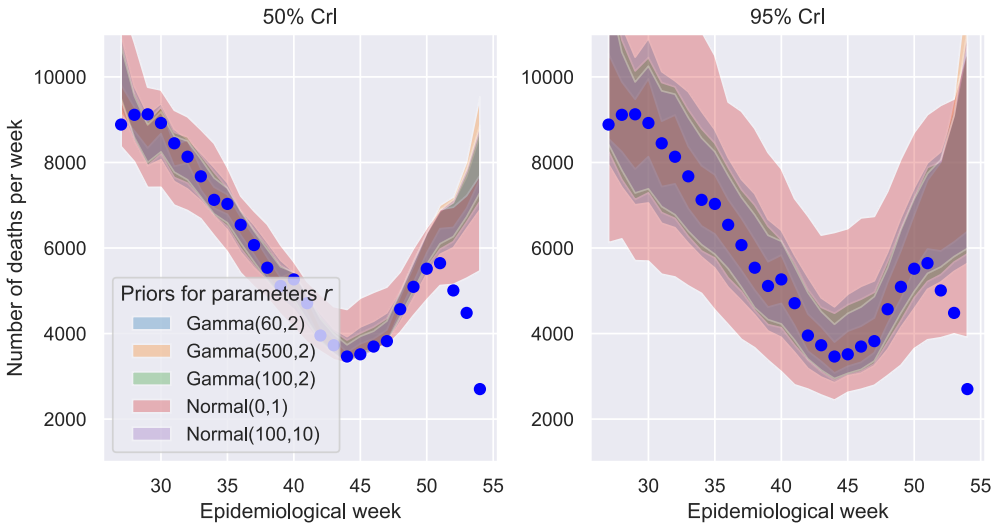


Figure 4.13: Model fits with different overdispersion r prior density for the 1D SE+SE data-split GP model.

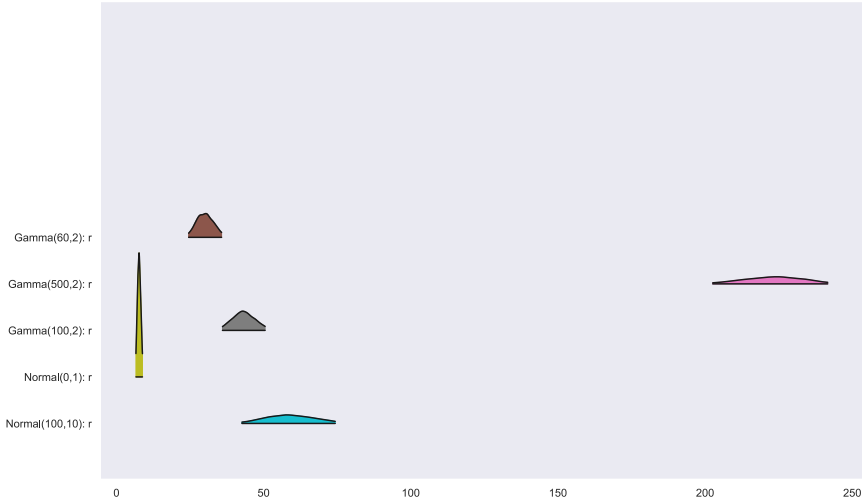


Figure 4.14: Posterior distribution of the overdispersion r parameter depending on the chosen prior density for 1D SE+SE data-split GP model.

4.5 Discussion

We proposed a new, flexible method for correcting reporting lags and demonstrated its applicability using COVID-19 mortality data in Brazil.

Summary of the findings

Applying nowcasting to surveillance data suffering from reporting delays is crucial to accurately track real-time epidemic dynamics. The limitations associated with using non-corrected data in epidemiological analyses are highlighted with our results of the R_t estimates shown in Fig. 4.1. Using this raw data leads to continued underestimation of R_t and predicts a declining epidemic. Specifically, in the month preceding the nowcast, the relative entropy value (see 1.3.5) for the ground truth and raw data R_t was on average 13.14 (max 43.8) and for ground truth and nowcasted data R_t only 0.26 (max 0.35), as shown in Fig. 4.15. By contrast, the ground truth results show that the epidemic remains uncontrolled, with R_t remaining above 1 — an important conclusion also captured by our nowcasting approach.

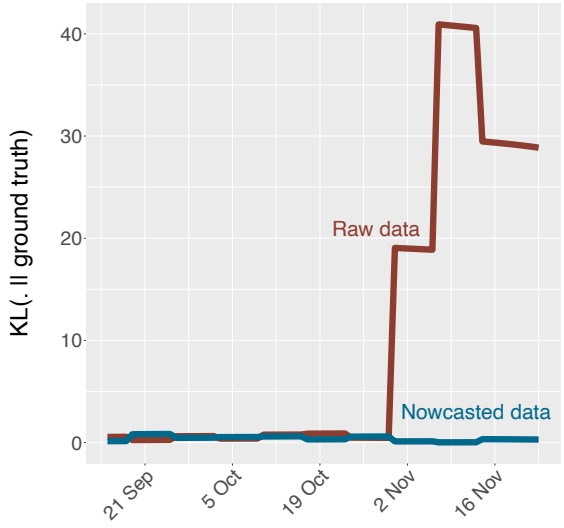


Figure 4.15: Kullback-Leibler divergence (relative entropy) between the R_t value estimated using raw data and nowcasted data.

Practical applications

The emergence of SAR-CoV-2 variants of concern with altered epidemiological characteristics, such as increased transmissibility [Volz et al., 2021] or partial evasion of immunity [Faria et al., 2021], emphasises the need for accurate and continued real-time epidemic tracking. To estimate COVID-19 mortality in Brazil, during a resurgent phase of the epidemic concurrent with the emergence of the Gamma (P.1) variant, we perform nowcasting using data released on the 8$^{\text{th}}$ Feb 2021 and compare the predictions to the updated numbers of deaths released on the 31$^{\text{st}}$ May 2021. The results of the nowcast for the

whole of Brazil are shown in Fig. 4.16. The reported data show a decline in the weekly number of deaths since epidemiological week 54 for Brazil, however, the nowcasted results show much higher numbers of estimated weekly deaths, closer to the true value. We performed a similar nowcasting analysis for numbers of COVID-19-related deaths in Manaus city in the Amazonas state in Brazil, which was used in Faria et al. [2021]. That nowcasted data was necessary, alongside genomic data and other known properties of the coronavirus, for mathematical modelling of a newly emerged Gamma variant of SARS-CoV-2, in order to compare the characteristics of the new variant to the previously circulating non-Gamma lineages.

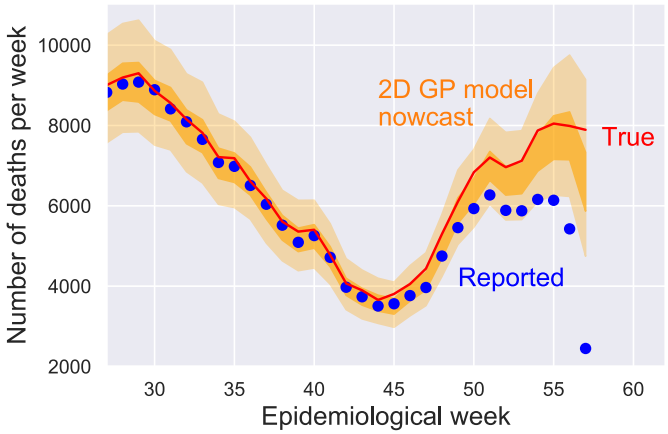


Figure 4.16: Nowcasted and reported deaths due to COVID-19 death for Brazil up to 8th Feb 2021 generated with the 2D additive GP model. 50% and 95% CrI for the GP model nowcasts are shown with the ribbon. True values were obtained from the SIVEP release on the 31st May 2021.

The GP nowcasting models introduced in this chapter can be readily used for real-time monitoring of the new outbreaks of diseases, as relatively few data points are required to train the model, ca. 3 months here. In other applications, more data may be required, depending on the distribution of reporting delays, variance of counts, and regularity and granularity of data. Although this study focuses on the application of the proposed GP-based nowcasting framework to the death counts for the whole country, the GP models can be applied at finer spatial scales, as illustrated for individual states of Brazil in Fig. 4.17. This flexibility is important due to the large heterogeneity of the healthcare system in the country [Baqui et al., 2020, Brizzi et al., 2022, Castro et al., 2021, Hawryluk et al., 2020].

Comparison with other nowcasting models

The GP nowcasting framework we introduced was benchmarked with the established NobBS method [McGough et al., 2020]. The main theoretical difference between the two proposed models is how the

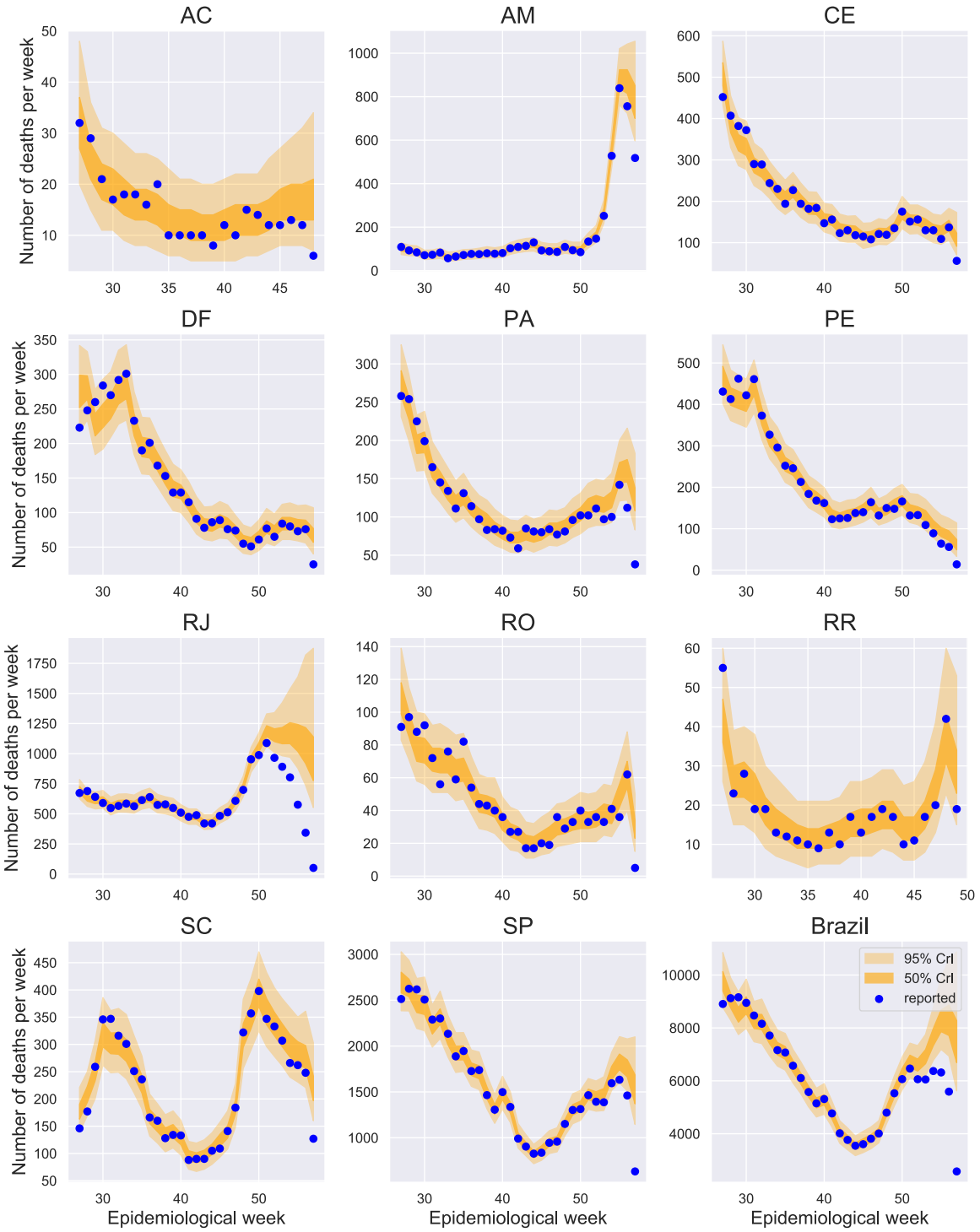


Figure 4.17: Nowcasts made by the 1D SE+SE data-split GP model, using data up to 8th Feb 2021 for Acre (AC), Amazonas (AM), Ceará (CE), Distrito Federal (DF), Pará (PA), Pernambuco (PE), Rio de Janeiro (RJ), Rondônia (RO), Roraima (RR), Santa Catarina (SC), São Paulo (SP) and whole Brazil.

mean of the negative-binomial distribution, describing the number of deaths per week, is modelled. Our model uses a latent Gaussian Process, whereas NobBS uses a first-order random walk. The 2D GP model improves on the RMSE for point predictions and CRPS for the distribution of samples compared to NobBS, as shown in Fig. 4.9, and provides more realistic uncertainty intervals (compare e.g. Figs. A.7 and A.8). As well as improving predictive performance on the missing parts of the reporting triangle, the GP framework provides a more expressive statistical model capable of better explaining the historical reporting data (see Fig. 4.5).

Limitations

One of the limitations of the approach described here is the dependence on the historical data and the regularity of the data releases, a limitation shared by many other nowcasting approaches. Additional challenges include variability in the distribution of reporting delays over time. For example, during the initial phase of the Brazilian COVID-19 epidemic, reporting delays were particularly severe. Delays in reporting are typically most extensive during outbreaks of a novel pathogen (such as SARS-CoV-2), due to the limitations in diagnostic availability and testing capacity. Relatedly, during epidemic peaks, the strain on healthcare systems and administrative staff due to increasing admissions can also lengthen reporting delays.

4.6 Conclusions

In this chapter, I have presented a new approach to modelling time-delay data, which can be used to nowcast online data streams that have statistically distributed delays. Our approach uses latent Gaussian Processes with additive kernels and gives a fully flexible and generic method to describe and predict the data for unknown delays. The method has been demonstrated for assessing mortality and estimating the effective reproduction number for COVID-19 reporting in Brazil, but can also be used for other contexts in which delays in a measurement process exist.

So far in this thesis I discussed a method for model selection, presented evidence for the importance of spatial heterogeneity, and introduced a method of correcting the reporting delays in the observed data. Altogether, this touches on various aspects of building and fitting the model to the data. However, the

practitioners often come across a situation, in which even with the correct data and reasonable, judiciously selected model, the inherent issues with the process of model fitting cause the results to be suspicious, or even wrong. In the next chapter, I investigate one such issue: identifiability, which is the ability of a model to uniquely determine the values of parameters of the model based on the data.

Chapter 5

Identifying importations in a renewal equation epidemic model

Identifiability issues affect statistical models across various disciplines, hindering the unique determination of parameters from available data. This problem often remains undetected, especially in models with non-interpretable parameters, where traditional diagnostics fail to signal parameter misidentification. Such issues are prominent in epidemiological modelling, where diverse factors can yield identical observed outcomes, complicating parameter estimation. Addressing this, we investigate the identifiability of importations in a discrete renewal equation-based epidemic model through simulated scenarios. Our study focuses on varying importation timing and magnitude, proposing a solution by constraining the variance of the reproduction number (R_t) using informative priors. By harnessing epidemiological insights suggesting limited daily R_t fluctuations, we hypothesise improved identifiability of importations. Through comprehensive modelling and analysis, we demonstrate the effectiveness of our approach in identifying importations, offering valuable insights for infectious disease modelling.

5.1 Introduction

The problem of identifiability is ubiquitous in statistical modelling. In general, it refers to the ability to uniquely identify the model's parameters using the available data [Ran and Hu, 2017]. Despite variations in data volume, quality, and model accuracy, some models may remain unidentifiable, which is attributed to the intricate interplay between data and parameters [Sameni, 2023]. For example, if the model includes two parameters which are confounded and only appear as a product, those cannot be uniquely identified [Ran and Hu, 2017]. Another example is a simple linear regression problem: consider a model with intercept β_0 and slope β_1 and only one data point known (x, y). The model is

$$y = \beta_0 + \beta_1 x + \epsilon \tag{5.1}$$

Given only one data point, there exists an infinite number of combinations of values of parameters (β_0, β_1) that can produce the same point (x, y), so the parameters (β_0, β_1) are unidentifiable.

The main difficulty in diagnosing this problem is that it often goes unnoticed, especially if the model contains non-interpretable parameters. Typically considered diagnostics for fitting statistical or machine learning models concern the convergence or the estimate errors, but those do not offer an explicit sign of a parameter misidentification.

This problem is widely observed in many scientific fields, including infectious disease modelling [Chowell, 2017, Reich et al., 2021, Tuncer and Le, 2018]. In epidemiological statistical models, this problem manifests itself when various mechanisms lead to the same observed outcome. For example, the number of observed new cases of a certain disease or a pathogen infection can be caused by a variety of reasons: influx of infected individuals from different locations, sudden increase in the reproduction number R_t, increased testing campaigns, etc. This is especially true for quantities like R_t, as they change over time and are not solely an inherent property of a pathogen, but rather an outcome of several variables, including the pathogen's transmissibility, social behaviour, immunity, and many others. One of the ways to avoid misidentification problems in applied Bayesian statistics is to use informative priors in the models and rigorously assess the results of the inference using the domain knowledge [Gelfand and Sahu, 1999].

5.1.1 Our contributions

In this study, we focus on the identifiability of importations in a simple renewal equation-based epidemic model. As mentioned, the increase in observed cases could be a result of importations from some location external to the system (location) considered, but could also be explained by increases in R_t. Here, we consider a number of simulated scenarios, where we vary the time and the magnitude of importations. We hypothesise that regularisation through constraining the R_t variance, by choosing a narrow informative prior, can aid the models to correctly identify and estimate the importations [Gelfand and Sahu, 1999]. The reasoning behind this intuition is the fact that, based on the epidemiological studies, the value of R_t should not vary drastically day to day (excluding particular scenarios where e.g. we observe a super spreading event or interventions aiming to curb the pathogen spread are implemented, impacting the connectivity patterns of both healthy and infected individuals).

The structure of the chapter is as follows: in Section 5.2 we introduce the model used for simulating the data and subsequently perform fitting. The results are presented in Section 5.3, where first we show three examples of a single double-spike importation scenario in subsection 5.3.1. Then, in subsection 5.3.2, we demonstrate the scenario where the simulated importation data has a seasonal character. Finally, in Section 5.4, we discuss the findings and their implications for infectious disease modelling.

5.2 Methods

In this section, we provide a detailed explanation of the methods employed in the study.

5.2.1 Model

The model used in this project for simulating the data and subsequent fitting is based on the widely used discrete formulation of the renewal equation model used for parametrisation of R_t [Cori et al., 2013, Flaxman et al., 2020, Fraser, 2007a, Mishra et al., 2020, Nouvellet et al., 2018].

We model the daily incidence y_t, that is the number of new cases on day t, with a Poisson distribution

with day-dependent mean $f(t)$. $f(t)$ is governed by a discrete renewal equation.

$$\begin{aligned} y_t &\sim \text{Poisson}(f(t)) \\ f(t) &= \mu_t + R_t \sum_{\tau<t} f(t-\tau)g(\tau) \end{aligned} \tag{5.2}$$

where y_t denotes the total number of new cases on day t (incidence), μ_t is a number of imported cases, R_t is an effective reproduction number on day t and $g(t)$ is a generation interval.

The R_t is parametrised using a 1-dimensional Gaussian Random Walk (GRW) (see 1.5.1). GRW is a random walk, where the step size is sampled from Normal distribution, that is:

$$\begin{aligned} GRW(\sigma, n) &= [x_1, x_2, ..., x_n] \\ x_{t+1} &= x_t + \epsilon, \ t = 1, ..., n \\ \epsilon &\sim \mathcal{N}(\sigma) \end{aligned} \tag{5.3}$$

where x_t is a location at time t (in our case the R_t) and ϵ is a step size (change in R_t between times t and $t+1$) and n is a total number of steps.

The GRW formulation for R_t includes n steps and σ_R variance as follows:

$$\begin{aligned} \sigma_R &\sim \text{Exp}(\theta_R) \\ R' &\sim \text{GRW}(\sigma_R, n) \\ R &= e^{R'} \end{aligned} \tag{5.4}$$

In the formulation above we used a shortened way of describing how the GRW is used for parametrisation of R_t, based on the GRW definition from Eq. 5.3. The reader should keep in mind here that in this notation both R' and R are vectors of dimension n (which is the number of GRW steps, or more specifically the number of days considered in our scenarios). Additionally, to prevent the negative values of R_t, intermediate values sampled from a GRW are exponentiated.

The hyperprior parameter θ_R is varied in the simulated scenarios, in order to inspect the impact of the variance prior on the model fit to the simulated data. We tested a range of values for the θ_R parameter,

from $\theta_R = 0.1$ (highest variance case) to $\theta_R = 100$ (lowest variance case). Intermediate values were tested as well. Those values were chosen based on the typical priors used by infectious disease modellers in similar models.

Similarly, importations are also modelled by a GRW. As an intermediate step, for each day we sample an α_i parameter from Relaxed Bernoulli (also known as Continuous Bernoulli) distribution [Loaiza-Ganem and Cunningham, 2019], which is then used as a multiplier for the value of μ. This is done to give more flexibility to μ, as the importations in our considered scenarios are not smooth day-to-day, as well as to improve the sampling as continuous parameters are easier to fit than discrete by the HMC algorithms. Relaxed Bernoulli is defined over the unit interval of $(0, 1)$, and the degree of relaxation is controlled by the temperature (*temp*) parameter (as the temperature approaches 0, the distribution becomes discrete). Examples of samples from Relaxed Bernoulli distributions with different sets of parameters are shown in Fig. 5.1

$$\begin{aligned} \theta_\mu, \text{temp} &\sim U(0,1) \\ \alpha &\sim \text{RelaxedBernoulli}(\text{temp}, \theta_\mu) \\ \sigma_\mu &\sim \text{Exp}(1) \\ \mu' &\sim \text{GRW}(\sigma_\mu, n) \\ \mu &= e^{\alpha\mu'} \end{aligned} \tag{5.5}$$

As before, notice that μ' and μ are both n-dimensional vectors.

5.2.2 Simulating the data

Typically, R_t does not stay constant but changes smoothly over time. For this reason, we chose to simulate the n-dimensional vector describing R_t by setting:

$$R = 1.15 + \sin(0.15 * [1, 2, ..., n]^T) \tag{5.6}$$

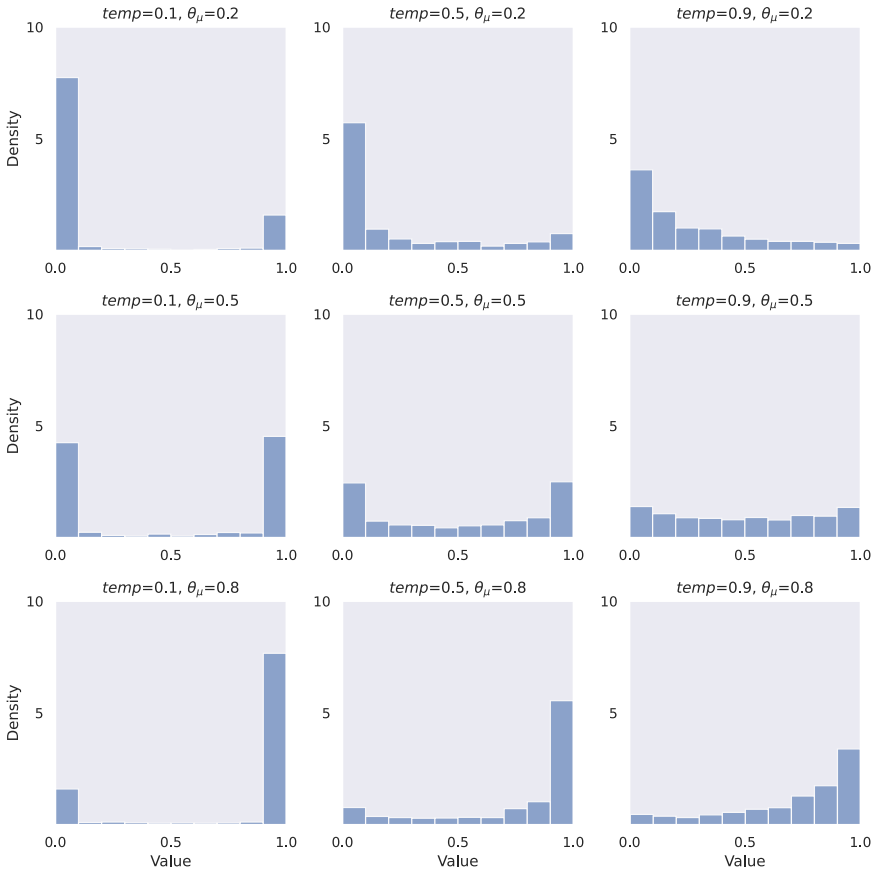


Figure 5.1: Example of Relaxed Bernoulli sampled values. As the *temp* parameter approaches 0, the distribution becomes discrete.

We chose the formulation in which R_t oscillates around the value of 1, as consistent lower values would quickly lead to the epidemic dying out, and higher ones would lead to the exponential growth of the case numbers, making it difficult to focus on numbers of imported cases (this will be in detail explained in the further sections).

We considered two main scenarios: Scenario 1 (subsection 5.3.1) considers a case where we observe two singular importation events. In this scenario, we chose two days where we want to observe those events and fix the number of new imported cases.

Scenario 2 (subsection 5.3.2) considers a case of importations occurring every day in a seasonal pattern. For simulation of imported cases, we use a sine or cosine function, similarly as for simulation of R_t in Equation 5.2.2.

Finally, to simulate the generation interval g we normalise n values from the probability density function

of the Gamma density with shape parameter $a = 3$ (as per Mishra et al. [2020]).

5.2.3 Model fitting

Once the data is simulated, we fit the model described in Section 5.2.1 to the daily incidence y_t, $t = 1, ...n$ (observed data). We also supply the simulated generation interval to the model.

In all scenarios, the simulations were run for $n = 50$ days. *High variance* examples were run with a value for R_t variance prior (Eq. 5.4) set to $\theta_R = 0.1$, that is

$$\begin{aligned} \sigma_R &\sim \text{Exp}(0.1) \\ R' &\sim \text{GRW}(\sigma_R, 50) \\ R &= e^{R'} \end{aligned} \tag{5.7}$$

The *low variance* examples were run with $\theta_R = 100$, that is

$$\begin{aligned} \sigma_R &\sim \text{Exp}(100) \\ R' &\sim \text{GRW}(\sigma_R, 50) \\ R &= e^{R'} \end{aligned} \tag{5.8}$$

Here we use a parametrisation of the exponential distribution using *rate*, which means that for $X \sim \text{Exp}(\theta)$, the expected value of X will be $\mathbb{E}[X] = \frac{1}{\theta}$. That means that for the high variance case, the mean-variance of the GRW step will be 10, whereas for the low variance, it will be 0.01. It is important to note, that although these values are not unrealistic, they were chosen to illustrate the problem tackled in this project in an educational and proof-of-concept way and we do not necessarily advocate using these specific priors for the epidemiological modelling studies.

Technical notes

All code required to reproduce the analysis presented in this chapter can be found on `https://github.com/ihawryluk/importations`.

Simulation of the data and fitting of the models were implemented in Python [Python Software Foundation], using a probabilistic programming library NumPyro [Phan et al., 2019]. All models were run using 4 chains, with 1000 warm-up steps and 20,000 further iterations, in order to obtain satisfactory convergence.

Each model fit was evaluated using the standard procedures for fitting Bayesian models. The convergence and correct mixing of the chains was evaluated by checking that $\hat{R} < 1.01$ for each parameter [Gelman and Rubin, 1992]. As the purpose of this study was to show how the choice of priors impacts the inference, rather than obtaining a perfect fit to the data, some of the model-priors pairs have not reached convergence. Traceplots and other MCMC diagnostics, such as Highest Posterior Density Interval (HPDI) and Effective Sample Size (ESS) were also analysed.

5.3 Results

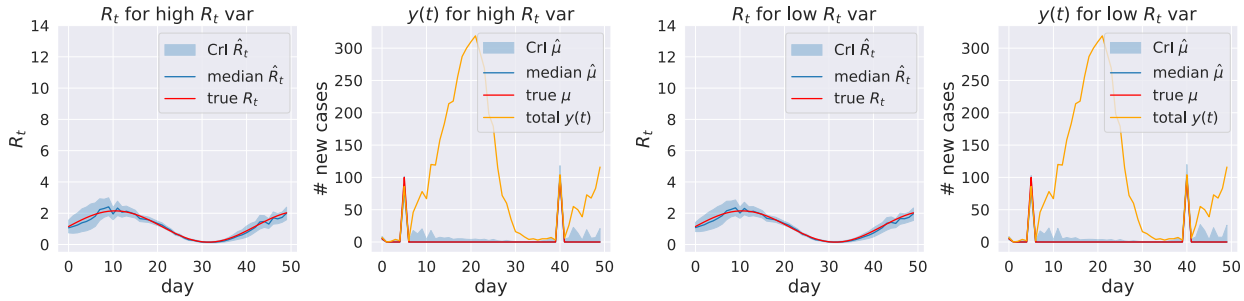
5.3.1 Scenario 1: double spike

In our simulations, we initially explored a scenario involving two isolated importation events. This occurs, for instance, when a pathogen is emerging in a specific location or has yet to be introduced, while simultaneously circulating in another area. Subsequently, individuals from a region experiencing an ongoing epidemic may collectively travel by plane or ferry to a location with minimal infections, potentially including some who are already infected with the pathogen.

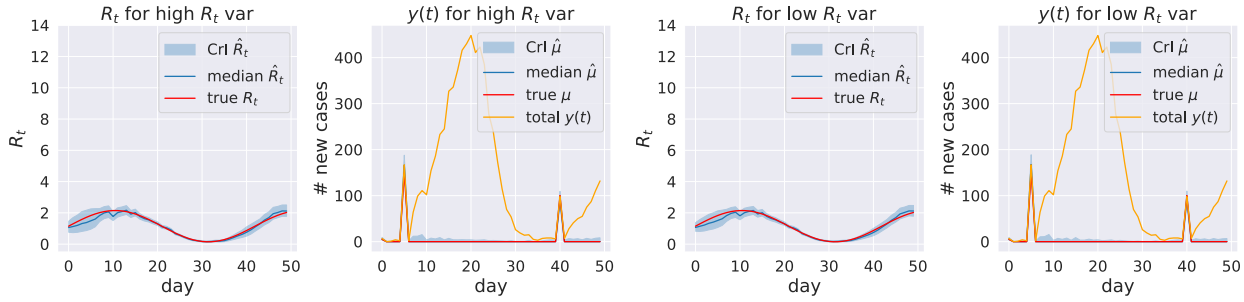
Second spike when R_t increasing but incidence close to 0

The first variation of this scenario is a case of two importation events taking place at the time where R_t is increasing, but the incidence is close to 0. The first spike happens at the beginning of the epidemic on day 5, and the second one during the second wave of R_t on day 40, while the values of μ_1 and μ_2 (numbers of imported cases at each of the events) are varied. The outbreak is initiated with 5 cases on day 0. The results of fitting the model to 4 examples of that are shown in Fig. 5.2, while the prediction errors are given in Fig. 5.3. Selected parameters' posterior distributions are compared in Fig. 5.4.

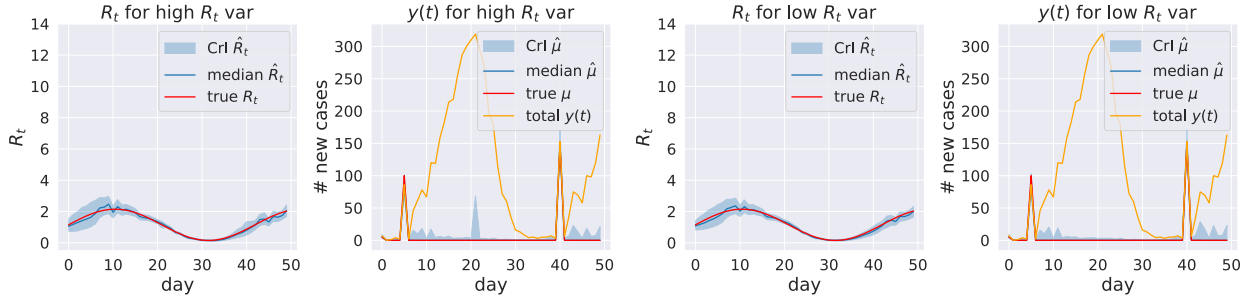
In the first 3 examples (Fig. 5.2 a-c), median posterior (blue line) recovers the true value (red line) for both R_t and μ estimates regardless of the chosen prior value for R_t variance. For both models, however,



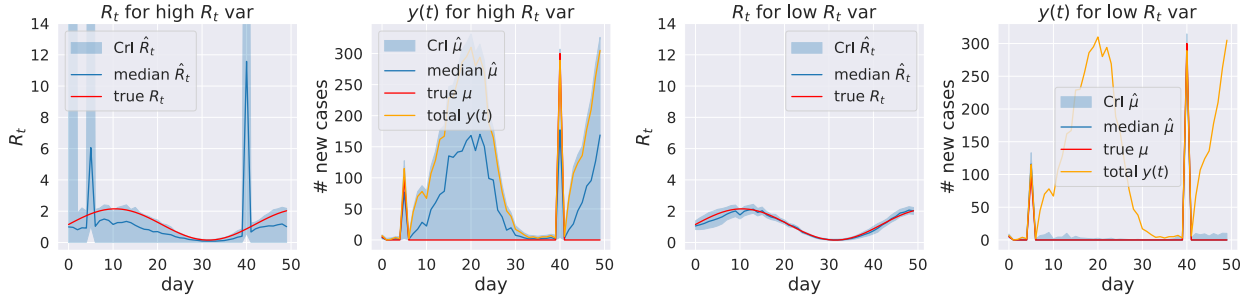
(a) Equal spikes: $\mu_1 = \mu_2 = 100$



(b) First spike higher: $\mu_1 = 150, \mu_2 = 100$



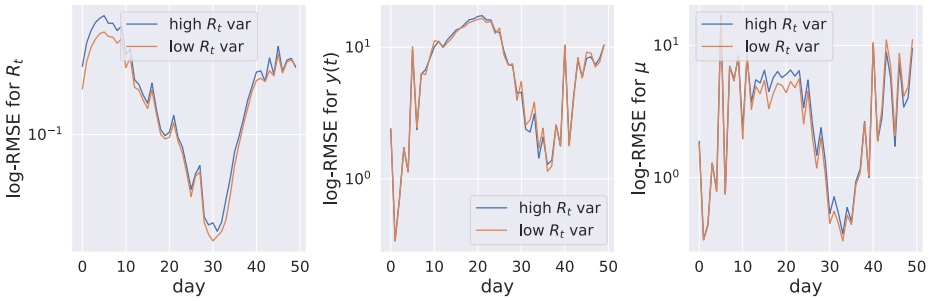
(c) Second spike higher: $\mu_1 = 100, \mu_2 = 150$



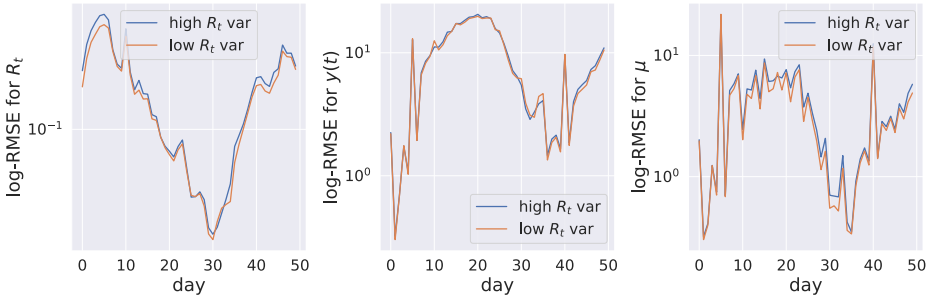
(d) Second spike higher: $\mu_1 = 100, \mu_2 = 300$

Figure 5.2: Simulation scenario with two importation events, first in the beginning of the epidemic (increasing R_t) and second one during the second increase in R_t, but when incidence is close to 0.

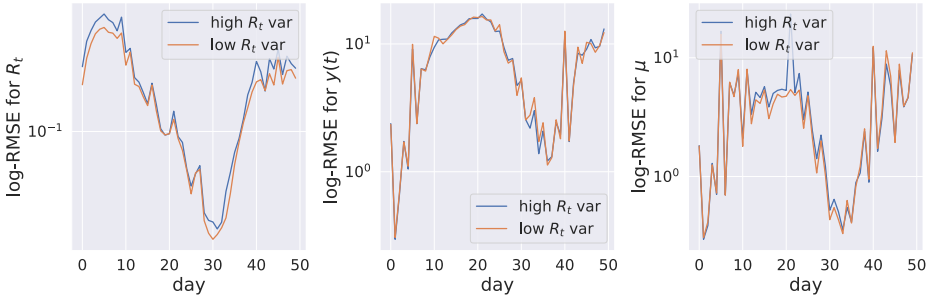
we can observe some wider credible intervals (CrI) overestimating the importations right after the actual importation takes place.



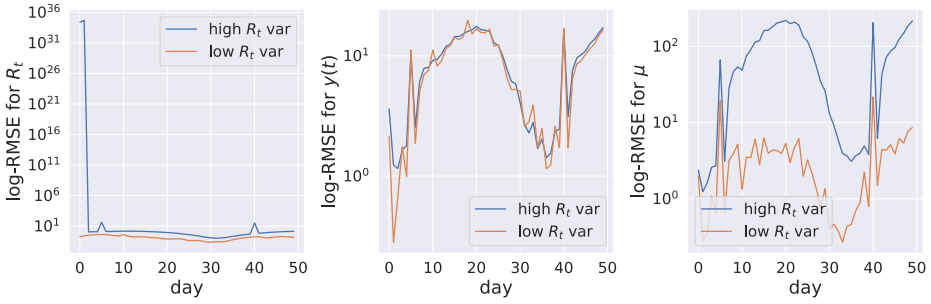
(a) Equal spikes: $\mu_1 = \mu_2 = 100$



(b) First spike higher: $\mu_1 = 150, \mu_2 = 100$



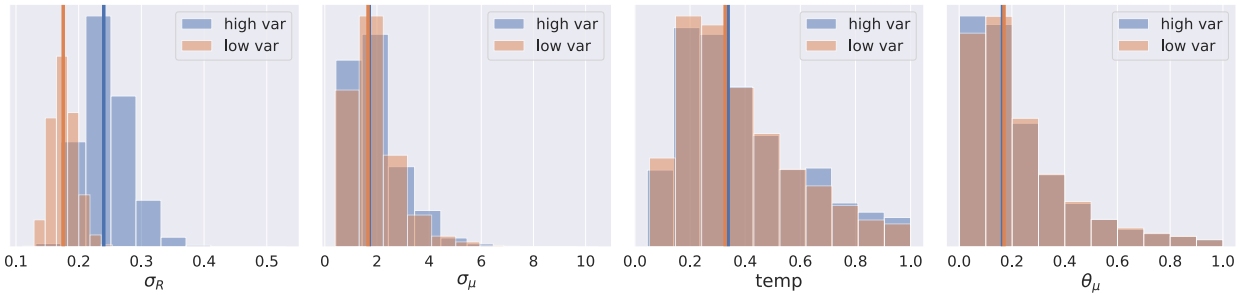
(c) Second spike higher: $\mu_1 = 100, \mu_2 = 150$



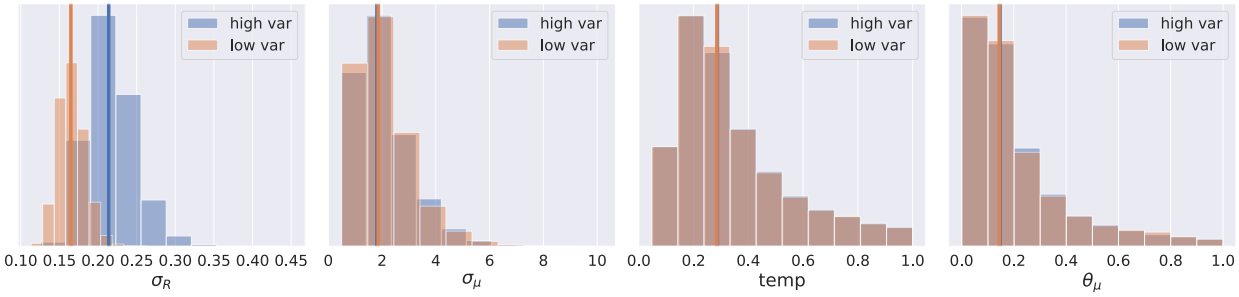
(d) Second spike higher: $\mu_1 = 100, \mu_2 = 300$

Figure 5.3: RMSEs for the simulation scenario with two importation events, first at the beginning of the epidemic (increasing R_t) and second one during the second increase in R_t, but when incidence is close to 0.

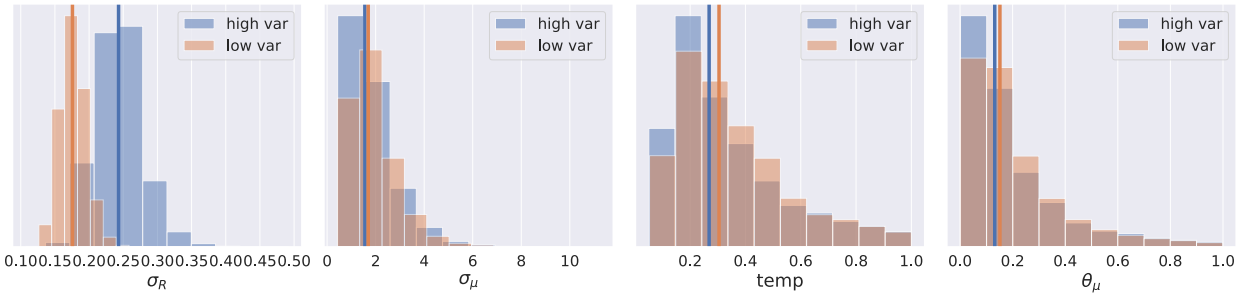
In the last example (Fig. 5.2d and Fig. 5.3d), where μ_2 was set to be higher than μ_1 ($\mu_1 = 100$, $\mu_2 = 300$), median posterior of the model fit with high variance underestimates both μ_1 and μ_2, fully attributing the



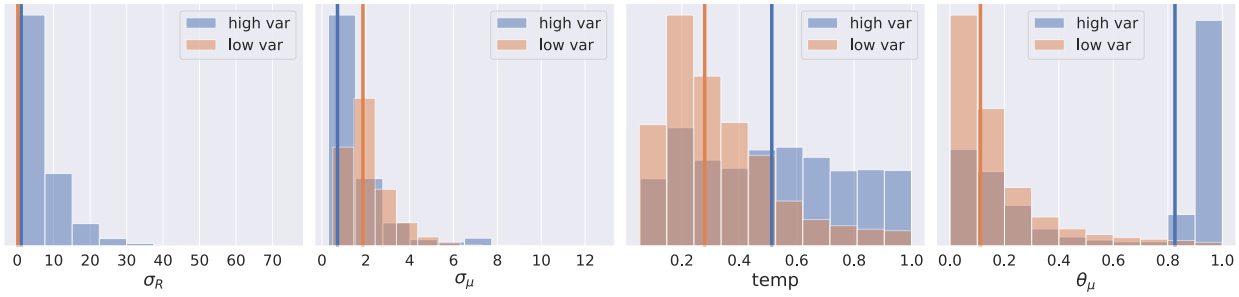
(a) Equal spikes: $\mu_1 = \mu_2 = 100$



(b) First spike higher: $\mu_1 = 150, \mu_2 = 100$



(c) Second spike higher: $\mu_1 = 100, \mu_2 = 150$



(d) Second spike higher: $\mu_1 = 100, \mu_2 = 300$

Figure 5.4: Selected parameters posterior distributions for the simulation scenario with two importation events, first at the beginning of the epidemic (increasing R_t) and second one during the second increase in R_t, but when incidence is close to 0.

increase in the number of observed cases to the sudden peak in R_t. Moreover, the model overestimates the total number of importations following the actual importations events, that is throughout the days 10-30

and after the μ_2 spike on day 40. Apart from the two sudden spikes in R_t, the R_t gets underestimated elsewhere, which implies most of the incidence is fuelled by the importations, rather than community transmission. Although it is not fully clear why the high variance model tackles well the case where $\mu_2 = 150$ but fails when $\mu_2 = 300$ (rows c and d), we hypothesise it is due to the fact that with this many new cases observed in the data at $t = 40$, the model considers a more plausible explanation the increase in R_t rather than importations. This however requires further investigation.

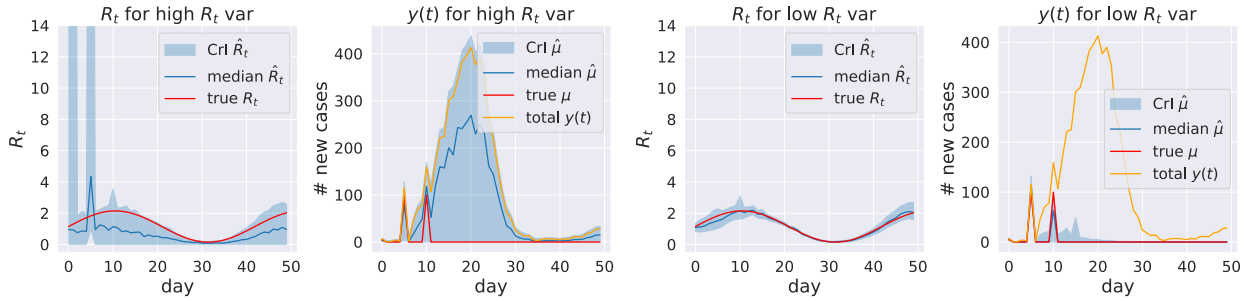
For that same example (Fig. 5.2 d), the model with lower R_t variance still fit the median posterior correctly.

In the simulations shown in Fig. 5.2, both importation events happen when the incidence is close to 0, which means that at the time of the importation, μ_1 and μ_2 will contribute to most (if not all) of the total cases. Because of that, it should be fairly easy for the model to correctly identify the importation rather than contributing the increase in observed cases to the rise in R_t. Still, we show that even in this simple example the prior selected for the R_t variance can have a big impact on the correct inference, which is especially pronounced in the last example in the figure. Note, that even with a lower variance prior for R_t, we still observe significantly wider credible intervals following importation spikes. This underscores the challenging nature of correctly identifying importations, even in this simplified scenario.

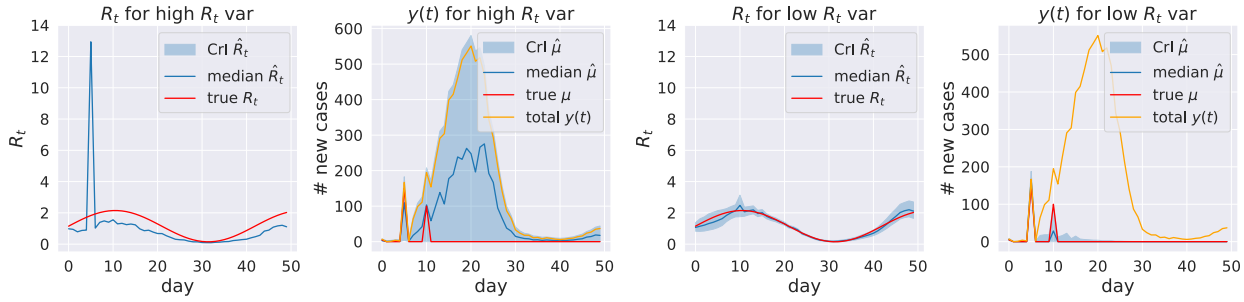
Second spike when R_t increasing

In the second variation of the two spikes scenario, we set the two importations events to take place at the beginning of the epidemic, during the increasing R_t, that is on days 5 and 10. Again, we varied the number of imported cases (μ_1 and μ_2) at each of the events. The results of fitting the model to those examples are shown in Fig. 5.5, the errors in Fig. 5.6, and the parameters posteriors in Fig. 5.7. The distinction between these examples and those in the previous subsection lies in the timing of the second importation event, occurring at a point when there are already more cases emerging from community transmission, as opposed to when incidence rates are close to zero.

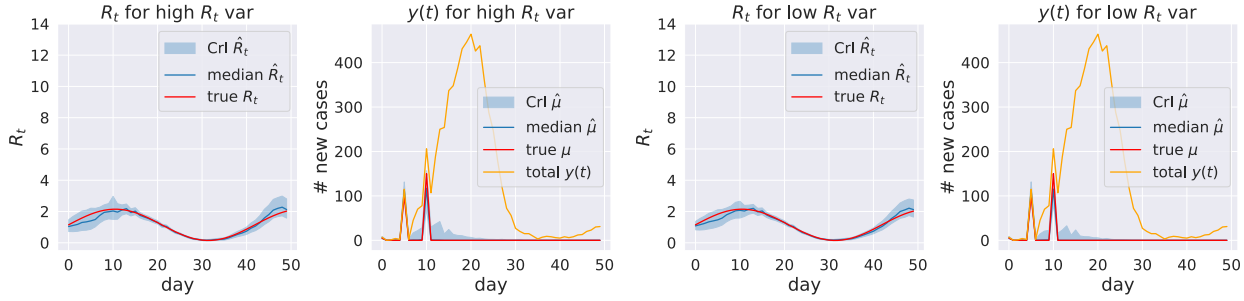
As we can see in Fig. 5.5, in the first two examples (a and b), when the R_t variance prior is high, the model fails to identify the importation events completely. R_t is massively overestimated around the time of the first spike, followed by the R_t being underestimated for the remainder of the simulated epidemic. Because R_t is underestimated, the model assumes that most observed cases are a result of importations, which we can observe in the plots in the second column, where the estimated curve of $\hat{\mu}$ is closer to the



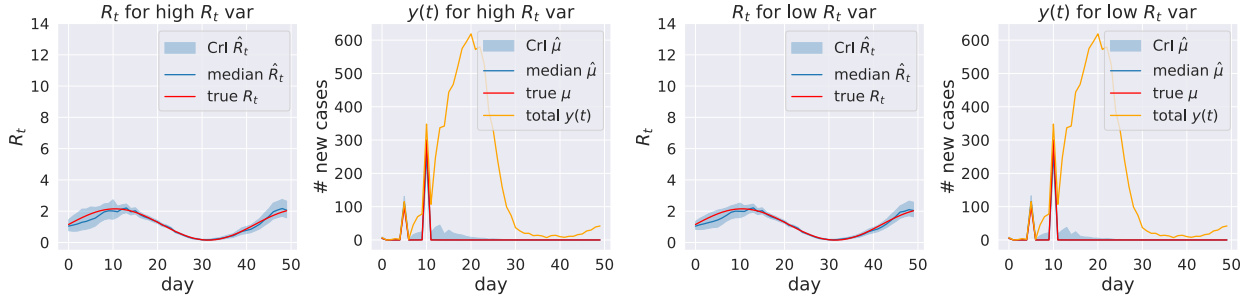
(a) Equal spikes: $\mu_1 = \mu_2 = 100$



(b) First spike higher: $\mu_1 = 150, \mu_2 = 100$



(c) Second spike higher: $\mu_1 = 100, \mu_2 = 150$

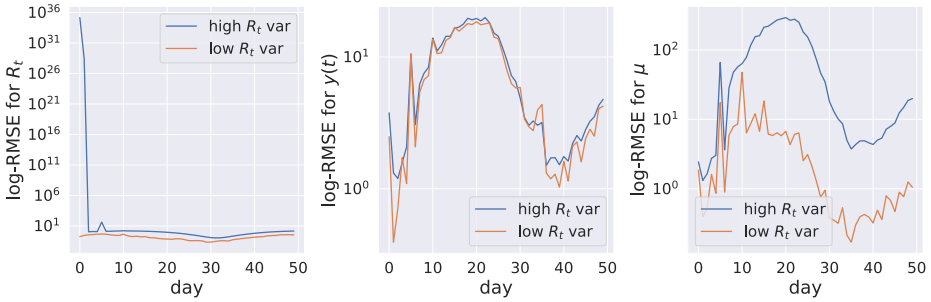


(d) Second spike higher: $\mu_1 = 100, \mu_2 = 300$

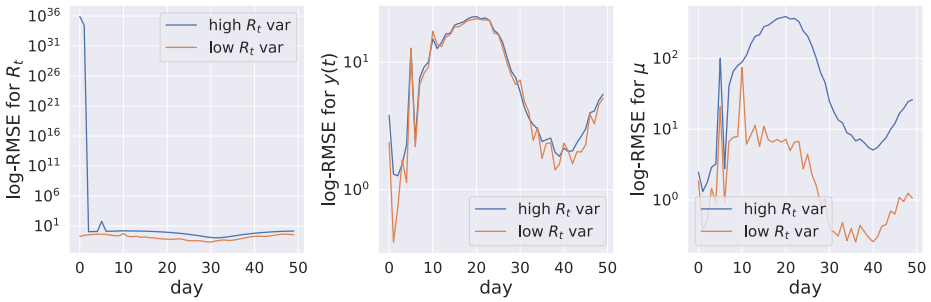
Figure 5.5: Simulation scenario with two importation events, both at the beginning of the epidemic (increasing R_t).

true incidence curve.

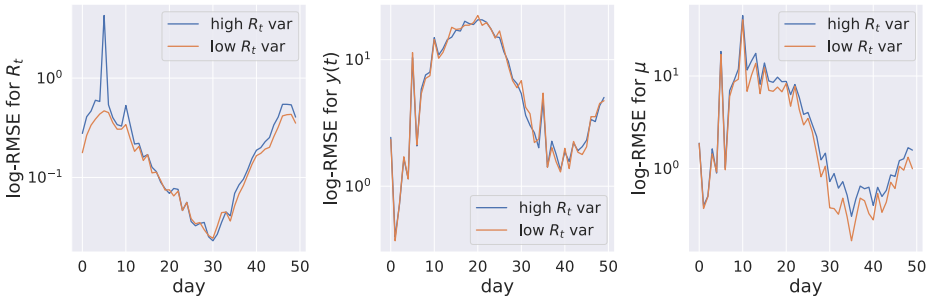
The same high R_t variance model manages better when the second importation spike is higher (examples



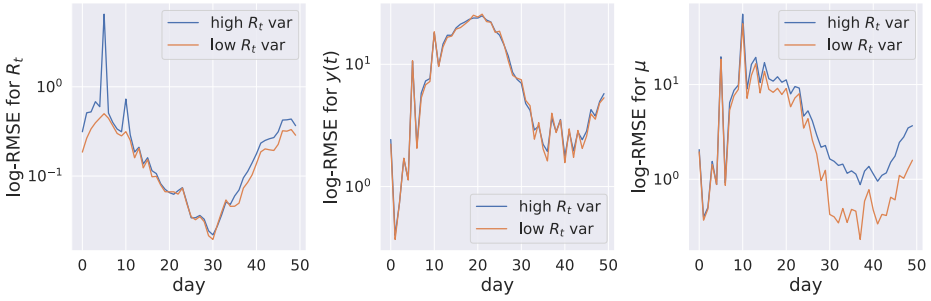
(a) Equal spikes: $\mu_1 = \mu_2 = 100$



(b) First spike higher: $\mu_1 = 150, \mu_2 = 100$



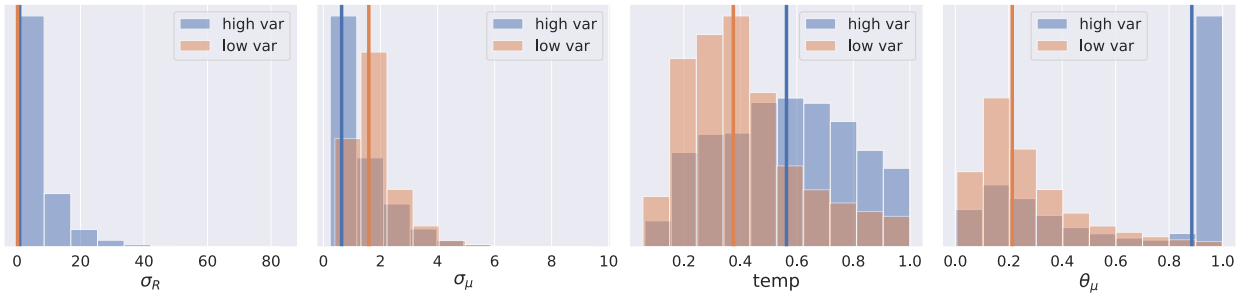
(c) Second spike higher: $\mu_1 = 100, \mu_2 = 150$



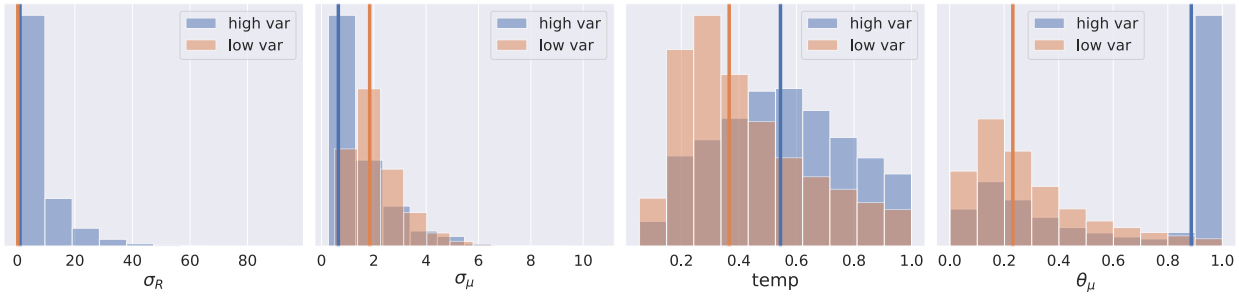
(d) Second spike higher: $\mu_1 = 100, \mu_2 = 300$

Figure 5.6: RMSEs for the simulation scenario with two importation events, both at the beginning of the epidemic (increasing R_t).

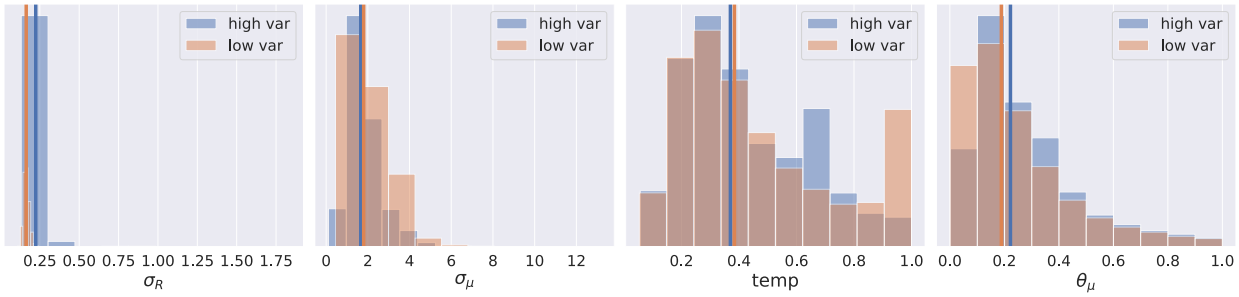
c and d). This is likely because in those examples, at the time of the second spike, μ_2 contributes to a substantial fraction of all new cases on that day, which makes it easier for the model to correctly identify the cases as imported (as shown in Fig. 5.2).



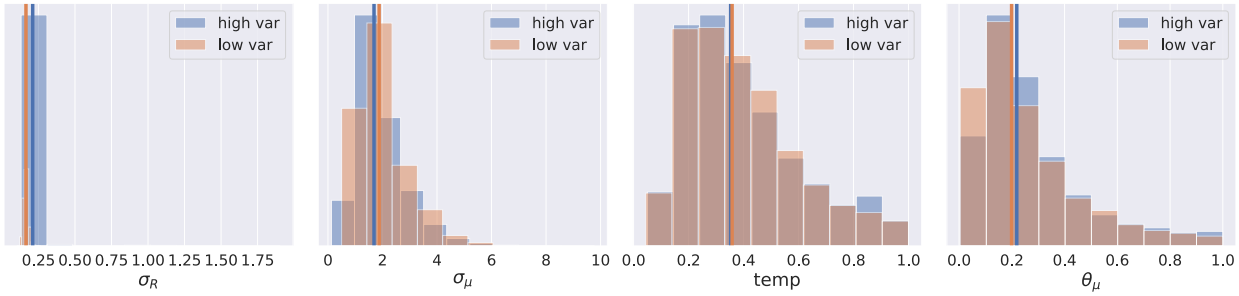
(a) Equal spikes: $\mu_1 = \mu_2 = 100$



(b) First spike higher: $\mu_1 = 150, \mu_2 = 100$



(c) Second spike higher: $\mu_1 = 100, \mu_2 = 150$



(d) Second spike higher: $\mu_1 = 100, \mu_2 = 300$

Figure 5.7: Selected parameters posterior distribution for the simulation scenario with two importation events, both at the beginning of the epidemic (increasing R_t).

The model with lower R_t variance again managed much better to recover the true curves. In all examples, the time of occurrence of the importations and their quantity were correctly identified. Although, even in these cases some of the posterior samples are still overestimating μ following the second spike. Despite

this, the posterior median presents a much better fit to the true data than the model with high variance.

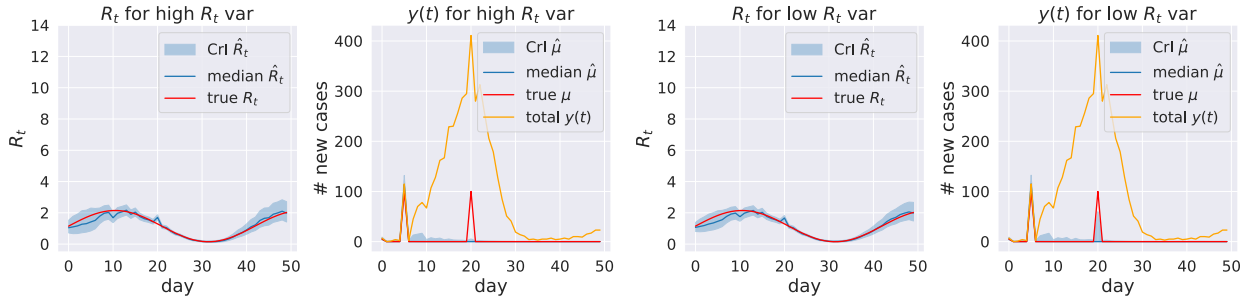
Second spike when R_t decreasing

In the third and final variation of the two spikes scenario, we kept the first importation event taking place at the beginning of the epidemic on day 5, during the increasing R_t, but we set the second one to occur on day 20 when R_t is decreasing. The results are shown in Fig. 5.8, the errors in Fig. 5.9, and the parameters posteriors in Fig. 5.10.

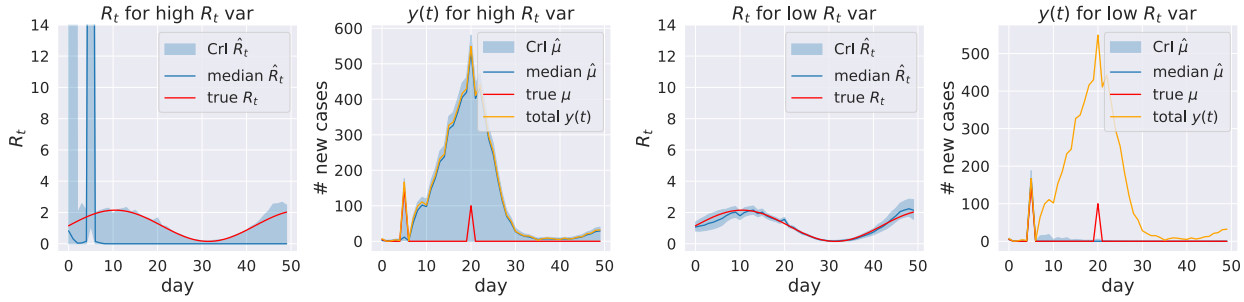
In three of the four examples, the high variance model attributes the spike in cases to the rise in R_t, causing a large overestimation of R_t around day 5 (time of the first importation event). In examples 5.8b and 5.8d, the R_t after day 5 get estimated as approximately 0, and all new cases are identified as importations throughout the time considered. Here, both importation peaks are ignored by the model. In both of those cases, the estimated $\hat{R}_t$ has the highest value reaching over 14, which is unrealistic for most of the pathogens. There was only one example, shown in Fig. 5.8a, in which the true R_t curve is not very far from the model's estimate, however even in that case only the first importation event is correctly picked up by the model, and the increase in cases around the time $t = 20$ is attributed to a short and small increase in R_t.

For the model with low variance, the estimated R_t curve closely follows the true curve, with a small bump in the posterior after the second importation event at $t = 20$. In all 4 examples, the first importation spike is correctly identified and estimated. Despite this, the model struggles to pick up the second spike, and instead, we can observe the bump in the estimate of R_t. The low R_t variance model manages to estimate the second spike correctly only in the last example, where $\mu_2 = 300$. In the other examples, the posterior median $\hat{\mu}$ is completely flat at the second importation peak, although in examples 5.8a and 5.8c the CrI covers the true peak.

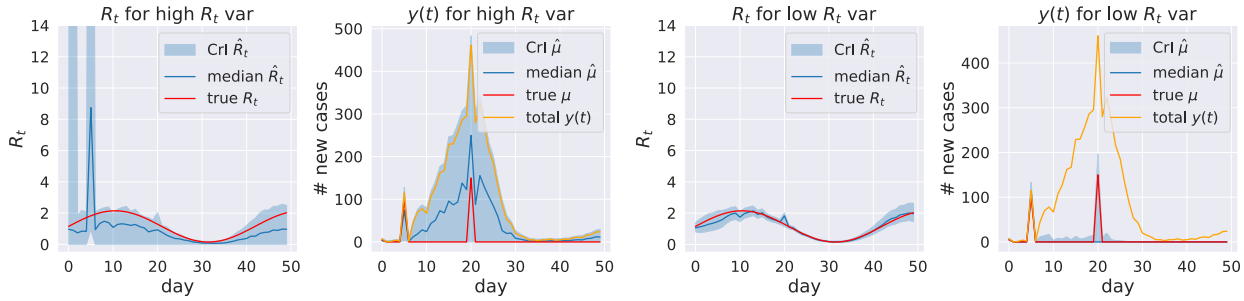
For both models correct identification of the second importation event was challenging. That is because the importation takes place when the community cases are still very high, following an increase in R_t, so even though R_t is decreasing at the time of μ_2, the numbers of imported cases are a smaller fraction of all new cases.



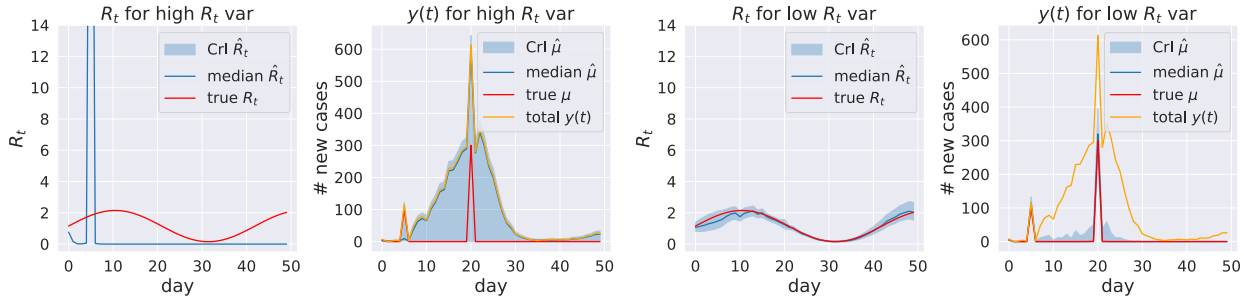
(a) Equal spikes: $\mu_1 = \mu_2 = 100$



(b) First spike higher: $\mu_1 = 150, \mu_2 = 100$

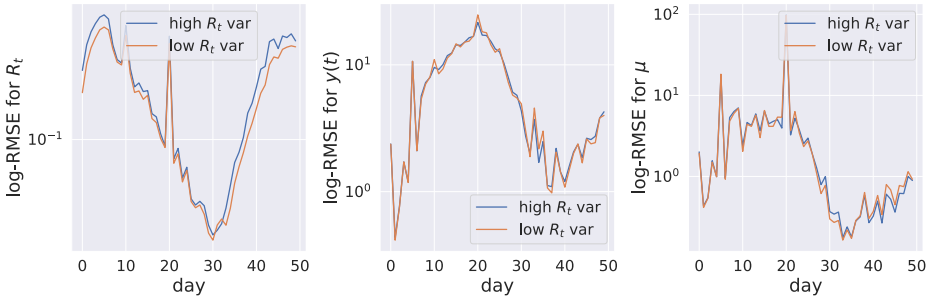


(c) Second spike higher: $\mu_1 = 100, \mu_2 = 150$

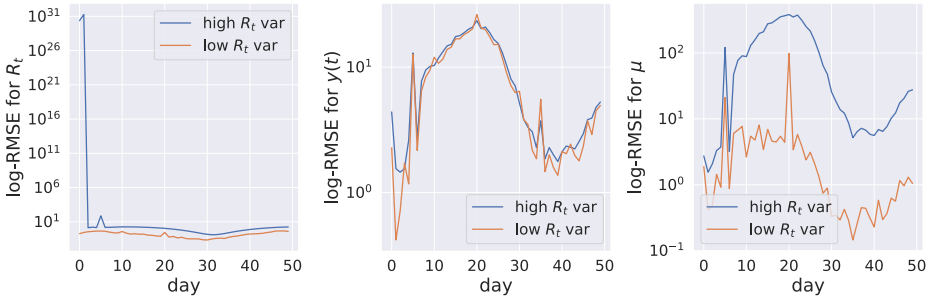


(d) Second spike higher: $\mu_1 = 100, \mu_2 = 300$

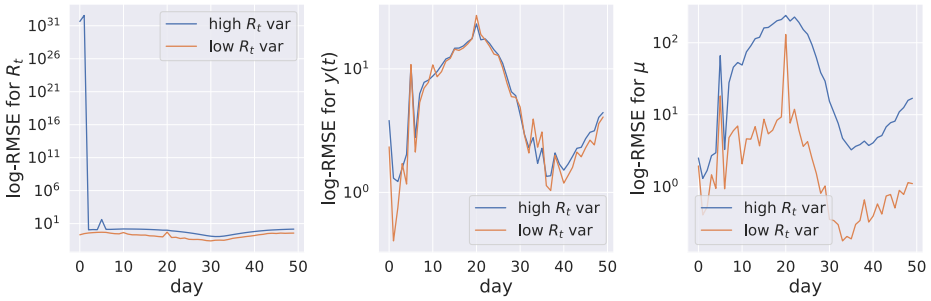
Figure 5.8: Simulation scenario with two importation events, the first one in the beginning of the epidemic (increasing R_t) and the second one in the middle (decreasing R_t).



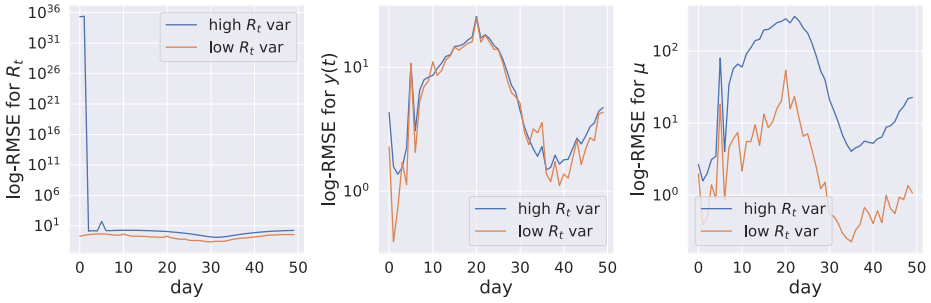
(a) Equal spikes: $\mu_1 = \mu_2 = 100$



(b) First spike higher: $\mu_1 = 150, \mu_2 = 100$

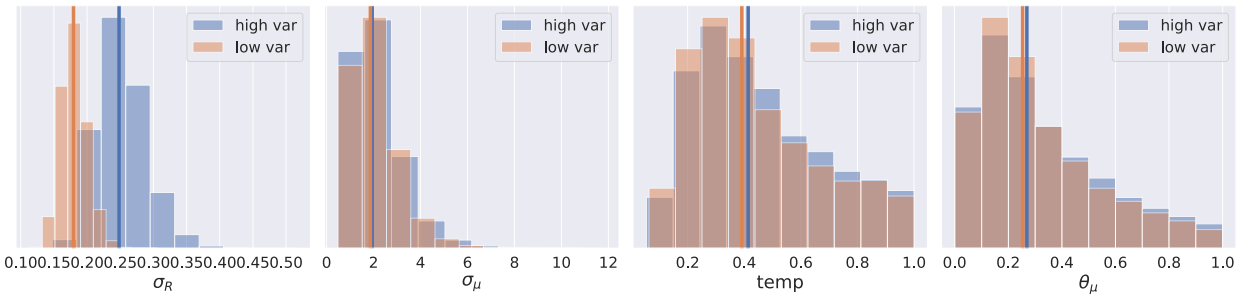


(c) Second spike higher: $\mu_1 = 100, \mu_2 = 150$

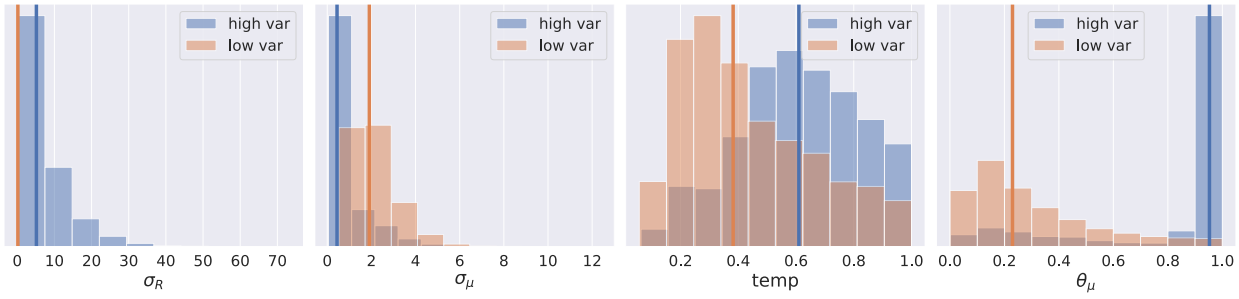


(d) Second spike higher: $\mu_1 = 100, \mu_2 = 300$

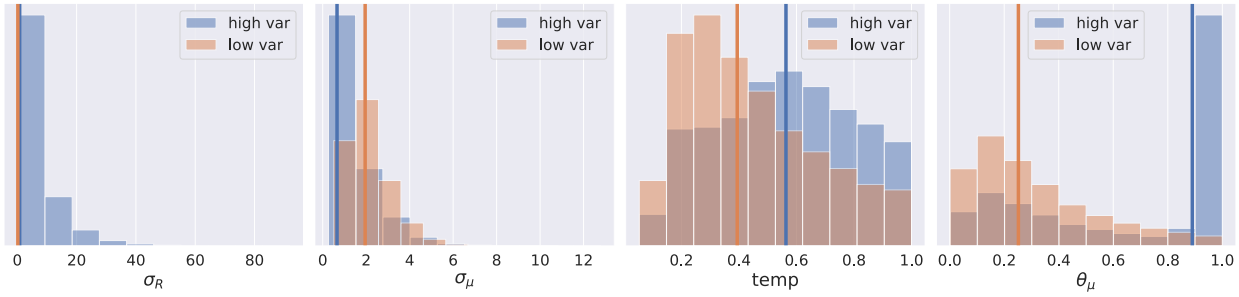
Figure 5.9: RMSEs for the simulation scenario with two importation events, the first one in the beginning of the epidemic (increasing R_t) and the second one in the middle (decreasing R_t).



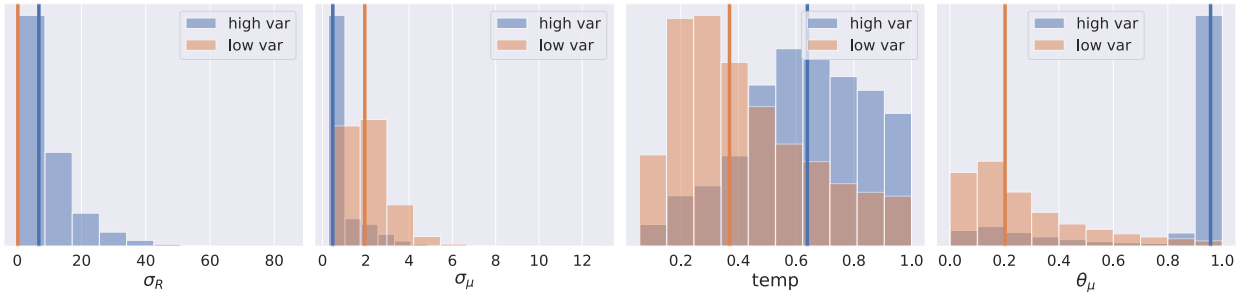
(a) Equal spikes: $\mu_1 = \mu_2 = 100$



(b) First spike higher: $\mu_1 = 150, \mu_2 = 100$



(c) Second spike higher: $\mu_1 = 100, \mu_2 = 150$



(d) Second spike higher: $\mu_1 = 100, \mu_2 = 300$

Figure 5.10: Selected parameters posterior distribution for the simulation scenario with two importation events, the first one at the beginning of the epidemic (increasing R_t) and the second one in the middle (decreasing R_t).

Converge diagnostics

In all scenarios considered so far, the RMSE for the observed incidence and the incidence estimated by the model was quite low. However, despite the high number of iterations, the models were run for, the convergence was not satisfactory for many of the parameters of the models. Table 5.1 shows the ranges of the $\hat{R}$ diagnostic for all considered scenarios for both high- and low- R_t variance models. We can see that the high variance model was less successful in achieving the convergence, which we could also observe in the wide estimates of the CrI for R_t for example in Fig. 5.8 b-d.

Table 5.1: Ranges of $\hat{R}$ diagnostics for the different models parameters in considered scenarios.

second spike when R_t ..	high R_t var	low R_t var
..increasing but inc ≈ 0, equal spikes	1.00 -1.02	1.00 - 1.02
..increasing but inc ≈ 0, first spike higher	1.00 - 1.00	1.00 - 1.04
..increasing but inc ≈ 0, second spike higher, $\mu_2 = 150$	1.01 - 1.22	1.00 - 1.04
..increasing but inc ≈ 0, second spike higher, $\mu_2 = 300$	1.03 - 10.08	1.00 - 1.08
..increasing, equal spikes	1.03 - 6.69	1.00 - 1.30
..increasing, first spike higher	1.01- 7.52	1.00 - 1.04
..increasing, second spike higher, $\mu_2 = 150$	1.00 - 1.17	1.00 - 1.27
..increasing, second spike higher, $\mu_2 = 300$	1.00 - 1.04	1.00 - 1.14
..decreasing, equal spikes	1.00 - 1.05	1.00 - 1.05
..decreasing, first spike higher	1.03 - 6.41	1.00 - 1.07
..decreasing, second spike higher, $\mu_2 = 150$	1.03 - 6.50	1.00 - 1.03
..decreasing, second spike higher, $\mu_2 = 300$	1.00 - 2.82	1.00 - 1.03

5.3.2 Scenario 2: seasonal importations

In the second simulation scenario, we explore seasonal importations. In this setup, a certain number of new cases are imported into a location daily, where the pathogen has not yet established circulation. Such occurrences may occur, for instance, when students return to their hometowns for holidays.

In this set of examples, the simulations included seasonal importations instead of single-day influxes of cases. That is, the importations were generated by setting

$$\mu_i = 100 * (1 + \sin(\frac{2\pi}{n/2} * i)) \tag{5.9}$$

Here, we used a sine function to generate daily new importations with seasonal patterns. In the equation above, n denotes the total number of days considered in the analysis, and factor $n/2$ ensures that there

are two periods of the sine function included.

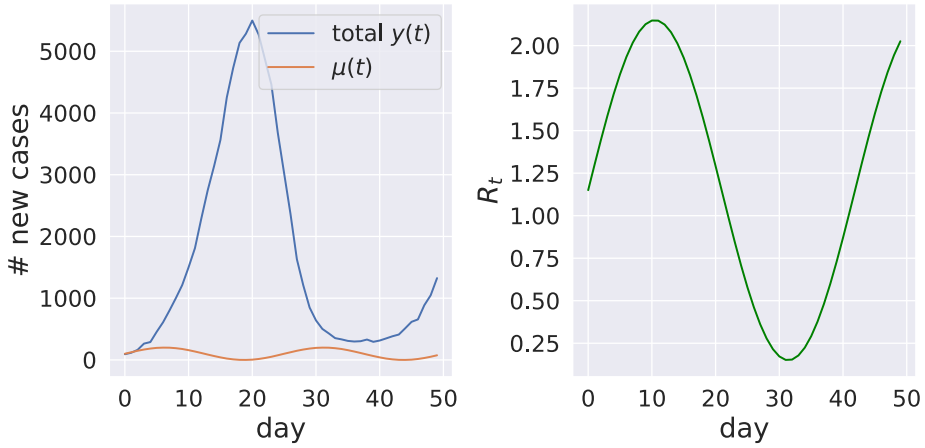
The examples of generated data are shown in Fig. 5.11

In Fig. 5.12 we show 3 examples of fitting the previously proposed model to the generated seasonal importations. Identifying importations using the same model as presented before, that is with GRW for μ, was unsuccessful regardless of the chosen priors. One reason for that could be that the incidence rises extremely fast in this case of new imported cases coming into the system every day, and then the new importations become a small fraction of all new infections (see Fig. 5.11), in which case our proposed methodology fails, as shown in the previous examples.

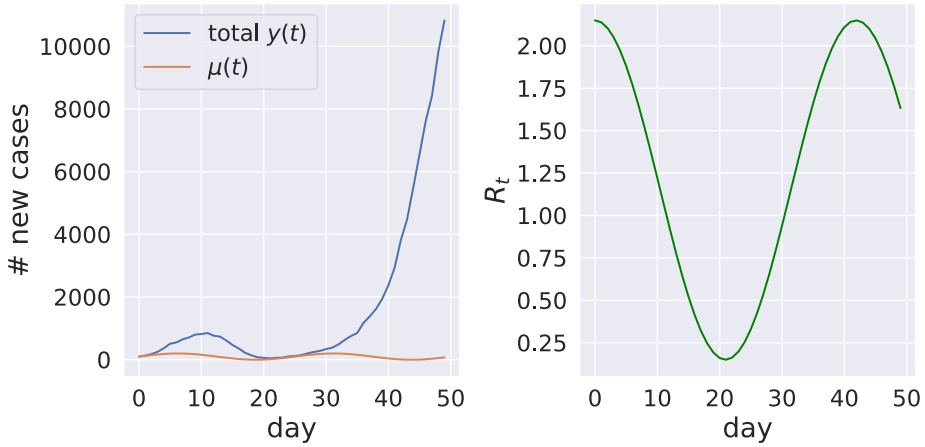
One solution is to include an explicit seasonal importations component in the model. The change in the model fitted to the simulated data is only in the μ component, and its modified version is as follows:

$$\begin{aligned} \sigma_\mu &\sim \mathcal{N}_+(1, 10) \\ \theta_\mu &\sim \mathcal{N}_+(1, 100) \\ \mu_t &= \theta_\mu * (1 + \sin(\frac{2\pi}{n/\sigma_\mu} * t)) \end{aligned} \tag{5.10}$$

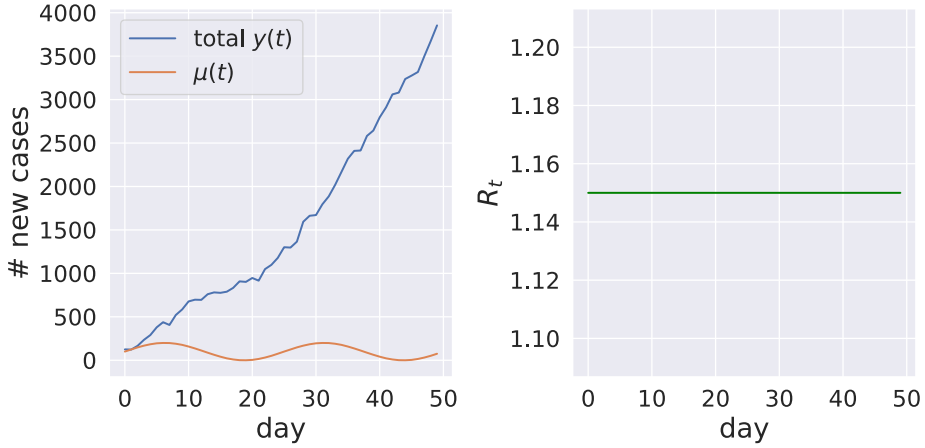
In Fig. 5.14 we show 3 examples of fitting the model with an explicit seasonal component for μ to the generated seasonal importations. As we can see, the new seasonal model is more accurate than the GRW model and manages to capture the seasonal behaviour. Even though it estimates R_t correctly, the importations are well estimated only in the case of constant R_t. Again, for the seasonal model, the choice of the prior for the R_t variance does not seem to have a significant impact on the results of the simulations.



(a) $R_t = sin()$

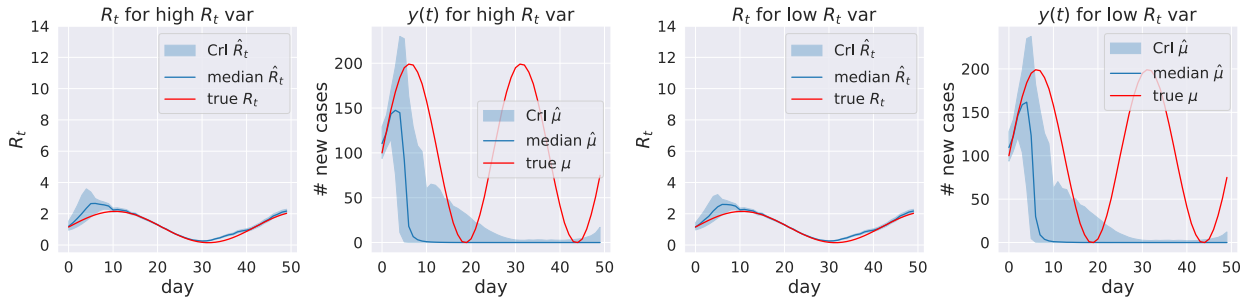


(b) $R_t = cos()$

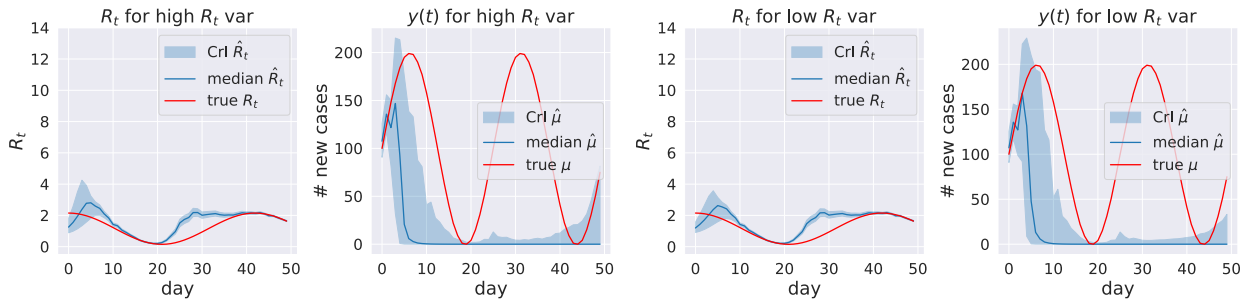


(c) $R_t = const$

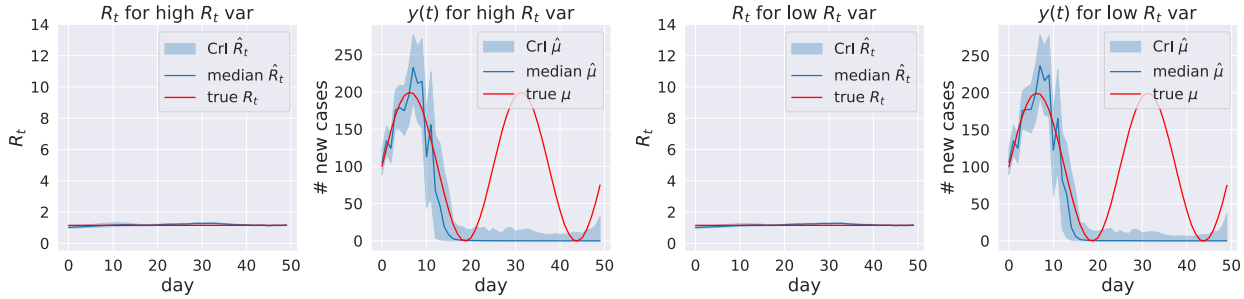
Figure 5.11: Simulation scenario with seasonal importations, incidence and R_t. Because of a high number of cumulative importations, the total incidence increases much faster than in the previous scenarios with singular importation events.



(a) $R_t = sin()$, μ inferred as GRW

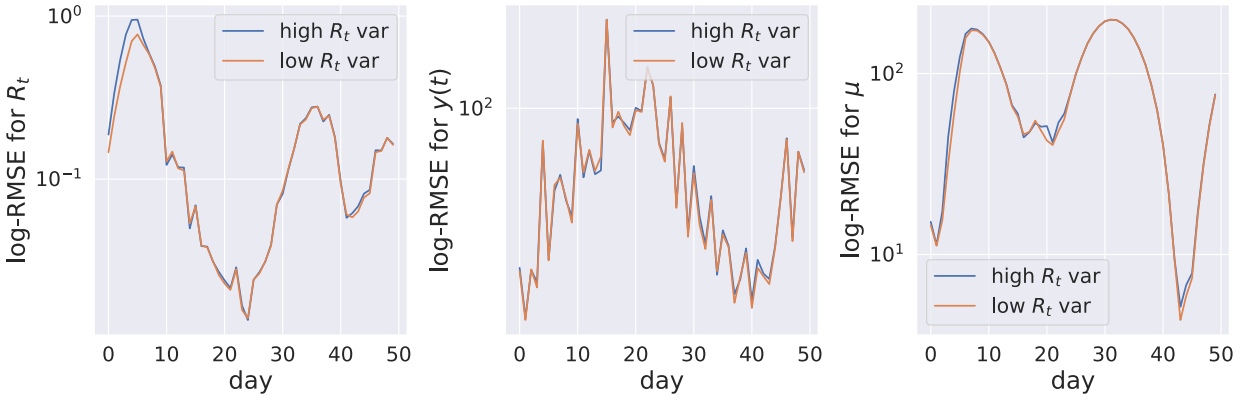


(b) $R_t = cos()$, μ inferred as GRW

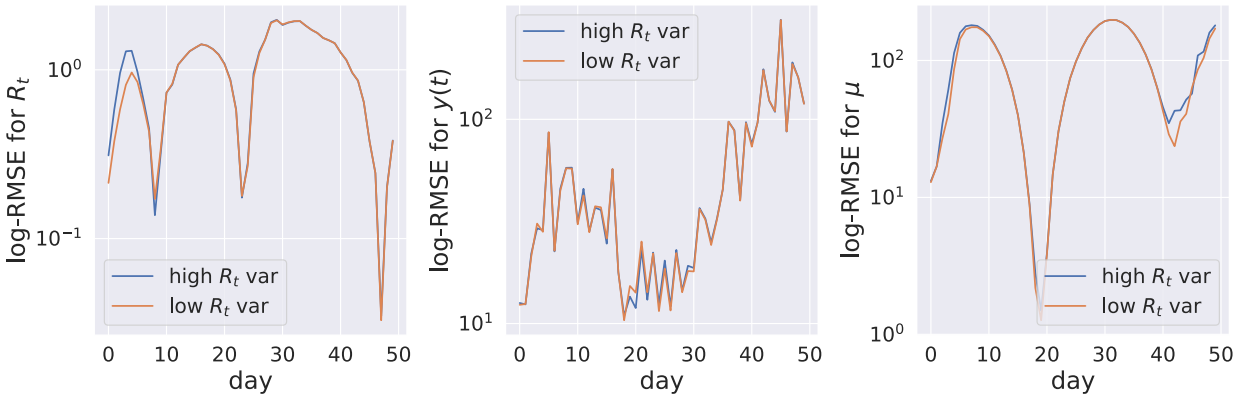


(c) $R_t = const$, μ inferred as GRW

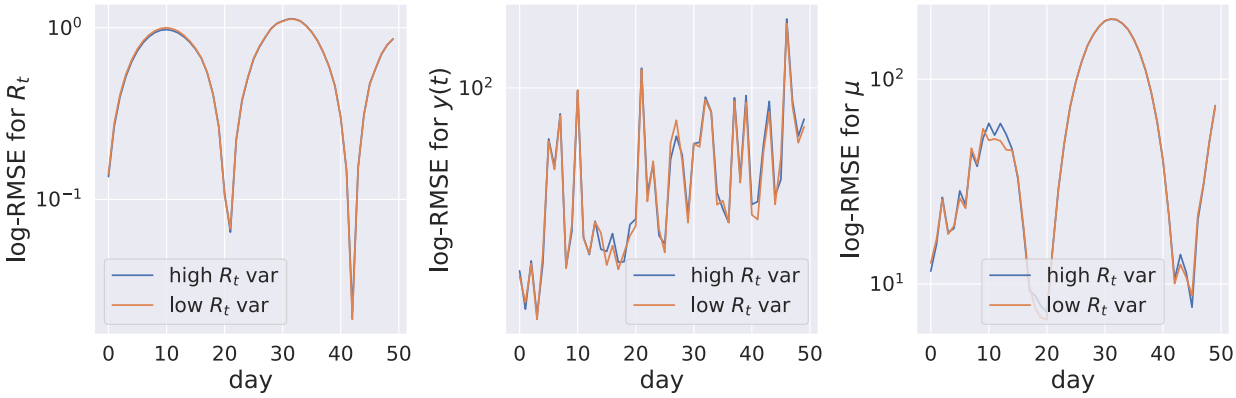
Figure 5.12: Simulation scenario with seasonal importations, where fitted model assumes μ is a GRW. Total incidence is not shown, as it is much higher than the numbers of importations, but can be seen in Fig. 5.11.



(a) $R_t = sin()$, μ inferred as GRW

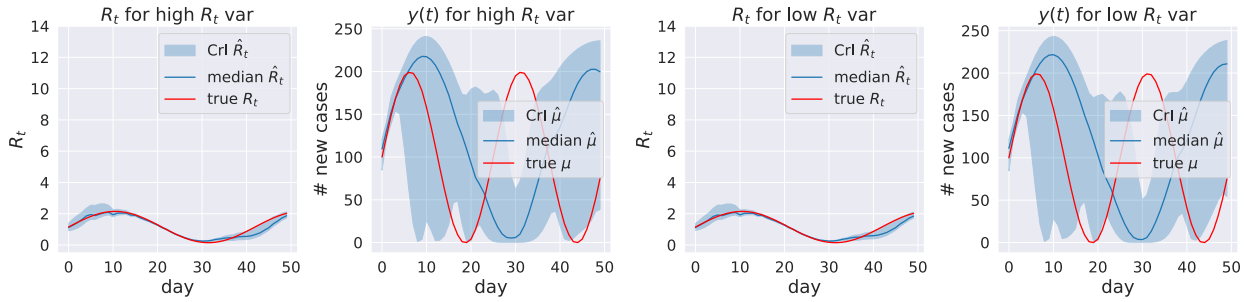


(b) $R_t = cos()$, μ inferred as GRW

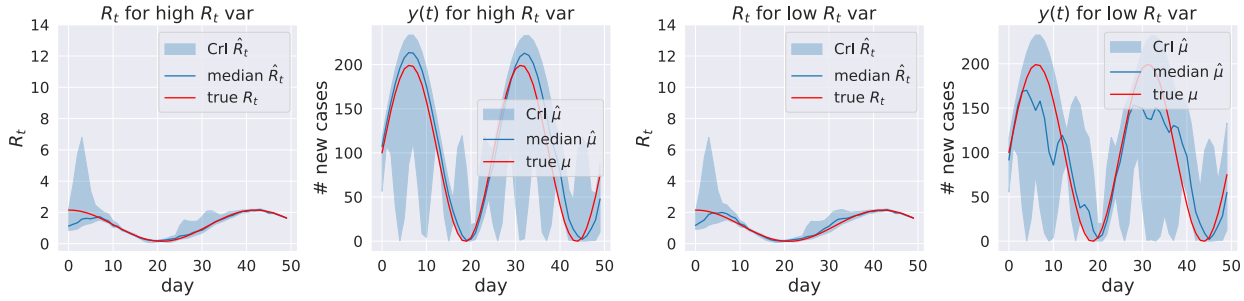


(c) $R_t = const$, μ inferred as GRW

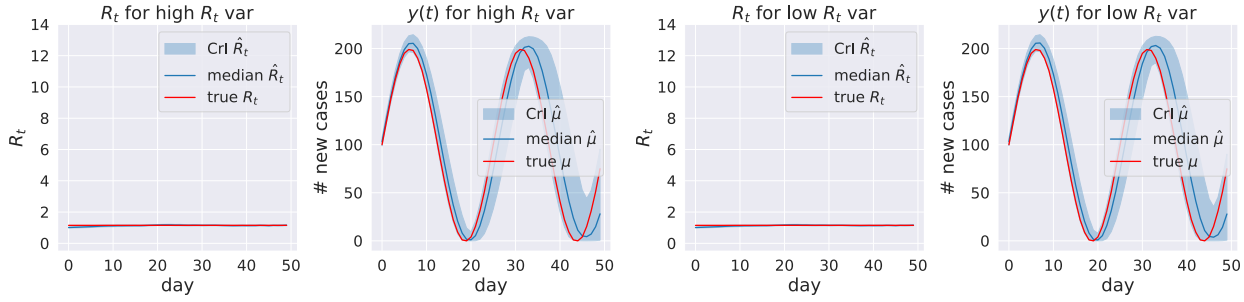
Figure 5.13: RMSEs for the simulation scenario with seasonal importations, where fitted model assumes μ is a GRW.



(a) $R_t = sin()$, μ inferred with explicit seasonality

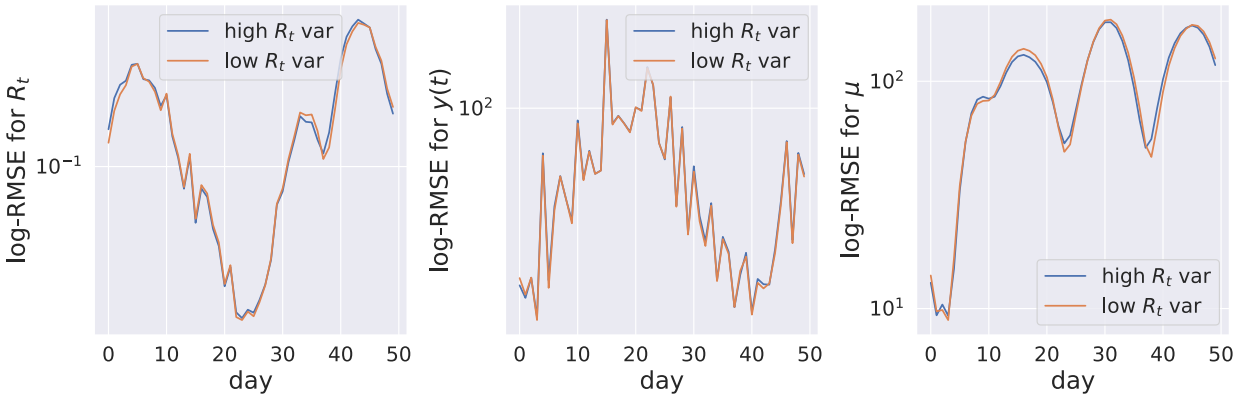


(b) $R_t = cos()$, μ inferred with explicit seasonality

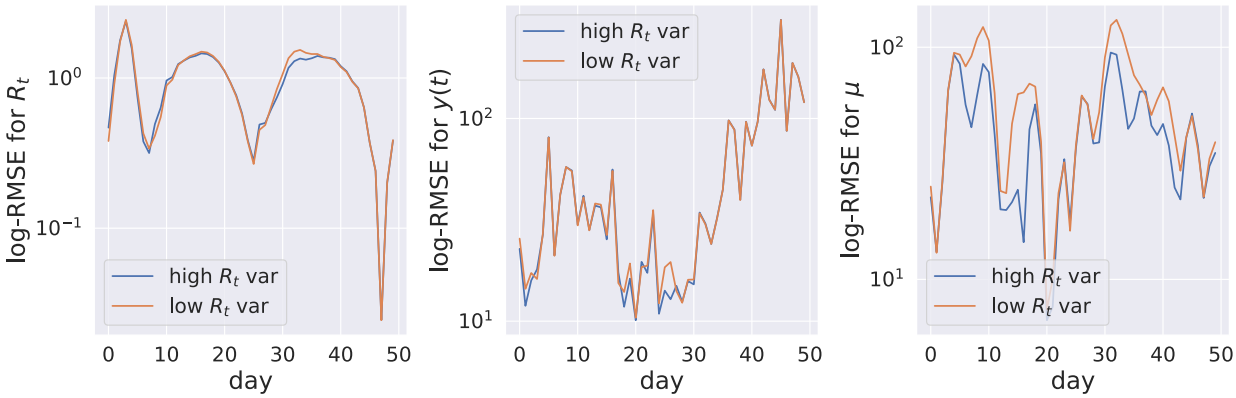


(c) $R_t = const$, μ inferred with explicit seasonality

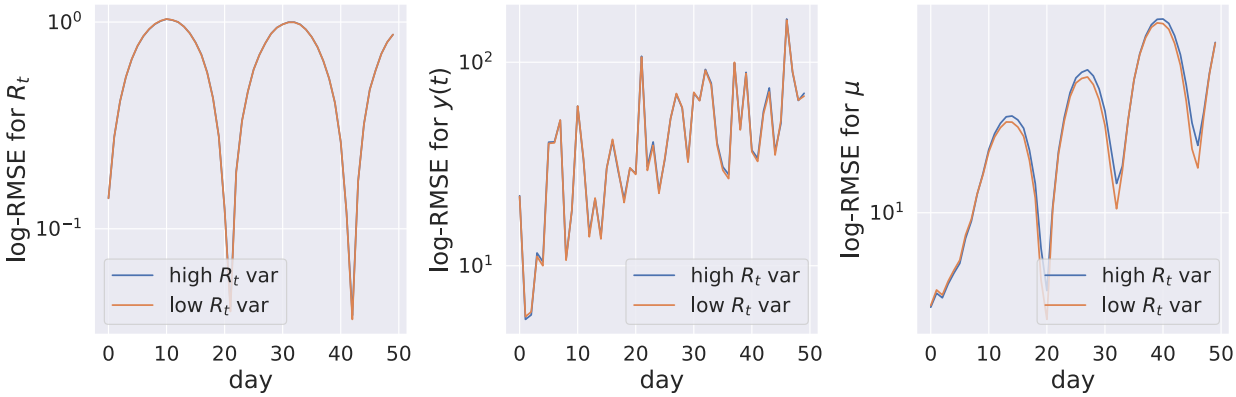
Figure 5.14: Simulation scenario with seasonal importations, where fitted model explicitly assumes μ is a seasonal and modelled with a $sin()$ function. Total incidence is not shown, as it is much higher than the numbers of importations, but can be seen in Fig. 5.11.



(a) $R_t = sin()$, μ inferred with explicit seasonality



(b) $R_t = cos()$, μ inferred with explicit seasonality



(c) $R_t = const$, μ inferred with explicit seasonality

Figure 5.15: RMSEs for the simulation scenario with seasonal importations, where fitted model explicitly assumes μ is a seasonal and modelled with a $sin()$ function.

5.4 Discussion

In this study, we investigated the impact of choosing the appropriate prior for the effective reproduction number R_t on the model fit for the renewal equation based epidemic model.

Summary of the findings

We empirically demonstrate that the selection of prior strongly influences the accuracy and alignment of epidemiological estimations with domain knowledge. It is crucial to underscore that in all scenarios considered, the model's fit to simulated observed data (daily new case counts) remained robust throughout the simulated epidemic period. Consequently, selecting the most suitable model based solely on fitting criteria becomes challenging (refer to Chapter 2). Although this project did not centre on a rigorous model selection process, our focus was on illustrating the pivotal role of prior choice in the renewal model of R_t through intuitive examples.

Implications for modelling

Our study holds several implications for infectious disease modelling. Firstly, we demonstrate that constraining the variance of R_t in the model often facilitates accurate inference of importation events. Specifically, when the prior variance of R_t is low, the model fitting process correctly identifies importations instead of attributing new cases to abrupt increases in R_t. This constraint aligns with epidemiological principles, as R_t typically exhibits minimal day-to-day variation, except in rare instances such as the implementation of restrictions (e.g. non-pharmaceutical interventions like business and school closures [Flaxman et al., 2020]), mass vaccination campaigns, or super-spreading events.

Accurately identifying importations at the onset of an epidemic holds significant importance for modellers and policymakers. This knowledge enables timely and targeted interventions to mitigate pathogen spread. For instance, optimal interventions may include travel restrictions, quarantine protocols, or health screenings upon arrival, rather than implementing broader measures such as lockdowns or business closures within the affected area [Reich et al., 2021, Wells et al., 2020].

Challenges and future work

In our study, we investigate the challenges associated with inferring importations in a discrete renewal equation based model, even in relatively straightforward simulated scenarios where μ constitutes a small fraction of new cases at time t. However, practical observations indicate that in such cases, the influence on accurately inferring R_t remains negligible provided that the variance of R_t is kept low (see Fig. 5.5). Consequently, in scenarios where μ is a minor contributor to new cases, the importance of accurately inferring the number of imported cases is arguably diminished, as most cases originate from community transmission. Thus, implementing travel restrictions in the area of interest may have a less pronounced impact on total incidence.

Our analysis is based on simulated data, encompassing various scenarios. However, there are additional potential scenarios worth investigating. For instance, while we focused on scenarios where the epidemic is in its early stages in the area of interest, exploring the identifiability of importations in ongoing epidemics could be informative. Nevertheless, our findings suggest that even with a more informative prior for R_t, the models we employed struggle to accurately capture importations when μ represents a minor fraction of all new cases. Furthermore, validating the method's performance on real datasets could offer valuable insights, although access to such data is often limited due to its nature.

Additionally, we opted for a simplistic parametrisation of R_t, contrasting with the more intricate models commonly utilised by epidemiological modellers, which incorporate disease- or population-specific dynamics (e.g. Ferretti et al. [2020], Flaxman et al. [2020], Kucharski et al. [2020]). Our rationale for this choice was to directly illustrate the potential impact of the priors selection on model fit. We aimed to address a broad audience by focusing on a general case applicable to various pathogens, emphasising that a model's fit to observed data does not necessarily guarantee accurate parameter inference. Domain knowledge remains essential for evaluating model performance and selecting the most appropriate model for further analysis.

5.5 Conclusions

In this chapter, we showed simulated examples of un-identifiability of importations in the renewal equation-based model of R_t. The study demonstrates that the choice of prior significantly influences the correct

inference of epidemiological quantities, emphasising the importance of aligning priors with domain knowledge. The findings suggest that constraining the variance of R_t aids in correctly inferring importations, aligning with epidemiological reasoning, and has implications for recommending optimal interventions in infectious disease modelling scenarios, such as travel restrictions and health checks.

In the next and final chapter, I reflect on the outcomes of the thesis and discuss their implications and limitations.

Chapter 6

Discussion

In this thesis, I have presented a suite of methods for infectious disease modelling. The methods I developed were necessitated by the ongoing COVID-19 pandemic and are focused on modelling emerging threats, although they can be applied to other scenarios, such as endemic diseases. The overarching theme of this thesis was to combine modern statistics and machine learning methods (e.g. Gaussian Processes) with more traditional approaches used in epidemiological modelling, such as renewal equation models.

In this final chapter, I will review the findings of the studies in this PhD thesis. I will also discuss their implications for infectious disease modelling, explore their limitations, and suggest directions for future research in the field.

6.1 Summary of findings

Chapter 2: Referenced Thermodynamic Integration

In Chapter 2 I proposed a method for calculating model evidence called referenced Thermodynamic Integration. Model evidence can be used to perform model selection in a fully Bayesian setting. The method is based on building a reference density, e.g. Gaussian, and then sampling from the geometric path between the reference and the original density. I presented the practical steps of applying the method to some pedagogical problems. I further presented the applicability of the method on a larger model of R_t for COVID-19 in South Korea, in which I also demonstrated the importance of performing model selection

rather than only relying on the model's fit to the observed data in order to pick the most reliable model.

The problem of model selection is ever-present in statistical modelling, and one that has been extensively studied. Still, it remains one of the most challenging aspects of data analysis [Jordan, 2011, Navarro, 2019]. Despite considerable efforts to develop robust methodologies and theoretical frameworks, there is a lack of consensus on which method of model selection is optimal, and the choice of one typically depends on the task at hand (prediction or explainability) as well as the practitioners' personal preference. In this chapter, I focused on only one of the available methods, Bayes factors, as it is rarely used in epidemiology, possibly due to its complexity. One of the main purposes of this project was to showcase the calculation of normalising constant in a very practical setting, using a method that could be relatively easily applied to a variety of problems, as its main step is to approximate the original density with a Gaussian one.

As I showed in the COVID-19 example in Section 2.3.4 and also in Chapter 5, it is not enough to validate the model's accuracy only by looking at its fit to the observed data. Rigorous model selection can help us prevent situations where even though the fit is correct, the estimates of the individual parameters remain unreasonable, e.g. implying large fluctuations in R_t, which are uncommon in reality.

The comprehensive Bayesian model selection methods, including referenced Thermodynamic Integration, still suffer from a range of issues, making it discouraging for practitioners to apply in their studies. Mainly, as they focus on calculations pertaining to full posteriors distributions of samples, they are significantly more computationally complex than simple, point-estimate based methods such as e.g. AIC or BIC. Although with the advances in computing these issues will gradually become less problematic, for this moment they still persist. Another major critique of the Bayes factors is their sensitivity to the priors. Many Bayesians argue that this is in fact a desired property, as the priors are an inherent part of any Bayesian model and should be taken into consideration when comparing different models. Navarro, 2019 also gives an example in which Bayes factors tend to give counterintuitive and wrong answers when both of the compared models are 'extremely bad'.

Moreover, model selection methods can compare different models and 'rate' them against one another. They are not however a method of model assessment, that is they do not tell us anything about an individual model. So even after fitting many models and comparing them using state-of-the-art methods to pick 'the best' one, the favoured model is always 'the best out of the models we tested'. We cannot ever know whether we could come up with a different model better explaining the phenomenon or a simpler

model.

We showed, that the referenced Thermodynamic Integration method for calculating model evidence has a faster convergence than competing methods like Power Posteriors. It still however requires fitting multiple models using the path between the original and reference density, and therefore a substantial computational power. The calculations for different λ placements can be parallelised as they do not depend on one another, to shorten the total compute time. For complex multidimensional models, even sampling from a single λ-path will still require a certain amount of time and resources. This could be further improved using the increasingly popular variational or approximate methods of inference, or normalising flows models.

The Bayesian modelling approach is increasingly favoured among infectious disease modellers (e.g. Aswi et al. [2019], Bhatt et al. [2015], Davies et al. [2021a], Faria et al. [2021], Flaxman et al. [2020]), owing to its natural ability to incorporate prior domain knowledge and accurately quantify uncertainty. A meticulous and informed model selection process should be integral to any Bayesian analysis. In epidemiological modelling, model selection extends beyond the final stages of analysis, where complex models are utilised to infer quantities such as R_t. It also encompasses intermediate steps, such as the selection of appropriate prior distributions for parameters like the generation interval or onset-to-death times, as demonstrated in the following chapter.

Chapter 3: Hospitalisation Distributions

The referenced Thermodynamic Integration method developed in Chapter 2 was further used in Chapter 3, where I fitted a range of probability density functions to COVID-19 hospitalisation distributions in Brazil. Typically, the densities used for estimating onset-to-death time, an important quantity in infectious disease modelling, are Gamma or Log-Normal. In this chapter, I showed that in some cases more flexible densities, such as generalised-Gamma, provide a better fit to the underlying data. I also give a framework for estimating those distributions on both national- and state-level, using a hierarchical Bayesian model. Through this, I showed that the estimated mean times of onset-to-death, admission-to-death and others are highly spatially heterogeneous. By applying a hierarchical framework, I was able to estimate those quantities even for states where limited data was available at the time of the study.

The epidemiological parameters estimated in that chapter are fundamental for understanding an ongoing epidemic. For the modellers, they are particularly important for estimating the reproduction number

and in the modelling of possible effects of interventions. Although paramount for the modelling, those parameters are also difficult to obtain at the early stages of the epidemic — for example, a widely cited paper by Verity et al. 2020 estimates the onset-to-death time in mainland China using data of only 24 deaths and onset-to-recovery using 169 cases outside the mainland China, as that was all that was available and reliable at the time of the study. Both of these estimates were used for estimating the Case Fatality Ratio, which was vital in those early stages of the epidemic not only for the modellers worldwide but also for policymakers and in public health campaigns.

However, these estimates were based on the very limited spatial region and no estimates were available for low- and middle-income countries (LMIC). As the data from countries around the world was extremely limited and the epidemic had not fully spread worldwide yet, the spatial heterogeneity of those distributions could only be hypothesised. In this study, I was able to leverage the extensive publicly available hospitalisation dataset for Brazil, to provide those estimates based on a substantially larger dataset, including large hospitals and small health centres, and spanning across the entire Brazil. This enabled us to get a grasp of the spatial heterogeneity of the character of infections in a LMIC. Moreover, even though I used a hierarchical Bayesian model and the size of the dataset was large, the approach I presented is fast and can be applied near real-time in the future epidemic.

As with many other studies, this approach is prone to biases in the data, which still need careful consideration. One of the main issues with treating the early-epidemic dataset is the censoring effect — if we correct for it, we are likely to have very few data points left, and if we do nothing, we are likely to arrive at the estimates of times shorter than they would be in reality. I touched upon these issues in the chapter's sensitivity analysis, but in the future, this approach should be further analysed and focused on those arising biases. Another possible bias in the data was the under-reporting of deaths and the reporting delays, which was the topic of the following chapter.

Chapter 4: Gaussian Process Nowcasting

Shifting the focus from the model properties to the data, in Chapter 4 I developed a method of correcting the reporting delays based on nowcasting using latent Gaussian Processes (GPs). This method is disease-agnostic and versatile, so it can be used in any setting where the stream of data suffers from reporting delays. In this project, I applied it to the weekly records of deaths due to COVID-19 in Brazilian hospitals.

GPs provide a flexible way of modelling the delays across the two dimensions of a reporting triangle, allowing us to estimate the numbers of deaths that occurred but have not been yet reported. The newly proposed method has been validated against a bench-marking method and provided favourable outputs both in terms of the overall accuracy and the estimate of the uncertainty of the results.

The delay in death reporting, an issue tackled in this chapter, is only one of many problems with the real-time data the modellers were faced with during the COVID-19 pandemic. The surveillance of cases and deaths due to the SARS-CoV-2 infections has been a tough challenge from the early stages of the pandemic when doctors and scientists worldwide faced the new pathogen, for which the diagnostic has not been fully developed and hence testing required time-consuming PCR [Msemburi et al., 2023, Wang et al., 2022]. This has not been available in every country due to various socio-economic reasons. Additionally, the standards of administration, including registering patients and issuing death certificates, varied from country to country and possibly even within the countries [Msemburi et al., 2023]. That became apparent to the scientists early on in the pandemic, and various methods of looking at the excess deaths instead of the reported COVID-19 deaths have been proposed. In Brazil, the SIVEP-Gripe database had been in place before the pandemic for recording patients with severe acute respiratory illness, including all health facilities and hospitals across the entire country. Because of that, the extent of the excess, underreported deaths in Brazil, was likely to be lower than in some other countries such as e.g. India, where the estimated excess deaths were huge [Jha et al., 2022, Msemburi et al., 2023]. But even with this infrastructure in place, the reporting delays remained a major problem, especially at times of higher burden on the hospitals, e.g. when the spike in hospitalisations occurred due to the new variants of concern [Faria et al., 2021].

Before the COVID-19 pandemic, the nowcasting models in epidemiology have not been widely popularised (with several exceptions, e.g. [Bastos et al., 2019], [McGough et al., 2020]). The most recent pandemic highlighted the issues with reporting delays and several new models have been proposed since. The novelty of our approach was using the latent Gaussian Process — a machine learning method. Although as with any nowcasting methods, their accuracy depends on the regularity of the reporting delays, GPs are highly flexible, which makes them a favourable candidate even if we do not have a lot of data or if the distribution of delays changes over time.

GPs have been utilised in many scientific fields, with a long history of use in geostatistics, physics, finance, bioinformatics and many others. In epidemiology, their main use has been in spatio-temporal statistics

(e.g. Bhatt et al. 2015), due to their remarkable interpolation properties and ability to capture complex relationships in the data. Other benefits of using GPs are uncertainty quantification, the ability to incorporate priors, reduced risk of overfitting, and no need to assume specific functional relationships. That makes them an excellent and versatile tool for solving many statistical questions.

Gaussian Processes and other machine learning methods have recently been gaining more traction among epidemiological modellers, even outside spatial modelling. Compared to more standard statistical methods like regression models, GPs can be considered somewhat less interpretable (especially in high-dimensional models, although there exists a level of interpretability through the kernel function) and more cumbersome to fit to the data and use for predictions. This is slowly being resolved by many approximate approaches, leveraging the GPU power, and many optimised algorithms provided by common libraries.

In this chapter, I demonstrate that a model incorporating a latent Gaussian Process can yield predictions comparable to, and in some cases, even more accurate than those of a group of human experts in epidemiology. While the use of crowd-sourced forecasts has been previously explored in machine learning (Abeliuk et al., 2020), our focus here is specifically on individuals who are trained and experienced at analysing epidemiological data. The decision to involve only experts in our human benchmarking exercise was motivated by the fact that these individuals are relied upon by authorities, policymakers, and governing bodies for their expertise. Therefore, we were particularly interested in assessing how our model's predictions compare to theirs. Further comparison of epidemiological models with human expert forecasts would not only be valuable for calibration but also potentially for demonstrating the reliability of mathematical models.

Chapter 5: Importations in renewal-equation model

Finally, coming back to the good model setting practices, in Chapter 5 I showed empirical examples of the influence of priors on the identifiability of importations in a discrete renewal equation parametrisation of R_t. This again underlines the importance of an appropriate and systematic model selection process, based not just on the goodness of fit to the observed data but including the domain knowledge regarding the model parameters. In that chapter, I fitted a discrete-renewal equation type model of R_t to simulated data, to show that a constraint on R_t is often necessary to correctly identify and quantify the imported cases of a disease or infection. In this project, this regularisation is carried out by assuming an informative, narrow

prior for R_t variance. As we show in the practical examples, if the model allows R_t to vary significantly, new cases of disease can be attributed to a large and sudden change in R_t, rather than the cases imported from a different location.

The approach presented in this chapter shows that through applying regularisation, we can identify the importation peaks in the simulated scenarios (subject to some constraints discussed in Section 5.4). Regularisation is a technique used across different statistical paradigms and machine learning (e.g. Lasso, Ridge regression, dropout, early stopping). In Bayesian settings, regularisation naturally arises through the use of informed priors, making it a technique routinely employed by Bayesian modellers.

In this chapter, I demonstrate a proof of principle by showcasing how regularisation techniques can facilitate the accurate identification of both the timing and magnitude of importations in an early epidemic scenario simulated experiment. While the task of identifying importations remains inherently challenging, we propose that integrating the effects of importations into standard modelling frameworks and packages should be feasible, provided that constraints on the R_t are maintained. This approach should be further tested by complementing it with real-world air passenger mobility data in future studies. Currently, the most reliable indication of whether an infection was imported or contracted locally can be obtained through genome sequencing data, so ideally, testing of our proposed method should be coupled with genomic datasets.

6.2 Overall conclusions

Statistical models have a rich history of application in modelling infectious disease dynamics. However, when faced with the emergence of a new pathogen, standard models are often insufficient in addressing the questions posed by epidemiologists. Innovative methods, leveraging the increasing computational capacity of modern machines, can substantially contribute to solving these modelling challenges.

Historically, Bayesian methods were primarily employed in scenarios with limited available data, where it was necessary to build models heavily relying on domain knowledge encapsulated in the priors. While the domain knowledge and priors remain fundamental, the emergence of a new pathogen introduces the possibility of misleading prior information. Conversely, although data availability was initially restricted at the onset of the COVID-19 pandemic, the volume of the datasets rapidly expanded with the global

spread of the virus.

Recognising the imperative of data-sharing, many governments, public health agencies, and research institutions promptly introduced regularly updated online dashboards, publicly accessible or restricted to researchers involved in pandemic response efforts. This led to an unprecedented increase in the volume of data available to modellers, presenting new challenges. Many Bayesian models, known for their computational complexity, became even more resource-intensive during the pandemic, where rapid response was paramount. Fortunately, the availability of libraries dedicated to optimised Probabilistic Programming is on the rise, facilitating computations using GPUs or parallel computing. This advancement also enables the incorporation of more sophisticated machine learning algorithms within a Bayesian framework.

Building upon this, I posit that combining standard statistical methods with modern machine learning techniques can offer fast and accurate frameworks for many epidemiological modellers. Machine learning has gained traction across various academic disciplines, enabling prompt processing of vast data volumes. Nonetheless, as emphasised throughout this thesis, the domain knowledge of the modellers will remain indispensable for model construction, identifying potential biases in the data, and comprehensively assessing the results.

Bibliography

E. Abdollahi, D. Champredon, J. M. Langley, A. P. Galvani, and S. M. Moghadas. Temporal estimates of case-fatality rate for COVID-19 outbreaks in Canada and the United States. *CMAJ*, 192(25):E666–E670, 2020. doi: 10.1503/cmaj.200711. URL https://doi.org/10.1503/cmaj.200711.

A. Abeliuk, D. M. Benjamin, F. Morstatter, and A. Galstyan. Quantifying machine influence over human forecasters. *Scientific Reports*, 10(1):15940, 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-72690-4. URL https://doi.org/10.1038/s41598-020-72690-4.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. doi: 10.1109/TAC.1974.1100705. URL https://oa.mg/work/10.1109/tac.1974.1100705.

A. Altmejd, J. Rocklöv, and J. Wallin. Nowcasting COVID-19 statistics reported with delay: a case-study of Sweden. *arXiv*, 2020. doi: arXiv:2006.06840. URL https://arxiv.org/abs/2006.06840.

R. Anderson and R. May. *Infectious Diseases of Humans: Dynamics and Control*. Infectious Diseases of Humans: Dynamics and Control. OUP Oxford, 1991. ISBN 9780198540403. URL https://books.google.co.uk/books?id=HT0--xXBguQC.

A. Aswi, S. M. Cramb, P. Moraga, and K. Mengersen. Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review. *Epidemiology and Infection*, 147:e33, 2019. doi: 10.1017/S0950268818002807. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6518570/.

D. Azzimonti and D. Ginsbourger. Estimating Orthant Probabilities of High-Dimensional Gaussian Vectors with An Application to Set Estimation. *Journal of Computational and Graphical Statistics*, 27:255 – 267, 2016. doi: 10.1080/10618600.2017.1360781. URL https://www.tandfonline.com/doi/full/10.1080/10618600.2017.1360781.

G. Baele, P. Lemey, and M. A. Suchard. Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty. *Systematic Biology*, 65(2):250–264, 2016. doi: 10.1093/sysbio/syv083. URL https://academic.oup.com/sysbio/article/65/2/250/2427272.

M. Bańbura and M. Modugno. Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160, 2014. URL https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1189.pdf.

M. Banbura, D. Giannone, and L. Reichlin. Nowcasting. *ECB Working Paper No. 1275*, 2010. URL https://ssrn.com/abstract=1717887.

P. Baqui, I. Bica, V. Marra, A. Ercole, and M. van der Schaar. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *The Lancet Global Health*, 8:e1018 – e1026, 2020. doi: 10.1016/S2214-109X(20)30285-0. URL https://doi.org/10.1016/S2214-109X(20)30285-0.

S. Baral, R. Chandler, R. G. Prieto, S. Gupta, S. Mishra, and M. Kulldorff. Leveraging epidemiological principles to evaluate Sweden's COVID-19 response. *Annals of Epidemiology*, 54:21–26, 2021. ISSN 1047-2797. doi: 10.1016/j.annepidem.2020.11.005. URL https://www.sciencedirect.com/science/article/pii/S1047279720304130.

L. V. Barrozo, M. Fornaciali, C. D. S. de André, G. A. Z. Morais, G. Mansur, W. Cabral-Miranda, M. J. de Miranda, J. R. Sato, and E. A. Júnior. GeoSES: A socioeconomic index for health and social research in Brazil. *PLoS ONE*, 15(4), 2020. doi: 10.1371/journal.pone.0232074. URL https://doi.org/10.1371/journal.pone.0232074.

L. S. Bastos, T. Economou, M. F. C. Gomes, D. A. M. Villela, F. C. Coelho, O. G. Cruz, O. Stoner, T. Bailey, and C. T. Codeço. A modelling approach for correcting reporting delays in disease surveillance data. *Statistics in Medicine*, 38(22):4363–4377, 2019. doi: 10.1002/sim.8303. URL https://doi.org/10.1002/sim.8303.

L. S. Bastos, R. P. Niquini, R. M. Lana, D. A. Villela, O. G. Cruz, F. C. Coelho, C. T. Codeço, and M. F. Gomes. COVID-19 and hospitalizations for SARI in Brazil: a comparison up to the 12th epidemiological week of 2020. *Cadernos de Saúde Pública*, 36, 2020. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2020000406001&nrm=iso.

C. H. Bennett. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, 22(2):245–268, 1976. ISSN 0021-9991. doi: 10.1016/0021-9991(76)90078-4. URL https://www.sciencedirect.com/science/article/pii/0021999176900784.

J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974. ISSN 00359246. URL http://www.jstor.org/stable/2984812.

S. Bhatt, D. J. Weiss, E. Cameron, D. Bisanzio, B. Mappin, U. Dalrymple, K. E. Battle, C. L. Moyes, A. Henry, P. A. Eckhoff, E. A. Wenger, O. Briët, M. A. Penny, T. A. Smith, A. Bennett, J. Yukich, T. P. Eisele, J. T. Griffin, C. A. Fergus, M. Lynch, F. Lindgren, J. M. Cohen, C. L. J. Murray, D. L. Smith, S. I. Hay, R. E. Cibulskis, and P. W. Gething. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*, 526(7572):207–211, October 2015. ISSN 1476-4687. doi: 10.1038/nature15535. URL https://doi.org/10.1038/nature15535.

Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, and T. Feng. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Disease*, 20:911–919, 2020. doi: 10.1016/S1473-3099(20)30287-5. URL https://doi.org/10.1016/S1473-3099(20)30287-5.

E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman. Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*, 2018. URL https://www.jmlr.org/papers/volume20/18-403/18-403.pdf.

P. Birrell, J. Blake, E. van Leeuwen, N. Gent, and D. De Angelis. Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1829):20200279, 2021. doi: 10.1098/rstb.2020.0279. URL https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2020.0279.

Brazilian Institute of Geography and Statistics. IBGE Projeções da População. URL https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html?=&t=resultados.

A. Brizzi, C. Whittaker, L. M. S. Servo, I. Hawryluk, C. A. Prete, W. M. de Souza, R. S. Aguiar, L. J. T. Araujo, L. S. Bastos, A. Blenkinsop, L. F. Buss, D. Candido, M. C. Castro, S. F. Costa, J. Croda, A. A.

de Souza Santos, C. Dye, S. Flaxman, P. L. C. Fonseca, V. E. V. Geddes, B. Gutierrez, P. Lemey, A. S. Levin, T. Mellan, D. M. Bonfim, X. Miscouridou, S. Mishra, M. Monod, F. R. R. Moreira, B. Nelson, R. H. M. Pereira, O. Ranzani, R. P. Schnekenberg, E. Semenova, R. Sonabend, R. P. Souza, X. Xi, E. C. Sabino, N. R. Faria, S. Bhatt, and O. Ratmann. Spatial and temporal fluctuations in COVID-19 fatality rates in Brazilian hospitals. *Nature Medicine*, 28(7):1476–1485, 07 2022. ISSN 1546-170X. doi: 10.1038/s41591-022-01807-1. URL https://doi.org/10.1038/s41591-022-01807-1.

M. Brown. A generalized error function in n dimensions. *U.S. Naval Missile Center, Theoretical Analysis Division*, 1963. URL https://apps.dtic.mil/dtic/tr/fulltext/u2/401722.pdf.

P.-C. Bürkner, J. Gabry, and A. Vehtari. Approximate leave-future-out cross-validation for Bayesian time series models. *Journal of Statistical Computation and Simulation*, 0(0):1–25, 2020. doi: 10.1080/00949655.2020.1783262. URL https://doi.org/10.1080/00949655.2020.1783262.

L. F. Buss, C. A. Prete, C. M. M. Abrahim, A. Mendrone, T. Salomon, C. de Almeida-Neto, R. F. O. França, M. C. Belotti, M. P. S. S. Carvalho, A. G. Costa, M. A. E. Crispim, S. C. Ferreira, N. A. Fraiji, S. Gurzenda, C. Whittaker, L. T. Kamaura, P. L. Takecian, P. da Silva Peixoto, M. K. Oikawa, A. S. Nishiya, V. Rocha, N. A. Salles, A. A. de Souza Santos, M. A. da Silva, B. Custer, K. V. Parag, M. Barral-Netto, M. U. G. Kraemer, R. H. M. Pereira, O. G. Pybus, M. P. Busch, M. C. Castro, C. Dye, V. H. Nascimento, N. R. Faria, and E. C. Sabino. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science*, 371(6526):288–292, 2021. doi: 10.1126/science.abe9728. URL https://www.science.org/doi/abs/10.1126/science.abe9728.

B. Calderhead and M. Girolami. Estimating Bayes factors via thermodynamic integration and population MCMC. *Computational Statistics & Data Analysis*, 53(12):4028 – 4045, 2009. ISSN 0167-9473. doi: 10.1016/j.csda.2009.07.025. URL http://www.sciencedirect.com/science/article/pii/S0167947309002722.

E. Cameron and A. Pettitt. Recursive Pathways to Marginal Likelihood Estimation with Prior-Sensitivity Analysis. *Statistical Science*, 29(3):397 – 419, 2014. doi: 10.1214/13-STS465. URL https://doi.org/10.1214/13-STS465.

B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A Probabilistic Programming Language. *Journal of Statistical Software, Articles*,

76(1):1–32, 2017. ISSN 1548-7660. doi: 10.18637/jss.v076.i01. URL https://www.jstatsoft.org/v076/i01.

M. C. Castro, S. Kim, L. Barberia, A. F. Ribeiro, S. Gurzenda, K. B. Ribeiro, E. Abbott, J. Blossom, B. Rache, and B. H. Singer. Spatiotemporal pattern of COVID-19 spread in Brazil. *Science*, 372(6544): 821–826, 2021. doi: 10.1126/science.abh1558. URL https://www.science.org/doi/abs/10.1126/science.abh1558.

T. Chen, D. Wu, H. Chen, W. Yan, D. Yang, G. Chen, K. Ma, D. Xu, H. Yu, H. Wang, T. Wang, W. Guo, J. Chen, C. Ding, X. Zhang, J. Huang, M. Han, S. Li, X. Luo, J. Zhao, and Q. Ning. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ*, 368, 2020. doi: 10.1136/bmj.m1091. URL https://doi.org/10.1136/bmj.m1091.

M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368 (6489):395–400, 2020. doi: 10.1126/science.aba9757. URL https://www.science.org/doi/abs/10.1126/science.aba9757.

G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3):379–398, 2017. ISSN 2468-0427. doi: https://doi.org/10.1016/j.idm.2017.08.001. URL https://www.sciencedirect.com/science/article/pii/S2468042717300234.

C. Codeco, F. Coelho, O. Cruz, S. Oliveira, T. Castro, and L. Bastos. Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil. *Revue d'Épidémiologie et de Santé Publique*, 66:S386, 2018. doi: 0.1016/j.respe.2018.05.408. URL https://doi.org/10.1016/j.respe.2018.05.408.

E. A. Codling, M. J. Plank, and S. Benhamou. Random walk models in biology. *Journal of The Royal Society Interface*, 5(25):813–834, 2008. doi: 10.1098/rsif.2008.0014. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2008.0014.

A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*, 178(9): 1505–1512, 09 2013. ISSN 0002-9262. doi: 10.1093/aje/kwt133. URL https://doi.org/10.1093/aje/kwt133.

R. N. Curnow and C. W. Dunnett. The Numerical Evaluation of Certain Multivariate Normal Integrals. *The Annals of Mathematical Statistics*, 33(2):571 – 579, 1962. doi: 10.1214/aoms/1177704581. URL https://doi.org/10.1214/aoms/1177704581

A. A. M. da Silva, L. G. L. Neto, C. de Maria Pedrozo, S. de Azevedo, L. M. M. da Costa, M. L. B. M. Bragança, A. K. D. Barros Filho, B. B. Wittlin, B. F. de Souza, B. L. C. A. de Oliveira, et al. Population-based seroprevalence of SARS-CoV-2 is more than halfway through the herd immunity threshold in the State of Maranhao, Brazil. *medRxiv*, 2020. doi: 10.1101/2020.08.28.20180463. URL https://www.medrxiv.org/content/early/2020/09/01/2020.08.28.20180463

S. Dana, A. B. Simas, B. A. Filardi, R. N. Rodriguez, L. L. da Costa Valiengo, and J. Gallucci-Neto. Brazilian Modeling of COVID-19 (bram-cod): a Bayesian Monte Carlo approach for COVID-19 spread in a limited data set context. *medRxiv*, 2020. doi: 10.1101/2020.04.29.20081174. URL https://doi.org/10.1101/2020.04.29.20081174.

N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, C. A. B. Pearson, B. J. Quilty, A. J. Kucharski, H. Gibbs, S. Clifford, A. Gimma, K. van Zandvoort, J. D. Munday, C. Diamond, W. J. Edmunds, R. M. G. J. Houben, J. Hellewell, T. W. Russell, S. Abbott, S. Funk, N. I. Bosse, Y. F. Sun, S. Flasche, A. Rosello, C. I. Jarvis, R. M. Eggo, and CMMID COVID-19 working group. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, 26(8):1205–1211, 2020a. doi: 10.1038/s41591-020-0962-9. URL https://doi.org/10.1038/s41591-020-0962-9

N. G. Davies, A. J. Kucharski, R. M. Eggo, A. Gimma, W. J. Edmunds, T. Jombart, K. O'Reilly, A. Endo, J. Hellewell, E. S. Nightingale, B. J. Quilty, C. I. Jarvis, T. W. Russell, P. Klepac, N. I. Bosse, S. Funk, S. Abbott, G. F. Medley, H. Gibbs, C. A. B. Pearson, S. Flasche, M. Jit, S. Clifford, K. Prem, C. Diamond, J. Emery, A. K. Deol, S. R. Procter, K. van Zandvoort, Y. F. Sun, J. D. Munday, A. Rosello, M. Auzenbergs, G. Knight, R. M. G. J. Houben, and Y. Liu. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, 5(7):e375–e385, 2020b. ISSN 2468-2667. doi: 10.1016/S2468-2667(20)30133-X. URL https://doi.org/10.1016/S2468-2667(20)30133-X.

N. G. Davies, S. Abbott, R. C. Barnard, C. I. Jarvis, A. J. Kucharski, J. D. Munday, C. A. B. Pearson, T. W. Russell, D. C. Tully, A. D. Washburne, T. Wenseleers, A. Gimma, W. Waites, K. L. M. Wong, K. van Zandvoort, J. D. Silverman, C. C.-. W. Group1‡, C.-. G. U. C.-U. Consortium‡, K. Diaz-Ordaz,

R. Keogh, R. M. Eggo, S. Funk, M. Jit, K. E. Atkins, and W. J. Edmunds. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538):eabg3055, 2021a. doi: 10.1126/science.abg3055. URL https://www.science.org/doi/abs/10.1126/science.abg3055.

N. G. Davies, R. C. Barnard, C. I. Jarvis, T. W. Russell, M. G. Semple, M. Jit, and W. J. Edmunds. Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: a modelling study. *The Lancet Infectious Diseases*, 21(4):482–492, 2021b. ISSN 1473-3099. doi: 10.1016/S1473-3099(20)30984-1. URL https://doi.org/10.1016/S1473-3099(20)30984-1.

W. de Souza, L. Buss, D. Candido, et al. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nature Human Behavior*, 4:856–865, 2020. doi: 10.1038/s41562-020-0928-4. URL https://doi.org/10.1038/s41562-020-0928-4.

T. J. Diciccio, R. E. Kass, A. Raftery, and L. Wasserman. Computing Bayes Factors by Combining Simulation and Asymptotic Approximations. *Journal of the American Statistical Association*, 92(439): 903–915, 1997. doi: 10.1080/01621459.1997.10474045. URL https://doi.org/10.1080/01621459.1997.10474045.

O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, June 1990. ISSN 1432-1416. doi: 10.1007/BF00178324. URL https://doi.org/10.1007/BF00178324.

J. Ding, V. Tarokh, and Y. Yang. Model Selection Techniques: An Overview. *IEEE Signal Processing Magazine*, 35(6):16–34, 2018. doi: 10.1109/MSP.2018.2867638. URL https://doi.org/10.1109/MSP.2018.2867638.

P. A. M. Dirac and N. H. D. Bohr. The quantum theory of the emission and absorption of radiation. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 114(767):243–265, 1927. doi: 10.1098/rspa.1927.0039. URL https://doi.org/10.1098/rspa.1927.0039.

C. A. Donnelly, A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, L. J. Abu-Raddad, L.-M. Ho, T.-Q. Thach, P. Chau, K.-P. Chan, L.-Y. T. Tai-Hing Lam, T. Tsang, S.-H. Liu, E. M. C. L. James H B Kong, N. M. Ferguson, and R. M. Anderson. Epidemiological determinants of spread of

causal agent of severe acute respiratory syndrome in Hong Kong. *The Lancet*, 361:1761–1766, 2003. doi: 10.1016/S0140-6736(03)13410-1. URL https://doi.org/10.1016/S0140-6736(03)13410-1.

A. I. Duff, T. Davey, D. Korbmacher, A. Glensk, B. Grabowski, J. Neugebauer, and M. W. Finnis. Improved method of calculating ab initio high-temperature thermodynamic properties with application to ZrC. *Physical Review B*, 91(21):214311, 2015. doi: 10.1103/PhysRevB.91.214311. URL https://doi.org/10.1103/PhysRevB.91.214311.

D. Duvenaud, J. Lloyd, R. Grosse, J. Tenenbaum, and G. Zoubin. Structure Discovery in Nonparametric Regression through Compositional Kernel Search. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1166–1174, Atlanta, Georgia, USA, 2013. PMLR. URL http://proceedings.mlr.press/v28/duvenaud13.html.

O. Eales and S. Riley. Differences between the true reproduction number and the apparent reproduction number of an epidemic time series. *Epidemics*, 46:100742, 2024. ISSN 1755-4365. doi: https://doi.org/10.1016/j.epidem.2024.100742. URL https://www.sciencedirect.com/science/article/pii/S1755436524000033.

Y. Fan, R. Wu, M.-H. Chen, L. Kuo, and P. O. Lewis. Choosing among partition models in Bayesian phylogenetics. *Molecular Biology and Evolution*, 28(1):523–532, 2011. doi: 10.1093/molbev/msq224. URL https://doi.org/10.1093/molbev/msq224.

N. R. Faria, T. A. Mellan, C. Whittaker, I. M. Claro, D. da S. Candido, S. Mishra, M. A. E. Crispim, F. C. S. Sales, I. Hawryluk, J. T. McCrone, R. J. G. Hulswit, L. A. M. Franco, M. S. Ramundo, J. G. de Jesus, P. S. Andrade, T. M. Coletti, G. M. Ferreira, C. A. M. Silva, E. R. Manuli, R. H. M. Pereira, P. S. Peixoto, M. U. G. Kraemer, N. Gaburo, C. da C. Camilo, H. Hoeltgebaum, W. M. Souza, E. C. Rocha, L. M. de Souza, M. C. de Pinho, L. J. T. Araujo, F. S. V. Malta, A. B. de Lima, J. do P. Silva, D. A. G. Zauli, A. C. de S. Ferreira, R. P. Schnekenberg, D. J. Laydon, P. G. T. Walker, H. M. Schlüter, A. L. P. dos Santos, M. S. Vidal, V. S. D. Caro, R. M. F. Filho, H. M. dos Santos, R. S. Aguiar, J. L. Proença-Modena, B. Nelson, J. A. Hay, M. Monod, X. Miscouridou, H. Coupland, R. Sonabend, M. Vollmer, A. Gandy, C. A. Prete, V. H. Nascimento, M. A. Suchard, T. A. Bowden, S. L. K. Pond, C.-H. Wu, O. Ratmann, N. M. Ferguson, C. Dye, N. J. Loman, P. Lemey, A. Rambaut, N. A. Fraiji, M. do P. S. S. Carvalho, O. G. Pybus, S. Flaxman, S. Bhatt, and E. C. Sabino. Genomics

and epidemiology of the p.1 sars-cov-2 lineage in manaus, brazil. *Science*, 372(6544):815–821, 2021. doi: 10.1126/science.abh2644. URL https://www.science.org/doi/abs/10.1126/science.abh2644.

N. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, et al. Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Technical report, Imperial College London, 2020. URL https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf.

L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936, 2020. doi: 10.1126/science.abb6936. URL https://www.science.org/doi/abs/10.1126/science.abb6936.

S. Flaxman, A. G. Wilson, D. Neill, H. Nickisch, and A. Smola. Fast Kronecker Inference in Gaussian Processes with non-Gaussian Likelihoods. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 607–616, Lille, France, 07–09 Jul 2015. PMLR. URL http://proceedings.mlr.press/v37/flaxman15.html.

S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, H. Coupland, T. A. Mellan, H. Zhu, T. Berah, J. W. Eaton, Imperial COVID-19 Reponse Team, A. Ghani, C. A. Donnelly, S. Riley, L. C. Okell, M. A. C. Vollmer, N. M. Ferguson, and S. Bhatt. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584:257—261, 2020. doi: 10.1038/s41586-020-2405-7. URL https://doi.org/10.1038/s41586-020-2405-7.

C. Fraser. Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. *PLOS ONE*, 2(8):1–12, 08 2007a. doi: 10.1371/journal.pone.0000758. URL https://doi.org/10.1371/journal.pone.0000758.

C. Fraser. Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. *PLOS ONE*, 2(8):1–12, 08 2007b. doi: 10.1371/journal.pone.0000758. URL https://doi.org/10.1371/journal.pone.0000758.

N. Friel and A. N. Pettitt. Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):589–607, 2008. doi: 10.1111/j.1467-9868.2007.00650.x. URL https://doi.org/10.1111/j.1467-9868.2007.00650.x.

N. Friel and J. Wyse. Estimating the evidence: a review. *Statistica Neerlandica*, 66(3):288–308, 2012. doi: 10.1111/j.1467-9574.2011.00515.x. URL https://doi.org/10.1111/j.1467-9574.2011.00515.x.

N. Friel, M. Hurn, and J. Wyse. Improving power posterior estimation of statistical evidence. *Statistics and Computing*, pages 1165–1180, 2017. doi: 10.1007/s11222-016-9678-6. URL https://doi.org/10.1007/s11222-016-9678-6.

S. Galmiche, T. Cortier, T. Charmet, L. Schaeffer, O. Chény, C. von Platen, A. Lévy, S. Martin, F. Omar, C. David, A. Mailles, F. Carrat, S. Cauchemez, and A. Fontanet. SARS-CoV-2 incubation period across variants of concern, individual factors, and circumstances of infection in France: a case series analysis from the ComCor study. *The Lancet Microbe*, 4(6):e409–e417, 2023. doi: 10.1016/S2666-5247(23)00005-8. URL https://doi.org/10.1016/S2666-5247(23)00005-8

T. Garske, J. Legrand, C. A. Donnelly, H. Ward, S. Cauchemez, C. Fraser, N. M. Ferguson, and A. C. Ghani. Assessing the severity of the novel influenza A/H1N1 pandemic. *BMJ*, 339, 2009. ISSN 0959-8138. doi: 10.1136/bmj.b2840. URL https://doi.org/10.1136/bmj.b2840.

H. Ge, K. Xu, and Z. Ghahramani. Turing: a language for flexible probabilistic inference. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, pages 1682–1690, 2018. URL http://proceedings.mlr.press/v84/ge18b.html.

A. E. Gelfand and S. K. Sahu. Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *Journal of the American Statistical Association*, 94(445):247–253, 1999. ISSN 01621459. URL http://www.jstor.org/stable/2669699.

M. Gell-Mann and K. A. Brueckner. Correlation energy of an electron gas at high density. *Physical Review*, 106(2):364–368, 1957. doi: 10.1103/PhysRev.106.364. URL https://doi.org/10.1103/PhysRev.106.364.

A. Gelman and X.-L. Meng. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*, 13(2):163 – 185, 1998. doi: 10.1214/ss/1028905934. URL https://doi.org/10.1214/ss/1028905934.

A. Gelman and D. B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statisti-*

cal Science, 7(4):457 – 472, 1992. doi: 10.1214/ss/1177011136. URL https://doi.org/10.1214/ss/1177011136.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition, 2013. doi: 10.1201/b16018. URL https://doi.org/10.1201/b16018.

A. Gelman, J. Hwang, and A. Vehtari. Understanding Predictive Information Criteria for Bayesian Models. *Statistics and Computing*, 24(6):997–1016, November 1 2014. ISSN 1573-1375. doi: 10.1007/s11222-013-9416-2. URL https://doi.org/10.1007/s11222-013-9416-2.

A. Gelman, A. Vehtari, D. Simpson, C. C. Margossian, B. Carpenter, Y. Yao, L. Kennedy, J. Gabry, P.-C. Bürkner, and M. Modrák. Bayesian Workflow. *arXiv preprint*, 2020. doi: 10.48550/arXiv.2011.01808. URL https://doi.org/10.48550/arXiv.2011.01808.

W. R. Gilks, A. Thomas, and D. J. Spiegelhalter. A Language and Program for Complex Bayesian Modelling. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 43(1):169–177, 1994. ISSN 00390526, 14679884. URL http://www.jstor.org/stable/2348941.

A. E. Gorbalenya, S. C. Baker, R. S. Baric, R. J. de Groot, C. Drosten, A. A. Gulyaeva, B. L. Haagmans, C. Lauber, A. M. Leontovich, B. W. Neuman, D. Penzar, S. Perlman, L. L. M. Poon, D. V. Samborskiy, I. A. Sidorov, I. Sola, J. Ziebuhr, and Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5(4):536–544, 2020. doi: 10.1038/s41564-020-0695-z. URL https://doi.org/10.1038/s41564-020-0695-z.

K. M. Gostic, L. McGough, E. B. Baskerville, S. Abbott, K. Joshi, C. Tedijanto, R. Kahn, R. Niehus, J. A. Hay, P. M. De Salazar, J. Hellewell, S. Meakin, J. D. Munday, N. I. Bosse, K. Sherrat, R. N. Thompson, L. F. White, J. S. Huisman, J. Scire, S. Bonhoeffer, T. Stadler, J. Wallinga, S. Funk, M. Lipsitch, and S. Cobey. Practical considerations for measuring the effective reproductive number, R_t. *PLOS Computational Biology*, 16(12):1–21, 12 2020. doi: 10.1371/journal.pcbi.1008409. URL https://doi.org/10.1371/journal.pcbi.1008409.

N. C. Grassly and C. Fraser. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology*, 6:477–487, 2008. doi: 10.1038/nrmicro1845. URL https://doi.org/10.1038/nrmicro1845.

R. B. Grosse, C. J. Maddison, and R. R. Salakhutdinov. Annealing between distributions by averaging moments. In *Advances in Neural Information Processing Systems*, volume 26, 2013. URL https://proceedings.neurips.cc/paper/2013/file/fb60d411a5c5b72b2e7d3527cfc84fd0-Paper.pdf.

W.-j. Guan, Z.-y. Ni, Y. Hu, W.-h. Liang, C.-q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. S. Hui, B. Du, L.-j. Li, G. Zeng, K.-Y. Yuen, R.-c. Chen, C.-l. Tang, T. Wang, P.-y. Chen, J. Xiang, S.-y. Li, J.-l. Wang, Z.-j. Liang, Y.-x. Peng, L. Wei, Y. Liu, Y.-h. Hu, P. Peng, J.-m. Wang, J.-y. Liu, Z. Chen, G. Li, Z.-j. Zheng, S.-q. Qiu, J. Luo, C.-j. Ye, S.-y. Zhu, and N.-s. Zhong. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*, 382(18):1708–1720, 2020. doi: 10.1056/NEJMoa2002032. URL https://doi.org/10.1056/NEJMoa2002032.

F. Günther, A. Bender, K. Katz, H. Küchenhoff, and M. Höhle. Nowcasting the COVID-19 pandemic in Bavaria. *Biometrical Journal*, 63(3):490–502, 2020. doi: 10.1002/bimj.202000112. URL https://doi.org/10.1002/bimj.202000112.

M. Habeck. Evaluation of marginal likelihoods via the density of states. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pages 486–494, La Palma, Canary Islands, 21–23 Apr 2012. PMLR. URL https://proceedings.mlr.press/v22/habeck12.html.

I. Hawryluk, T. A. Mellan, H. H. Hoeltgebaum, R. P. Schnekenberg, S. Mishra, C. Whittaker, H. Zhu, C. Donnelly, A. Gandy, S. Flaxman, and S. Bhatt. Inference of COVID-19 epidemiological distributions from Brazilian hospital data. *Journal of The Royal Society Interface*, 17(172):20200596, 2020. doi: 10.1098/rsif.2020.0596. URL https://doi.org/10.1098/rsif.2020.0596.

I. Hawryluk, H. Hoeltgebaum, S. Mishra, X. Miscouridou, R. P. Schnekenberg, C. Whittaker, M. Vollmer, S. Flaxman, S. Bhatt, and T. A. Mellan. Gaussian process nowcasting: application to COVID-19 mortality reporting. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 1258–1268. PMLR, 27–30 Jul 2021. URL https://proceedings.mlr.press/v161/hawryluk21a.html.

I. Hawryluk, S. Mishra, S. Flaxman, S. Bhatt, and T. A. Mellan. Application of referenced thermodynamic integration to Bayesian model selection. *PLOS ONE*, 18(8):1–16, 08 2023. doi: 10.1371/journal.pone.0289889. URL https://doi.org/10.1371/journal.pone.0289889.

J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, S. Flasche, B. J. Quilty, N. Davies, Y. Liu, S. Clifford, P. Klepac, M. Jit, C. Diamond, H. Gibbs, K. van Zandvoort, S. Funk, and R. M. Eggo. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 8(4):e488–e496, 2020. ISSN 2214-109X. doi: 10.1016/S2214-109X(20)30074-7. URL https://doi.org/10.1016/S2214-109X(20)30074-7.

M. D. Hoffman and A. Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014. URL http://jmlr.org/papers/v15/hoffman14a.html.

R. P. Inward, F. Jackson, A. Dasgupta, G. Lee, A. L. Battle, K. V. Parag, and M. U. Kraemer. Impact of spatiotemporal heterogeneity in COVID-19 disease surveillance on epidemiological parameters and case growth rates. *Epidemics*, 41:100627, 2022. ISSN 1755-4365. doi: https://doi.org/10.1016/j.epidem.2022.100627. URL https://www.sciencedirect.com/science/article/pii/S1755436522000676.

W. H. Jefferys and J. O. Berger. Ockham's Razor and Bayesian Analysis. *American Scientist*, 80(1):64–72, 1992. URL http://www.jstor.org/stable/29774559.

H. Jeffreys. *Theory of Probability*. Oxford University Press, 3 edition, 1965. URL https://books.google.co.uk/books/about/Theory_of_Probability.html?id=EbodAQAAMAAJ&redir_esc=y.

P. Jha, Y. Deshmukh, C. Tumbe, W. Suraweera, A. Bhowmick, S. Sharma, P. Novosad, S. H. Fu, L. Newcombe, H. Gelband, and P. Brown. COVID mortality in India: National survey data and health facility deaths. *Science*, 375(6581):667–671, 2022. doi: 10.1126/science.abm5154. URL https://www.science.org/doi/abs/10.1126/science.abm5154.

T. Jombart, K. Van Zandvoort, T. W. Russell, C. I. Jarvis, A. Gimma, S. Abbott, S. Clifford, S. Funk, H. Gibbs, Y. Liu, et al. Inferring the number of COVID-19 cases from recently reported deaths. *Wellcome Open Research*, 5, 2020. doi: 10.12688/wellcomeopenres.15786.1. URL https://doi.org/10.12688/wellcomeopenres.15786.1.

M. I. Jordan. What are the open problems in Bayesian statistics? *ISBA Bulletin*, 18(1), 2011. URL https://www4.stat.ncsu.edu/~reich/st740/Bayesian_open_problems.pdf.

K. S. Kaminsky. Prediction of IBNR claim counts by modelling the distribution of report lags. *Insurance: Mathematics and Economics*, 6(2):151 – 159, 1987. doi: 10.1016/0167-6687(87)90024-2. URL https://doi.org/10.1016/0167-6687(87)90024-2.

R. E. Kass and A. E. Raftery. Bayes Factors. *Journal of the American Statistical Association*, 90(430): 773–795, 1995. URL https://www.stat.cmu.edu/~kass/papers/bayesfactors.pdf.

W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London*, 115(772):700–721, 1927. doi: 10.1098/rspa.1927.0118. URL https://doi.org/10.1098/rspa.1927.0118.

J. G. Kirkwood. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics*, 3(5):300–313, 1935. doi: 10.1063/1.1749657. URL https://doi.org/10.1063/1.1749657.

S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, and M. Lipsitch. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368(6493):860–868, 2020. doi: 10.1126/science.abb5793. URL https://www.science.org/doi/abs/10.1126/science.abb5793.

A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, N. Davies, A. Gimma, K. van Zandvoort, H. Gibbs, J. Hellewell, C. I. Jarvis, S. Clifford, B. J. Quilty, N. I. Bosse, S. Abbott, P. Klepac, and S. Flasche. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(5):553–558, 2020. ISSN 1473-3099. doi: 10.1016/S1473-3099(20)30144-4. URL https://doi.org/10.1016/S1473-3099(20)30144-4.

S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22 (1):79 – 86, 1951. doi: 10.1214/aoms/1177729694. URL https://doi.org/10.1214/aoms/1177729694.

M. Kulldorff, S. Gupta, and J. Bhattacharya. The Great Barrington Declaration. https://gbdeclaration.org/, 2020.

B. Lambert. *A Student's Guide to Bayesian Statistics*. SAGE Publications Ltd, London, 2018. ISBN 9781526418289. URL https://study.sagepub.com/lambert/student-resources.

M. M. Lamers and B. L. Haagmans. SARS-CoV-2 pathogenesis. *Nature Reviews Microbiology*, 20(5):270–284, 2022. doi: 10.1038/s41579-022-00713-0. URL https://doi.org/10.1038/s41579-022-00713-0.

N. Lartillot and H. Philippe. Computing Bayes Factors Using Thermodynamic Integration. *Systematic Biology*, 55(2):195–207, 2006. ISSN 1063-5157. doi: 10.1080/10635150500433722. URL https://doi.org/10.1080/10635150500433722.

J. F. Lawless. Adjustments for Reporting Delays and the Prediction of Occurred but Not Reported Events. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 22(1):15–31, 1994. URL http://www.jstor.org/stable/3315820.

G. Lefebvre, R. Steele, and A. C. Vandal. A path sampling identity for computing the Kullback–Leibler and J divergences. *Computational Statistics & Data Analysis*, 54(7):1719–1731, 2010. ISSN 0167-9473. doi: 10.1016/j.csda.2010.01.018. URL https://www.sciencedirect.com/science/article/pii/S0167947310000332

M. Li, P. Chen, Q. Yuan, B. Song, and J. Ma. Transmission characteristics of the COVID-19 outbreak in China: a study driven by data. *medRxiv*, 2020a. URL https://www.medrxiv.org/content/10.1101/2020.02.26.20028431v1.full.pdf.

Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*, 382(13):1199–1207, 2020b. doi: 10.1056/NEJMoa2001316. URL https://doi.org/10.1056/NEJMoa2001316.

N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of Clinical Medicine*, 9(2):538, 2020. doi: 10.3390/jcm9020538. URL https://doi.org/10.3390/jcm9020538.

H. Liu, Y.-S. Ong, X. Shen, and J. Cai. When Gaussian Process Meets Big Data: A Review of Scalable GPs. *IEEE Transactions on Neural Networks and Learning Systems*, 31(11):4405–4423, 2020. doi: 10.1109/TNNLS.2019.2957109. URL https://doi.org/10.1109/TNNLS.2019.2957109.

F. Llorente, L. Martino, D. Delgado, and J. López-Santiago. Marginal Likelihood Computation for Model Selection and Hypothesis Testing: An Extensive Review. *SIAM Review*, 65(1):3–58, 2023. doi: 10.1137/20M1310849. URL https://doi.org/10.1137/20M1310849.

G. Loaiza-Ganem and J. P. Cunningham. The continuous Bernoulli: fixing a pervasive error in variational autoencoders. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/f82798ec8909d23e55679ee26bb26437-Paper.pdf.

R. McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC, 1st edition, 2016. doi: 10.1201/9781315372495. URL https://doi.org/10.1201/9781315372495.

S. F. McGough, M. A. Johansson, M. Lipsitch, and N. A. Menzies. Nowcasting by Bayesian Smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLoS computational biology*, 16(4): e1007735, 2020. doi: 10.1371/journal.pcbi.1007735. URL https://doi.org/10.1371/journal.pcbi.1007735.

T. A. Mellan, H. H. Hoeltgebaum, S. Mishra, C. Whittaker, R. P. Schnekenberg, A. Gandy, H. J. T. Unwin, M. A. Vollmer, H. Coupland, I. Hawryluk, et al. Report 21: Estimating COVID-19 cases and reproduction number in Brazil. *medRxiv*, 2020. URL https://doi.org/10.25561/78872.

X.-L. Meng and W. H. Wong. Simulating rations of normalising constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 6(4):831–860, 1996a. ISSN 10170405, 19968507. URL http://www.jstor.org/stable/24306045.

X.-L. Meng and W. H. Wong. Simulationg ratios of normalising constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 6(4):831–860, 1996b. ISSN 10170405, 19968507. URL http://www.jstor.org/stable/24306045.

Ministério da Saúde. SRAG 2020 - Banco de Dados de Síndrome Respiratória Aguda Grave, 2020. URL https://opendatasus.saude.gov.br/dataset/bd-srag-2020.

S. Mishra, T. Berah, T. A. Mellan, H. J. T. Unwin, M. A. Vollmer, K. V. Parag, A. Gandy, S. Flaxman, and S. Bhatt. On the derivation of the renewal equation from an age-dependent branching process: an

epidemic modelling perspective. *arXiv preprint*, 2020. doi: 10.48550/arXiv.2006.16487. URL https://doi.org/10.48550/arXiv.2006.16487.

T. Miwa, A. J. Hayter, and S. Kuriki. The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):223–234, 2003. doi: 10.1111/1467-9868.00382. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00382.

K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*, 25(10):2000180, 2020. doi: 10.2807/1560-7917.ES.2020.25.10.2000180. URL https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.10.2000180.

W. Msemburi, A. Karlinsky, V. Knutson, S. Aleshin-Guendel, S. Chatterji, and J. Wakefield. The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature*, 613(7942):130–137, January 2023. ISSN 1476-4687. doi: 10.1038/s41586-022-05522-2. URL https://doi.org/10.1038/s41586-022-05522-2.

D. J. Navarro. Between the Devil and the Deep Blue Sea: Tensions Between Scientific Judgment and Statistical Model Selection. *Computational Brain & Behavior*, 2(1):28–34, March 2019. ISSN 2522-087X. doi: 10.1007/s42113-018-0019-z. URL https://doi.org/10.1007/s42113-018-0019-z.

R. M. Neal. Probabilistic Inference Using Markov Chain Monte Carlo Methods , 1993. URL https://bayes.wustl.edu/Manual/RadfordNeal.review.pdf.

R. M. Neal. Annealed importance sampling. *Statistics and Computing*, 11:125–139, 2001. doi: 10.1023/A:1008923215028. URL https://doi.org/10.1023/A:1008923215028.

R. P. Niquini, R. M. Lana, A. G. Pacheco, O. G. Cruz, F. C. Coelho, L. M. Carvalho, D. A. M. Villela, M. F. d. C. Gomes, and L. S. Bastos. Description and comparison of demographic characteristics and comorbidities in SARI from COVID-19, SARI from influenza, and the Brazilian general population. *Cadernos de Saúde Pública*, 36, 00 2020. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2020000705013&nrm=iso.

P. Nouvellet, A. Cori, T. Garske, I. M. Blake, I. Dorigatti, W. Hinsley, T. Jombart, H. L. Mills, G. Nedjati-Gilani, M. D. Van Kerkhove, C. Fraser, C. A. Donnelly, N. M. Ferguson, and S. Riley. A simple

approach to measure transmissibility and forecast incidence. *Epidemics*, 22:29–35, 2018. ISSN 1755-4365. doi: 10.1016/j.epidem.2017.02.012. URL https://www.sciencedirect.com/science/article/pii/S1755436517300245.

D. B. Owen. Orthant probabilities. *Wiley StatsRef: Statistics Reference Online*, 2014. doi: 10.1002/9781118445112.stat01097. URL https://doi.org/10.1002/9781118445112.stat01097.

M. Pakkanen, X. Miscouridou, M. Penn, et al. Unifying Incidence and Prevalence Under a Time-Varying General Branching Process. *Journal of Mathematical Biology*, 87(35), 2023. doi: 10.1007/s00285-023-01958-w. URL https://doi.org/10.1007/s00285-023-01958-w.

K. V. Parag and C. A. Donnelly. Using information theory to optimise epidemic models for real-time prediction and estimation. *PLOS Computational Biology*, 16(7):1–20, 07 2020. doi: 10.1371/journal.pcbi.1007990. URL https://doi.org/10.1371/journal.pcbi.1007990.

D. Phan, N. Pradhan, and M. Jankowiak. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. *arXiv preprint*, 2019. doi: 10.48550/arXiv.1912.11554. URL https://doi.org/10.48550/arXiv.1912.11554.

R. Piessens, E. deDoncker Kapenga, C. Ueberhuber, and D. Kahaner. *QUADPACK: A Subroutine Package for Automatic Integration*. Springer, 1983. ISBN 3540125531. URL https://books.google.co.uk/books/about/Quadpack.html?id=m2AZAQAAIAAJ&redir_esc=y.

C. M. Pooley and G. Marion. Bayesian model evidence as a practical alternative to deviance information criterion. *Royal Society Open Science*, 5, 2018. doi: 10.1098/rsos.171519. URL https://doi.org/10.1098/rsos.171519.

K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. B. Pearson, J. D. Munday, S. Abbott, H. Gibbs, A. Rosello, B. J. Quilty, T. Jombart, F. Sun, C. Diamond, A. Gimma, K. van Zandvoort, S. Funk, C. I. Jarvis, W. J. Edmunds, N. I. Bosse, J. Hellewell, M. Jit, and P. Klepac. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, 5(5):e261–e270, 2020. doi: 10.1016/S2468-2667(20)30073-6. URL https://doi.org/10.1016/S2468-2667(20)30073-6.

Python Software Foundation. Python Software Foundation. URL https://www.python.org/.

Z. Qian, A. M. Alaa, and M. van der Schaar. When and How to Lift the Lockdown? Global COVID-19 Scenario Analysis and Policy Assessment using Compartmental Gaussian Processes. In *Advances in Neural Information Processing Systems*, volume 33, pages 10729–10740, 2020. URL https://proceedings.neurips.cc/paper/2020/file/79a3308b13cd31f096d8a4a34f96b66b-Paper.pdf.

Z.-Y. Ran and B.-G. Hu. Parameter Identifiability in Statistical Machine Learning: A Review. *Neural Computation*, 29(5):1151–1203, 05 2017. ISSN 0899-7667. doi: 10.1162/NECO_a_00947. URL https://doi.org/10.1162/NECO_a_00947.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. URL www.GaussianProcess.org/gpml.

N. G. Reich, J. Lessler, D. A. T. Cummings, and R. Brookmeyer. Lack of practical identifiability may hamper reliable predictions in COVID-19 epidemic models. *Science Advances*, 7(36):eabg5234, 2021. doi: 10.1126/sciadv.abg5234. URL https://doi.org/10.1126/sciadv.abg5234.

J. Ridgway. Computation of Gaussian orthant probabilities in high dimension. *Statistics and computing*, 26(4):899–916, 2016. doi: 10.1007/s11222-015-9578-1. URL https://doi.org/10.1007/s11222-015-9578-1.

G. Riutort-Mayol, P.-C. Bürkner, M. R. Andersen, A. Solin, and A. Vehtari. Practical Hilbert space approximate Bayesian Gaussian processes for probabilistic programming. *Statistics and Computing*, 33 (1):17, 2022. ISSN 1573-1375. doi: 10.1007/s11222-022-10167-2. URL https://doi.org/10.1007/s11222-022-10167-2.

H. Ruben. An asymptotic expansion for the multivariate normal distribution and Mills' ratio. *Journal of Research of the National Bureau of Standards Section B Mathematics and Mathematical Physics*, page 3, 1964. doi: 0.1017/S1446788700026872. URL https://api.semanticscholar.org/CorpusID:56130390.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009. doi: 10.1111/j.1467-9868.2008.00700.x. URL https://doi.org/10.1111/j.1467-9868.2008.00700.x.

H. Salje, C. T. Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. L. Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle, and S. Cauchemez. Estimating the burden of SARS-CoV-2 in France. *Science*, 369(6500):208–211, 2020a. doi: 10.1126/science.abc3517. URL https://www.science.org/doi/abs/10.1126/science.abc3517.

H. Salje, C. Tran Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. Le Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle, and S. Cauchemez. Estimating the burden of SARS-CoV-2 in France. *Science*, 369(6500):208–211, 2020b. ISSN 0036-8075. doi: 10.1126/science.abc3517. URL https://science.sciencemag.org/content/369/6500/208.

J. Salvatier, T. V. Wiecki, and C. Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55, 2016. ISSN 2376-5992. doi: 10.7717/peerj-cs.55. URL https://doi.org/10.7717/peerj-cs.55.

R. Sameni. Beyond Convergence: Identifiability of Machine Learning and Deep Learning Models. *arXiv preprint*, 2023. doi: arXiv:2307.11332. URL https://doi.org/10.48550/arXiv.2307.11332.

G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978. ISSN 00905364. URL http://www.jstor.org/stable/2958889.

S. Seaman and D. De Angelis. Challenges in estimating the distribution of delay from COVID-19 death to report of death, 2020. URL https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/2020/05/Challenges-in-estimating-the-distribution-of-delay-from-COVID-19-death-to-report-of-death-1.pdf.

R. Sender, Y. Bar-On, S. W. Park, E. Noor, J. Dushoff, and R. Milo. The unmitigated profile of COVID-19 infectiousness. *eLife*, 11:e79134, 2022. ISSN 2050-084X. doi: 10.7554/eLife.79134. URL https://doi.org/10.7554/eLife.79134.

X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf.

B. Singh, K. K. Sharma, S. Rathi, and G. Singh. A generalized log-normal distribution and its goodness of fit to censored data. *Computational Statistics*, 27:51–67, 2012. doi: 10.1007/s00180-011-0233-9. URL https://doi.org/10.1007/s00180-011-0233-9.

D. Sivia and J. Skilling. *Data analysis: a Bayesian tutorial*. OUP Oxford, 2006. URL https://books.google.co.uk/books/about/Data_Analysis.html?id=lYMSDAAAQBAJ&redir_esc=y.

J. Skilling. Nested Sampling for General Bayesian Computation. *Bayesian Analysis*, 1(4):833–860, 2006. URL https://projecteuclid.org/download/pdf_1/euclid.ba/1340370944.

A. F. Smith and D. J. Spiegelhalter. Bayes Factors and Choice Criteria for Linear Models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):213–220, 1980. URL https://www.jstor.org/stable/2984964.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4): 583–639, 2002. doi: 10.1111/1467-9868.00353. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00353.

E. W. Stacy. A Generalization of the Gamma Distribution. *The Annals of Mathematical Statistics*, 33(3): 1187–1192, 1962. URL http://www.jstor.org/stable/2237889.

Stan Development Team. PyStan: the Python interface to Stan. URL http://mc-stan.org.

T. Stocks, T. Britton, and M. Höhle. Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany. *Biostatistics*, 21(3):400–416, 2018. ISSN 1465-4644. doi: 10.1093/biostatistics/kxy057. URL https://doi.org/10.1093/biostatistics/kxy057.

L. Sun, C. Lee, and J. A. Hoeting. Parameter inference and model selection in deterministic and stochastic dynamical models via approximate Bayesian computation: modeling a wildlife epidemic. *Environmetrics*, 26(7):451–462, 2015. doi: 10.1002/env.2353. URL https://doi.org/10.1002/env.2353.

The Lancet. COVID-19 in Brazil: "So what?". *The Lancet*, 395:1461, 2020. doi: 10.1016/S0140-6736(20)31095-3. URL https://doi.org/10.1016/S0140-6736(20)31095-3.

L. Tierney and J. B. Kadane. Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986. doi: 10.1080/01621459.1986.10478240. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478240.

M. Titsias. Variational Learning of Inducing Variables in Sparse Gaussian Processes. In *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 567–574. PMLR, 16–18 Apr 2009. URL https://proceedings.mlr.press/v5/titsias09a.html.

D. Tran, A. Kucukelbir, A. B. Dieng, M. Rudolph, D. Liang, and D. M. Blei. Edward: A library for probabilistic modeling, inference, and criticism. *arXiv preprint*, 2016. doi: arXiv:1610.09787. URL https://doi.org/10.48550/arXiv.1610.09787.

N. Tuncer and T. T. Le. Structural and practical identifiability analysis of outbreak models. *Mathematical Biosciences*, 299:1–18, 2018. ISSN 0025-5564. doi: 10.1016/j.mbs.2018.02.004. URL https://www.sciencedirect.com/science/article/pii/S0025556417303164.

A. Vehtari and J. Ojanen. A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, 6:1935–7516, 2012. doi: 10.1214/12-SS102. URL https://doi.org/10.1214/12-SS102.

A. Vehtari, A. Gelman, and J. Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27:1413–1432, 2017. doi: 10.1007/s11222-016-9696-4. URL https://doi.org/10.1007/s11222-016-9696-4.

A. Vehtari, D. P. Simpson, Y. Yao, and A. Gelman. Limitations of "Limitations of Bayesian Leave-one-out Cross-Validation for Model Selection". *Computational Brain & Behavior*, 2:22–27, 2019. doi: 10.1007/s42113-018-0020-6. URL https://doi.org/10.1007/s42113-018-0020-6.

A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner. Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2), 2021. doi: 10.1214/20-ba1221. URL https://doi.org/10.1214/20-ba1221.

R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, T. Jamie, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, and G. Nedjati. Estimates of the severity of COVID-19 disease. *The Lancet Infec-*

tious Diseases, 20(6), 2020. doi: 10.1016/S1473-3099(20)30243-7. URL https://doi.org/10.1016/S1473-3099(20)30243-7.

D. Villela. How limitations in data of health surveillance impact decision making in the COVID-19 epidemic. *SciELO Preprints*, 2020. doi: 10.1590/SciELOPreprints.1313. URL https://doi.org/10.1590/SciELOPreprints.1313.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. URL https://doi.org/10.1038/s41592-019-0686-2.

S. Vitoratou and I. Ntzoufras. Thermodynamic Bayesian model comparison. *Statistics and Computing*, pages 1165–1180, 2017. doi: 10.1007/s11222-016-9678-6. URL https://doi.org/10.1007/s11222-016-9678-6.

L. Vočadlo and D. Alfè. Ab initio melting curve of the fcc phase of aluminum. *Phys. Rev. B*, 65, 2002. doi: 10.1103/PhysRevB.65.214105. URL https://link.aps.org/doi/10.1103/PhysRevB.65.214105.

E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, 593:266—269, 2021. doi: 10.1038/s41586-021-03470-x. URL https://doi.org/10.1038/s41586-021-03470-x.

P. G. T. Walker, C. Whittaker, O. J. Watson, M. Baguelin, P. Winskill, A. Hamlet, B. A. Djafaara, Z. Cucunubá, D. Olivera Mesa, W. Green, H. Thompson, S. Nayagam, K. E. C. Ainslie, S. Bhatia, S. Bhatt, A. Boonyasiri, O. Boyd, N. F. Brazeau, L. Cattarino, G. Cuomo-Dannenburg, A. Dighe, C. A. Donnelly, I. Dorigatti, S. L. van Elsland, R. FitzJohn, H. Fu, K. A. Gaythorpe, L. Geidelberg, N. Grassly, D. Haw, S. Hayes, W. Hinsley, N. Imai, D. Jorgensen, E. Knock, D. Laydon, S. Mishra, G. Nedjati-Gilani, L. C. Okell, H. J. Unwin, R. Verity, M. Vollmer, C. E. Walters, H. Wang, Y. Wang, X. Xi, D. G. Lalloo, N. M. Ferguson, and A. C. Ghani. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science*, 369(6502):413–422, 2020. ISSN 0036-8075. doi: 10.1126/science.abc0035. URL https://www.science.org/doi/10.1126/science.abc00350.

J. Wallinga and M. Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, 2007. doi: 10.1098/rspb.2006.3754. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2006.3754.

H. Wang, K. R. Paulson, S. A. Pease, S. Watson, H. Comfort, et al. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *The Lancet*, 399(10334):1513–1536, April 16 2022. ISSN 0140-6736. doi: 10.1016/S0140-6736(21)02796-3. URL https://doi.org/10.1016/S0140-6736(21)02796-3.

K. Wang, G. Pleiss, J. Gardner, S. Tyree, K. Q. Weinberger, and A. G. Wilson. Exact Gaussian Processes on a Million Data Points. 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. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/01ce84968c6969bdd5d51c5eeaa3946a-Paper.pdf.

S. Watanabe. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J. Mach. Learn. Res.*, 11:3571–3594, dec 2010. ISSN 1532-4435. URL https://www.jmlr.org/papers/volume11/watanabe10a/watanabe10a.pdf.

C. R. Wells, P. Sah, S. M. Moghadas, A. Pandey, A. Shoukat, Y. Wang, Z. Wang, L. A. Meyers, B. H. Singer, and A. P. Galvani. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proceedings of the National Academy of Sciences*, 117 (13):7504–7509, 2020. doi: 10.1073/pnas.2002616117. URL https://www.pnas.org/doi/abs/10.1073/pnas.2002616117.

WHO. Novel Coronavirus(2019-nCoV) Situation Report – 22. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2. 2019. Accessed on: March 5, 2024.

E. J. Williams. *Regression Analysis*. Wiley, 1959. URL https://archive.org/details/regressionanalys0000will.

A. G. Wilson and R. Adams. Gaussian Process Kernels for Pattern Discovery and Extrapolation. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of*

Machine Learning Research, pages 1067–1075, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL http://proceedings.mlr.press/v28/wilson13.html.

A. G. Wilson, Z. Hu, R. Salakhutdinov, and E. P. Xing. Deep Kernel Learning. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 370–378, Cadiz, Spain, 09–11 May 2016. PMLR. URL http://proceedings.mlr.press/v51/wilson16.htm.

D. Wolffram, S. Abbott, M. an der Heiden, S. Funk, F. Günther, D. Hailer, S. Heyder, T. Hotz, J. van de Kassteele, H. Küchenhoff, S. Müller-Hansen, D. Syliqi, A. Ullrich, M. Weigert, M. Schienle, and J. Bracher. Collaborative nowcasting of COVID-19 hospitalization incidences in Germany. *PLOS Computational Biology*, 19(8):1–25, 08 2023. doi: 10.1371/journal.pcbi.1011394. URL https://doi.org/10.1371/journal.pcbi.1011394.

World Health Organisation. WHO, Coronavirus disease 2019 (COVID-19) Situation Report – 1, 2020a. URL https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4.

World Health Organisation. WHO, Coronavirus disease 2019 (COVID-19) Situation Report – 11, 2020b. URL https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4.

World Health Organisation. WHO, Coronavirus disease 2019 (COVID-19) Weekly Epidemiological Update 14th September, 2020c. URL https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200914-weekly-epi-update-5.pdf?sfvrsn=cf929d04_2.

J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, B. J. C. Pablo M. de Salazar, M. Lipsitch, and G. M. Leung. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine*, 26:506–510, 2020a. doi: 10.1038/s41591-020-0822-7. URL https://doi.org/10.1038/s41591-020-0822-7.

J. T. Wu, K. Leung, and G. M. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395(10225):689–697, 2020b. doi: 10.1016/S0140-6736(20)30260-9. URL https://doi.org/10.1016/S0140-6736(20)30260-9.

J. T. Wu, K. Leung, N. K. Mary Bushman, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, and G. M. Leung. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *The Lancet*, 395:1382–1393, 2020c. doi: 10.1016/S0140-6736(20)30746-7. URL https://doi.org/10.1016/S0140-6736(20)30746-7.

Y. Wu, L. Kang, Z. Guo, J. Liu, M. Liu, and W. Liang. Incubation Period of COVID-19 Caused by Unique SARS-CoV-2 Strains: A Systematic Review and Meta-analysis. *JAMA Network Open*, 5(8): e2228008–e2228008, 08 2022. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2022.28008. URL https://doi.org/10.1001/jamanetworkopen.2022.28008.

W. Xie, P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic biology*, 60(2):150–160, 2011. doi: 10.1093/sysbio/syq085. URL https://doi.org/10.1093/sysbio/syq085.

X. Yang, Y. Yu, J. Xu, H. Shu, J. Xia, H. Liu, Y. Wu, L. Zhang, Z. Yu, M. Fang, T. Yu, Y. Wang, S. Pan, X. Zou, S. Yuan, and Y. Shang. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *The Lancet Respiratory Medicine*, 8, 2020. doi: 10.1016/S2213-2600(20)30079-5. URL https://doi.org/10.1016/S2213-2600(20)30079-5.

Y. Yang. Can the strengths of AIC and BIC be shared? A Conflict between Model Indentification and Regression Estimation. *Biometrika*, 92(4):937–950, 2005. URL http://www.jstor.org/stable/20441246.

F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y. Zhang, H. Chen, and B. Cao. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395:1054–1062, 2020. doi: 10.1016/S0140-6736(20)30566-3. URL https://doi.org/10.1016/S0140-6736(20)30566-3.

N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, and W. Tan. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*, 382(8):727–733, 2020. doi: 10.1056/NEJMoa2001017. URL https://doi.org/10.1056/NEJMoa2001017.

Appendix A

Appendix: Gaussian Process Nowcasting: Application to COVID-19 Mortality Reporting

Additional materials for Chapter 4.

A.1 Retrospective testing

The figures below show the additional retrospective tests as mentioned in the main text.

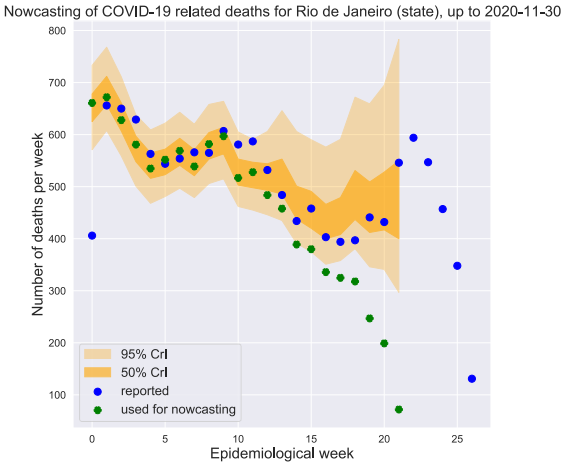


Figure A.1: Nowcasted and reported deaths due to COVID-19 death for Rio de Janeiro (state), generated with a 1D SE+SE data-split GP model. Reported deaths are shown in blue, nowcasted CrI in orange. Here the nowcasting was performed with all data available up till the SIVEP data release on the 30$^{\text{th}}$ Nov 2020. At that point, looking only at the reported data might indicate that the number of deaths keeps decreasing, however using nowcasting would have revealed the uptick in the number of deaths, which was not yet observed in the data at the time of the 30$^{\text{th}}$ Nov 2020 release.

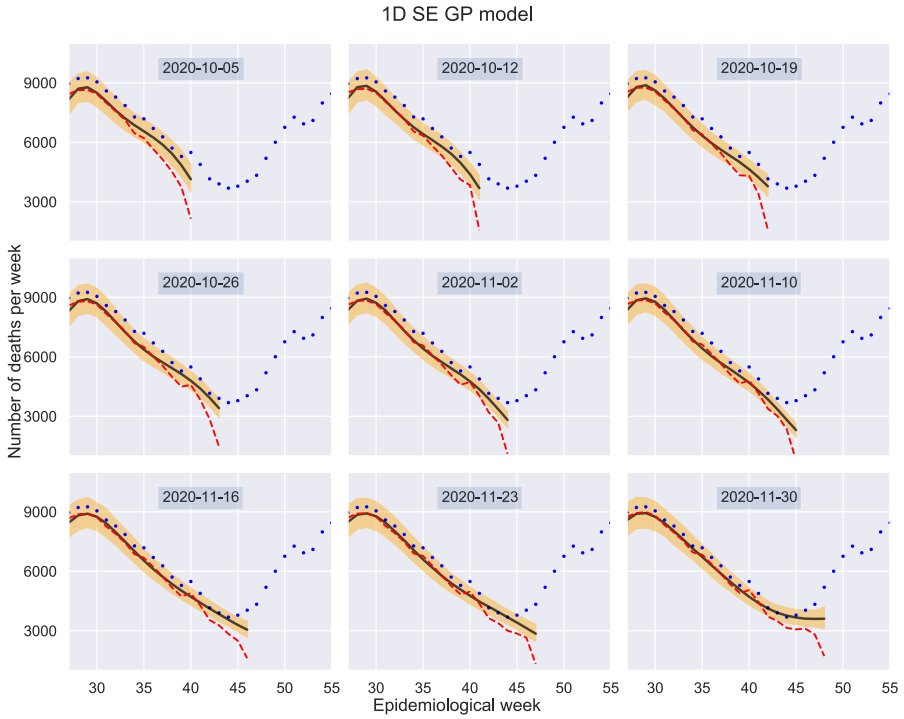


Figure A.2: Retrospective tests for the 1D SE GP nowcasting model. Deaths reported in the 31$^{\text{st}}$ May 2021 release are shown with blue dots, data available at the time of nowcasting with a red dashed line, nowcasted mean values with a black solid line and 95% CrI with orange ribbon.

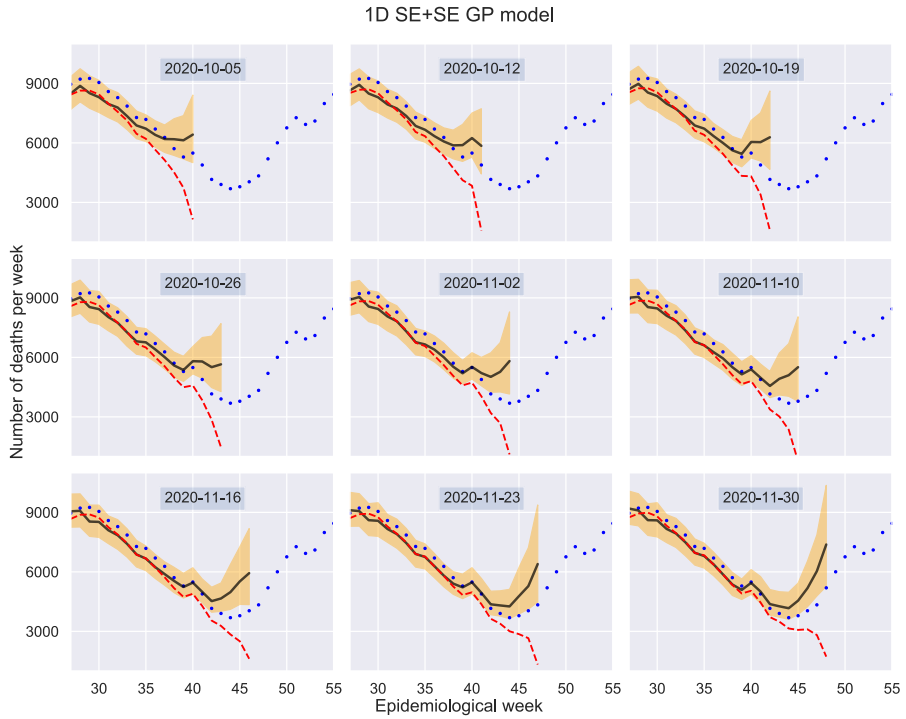


Figure A.3: Retrospective tests for the 1D SE+SE GP model.

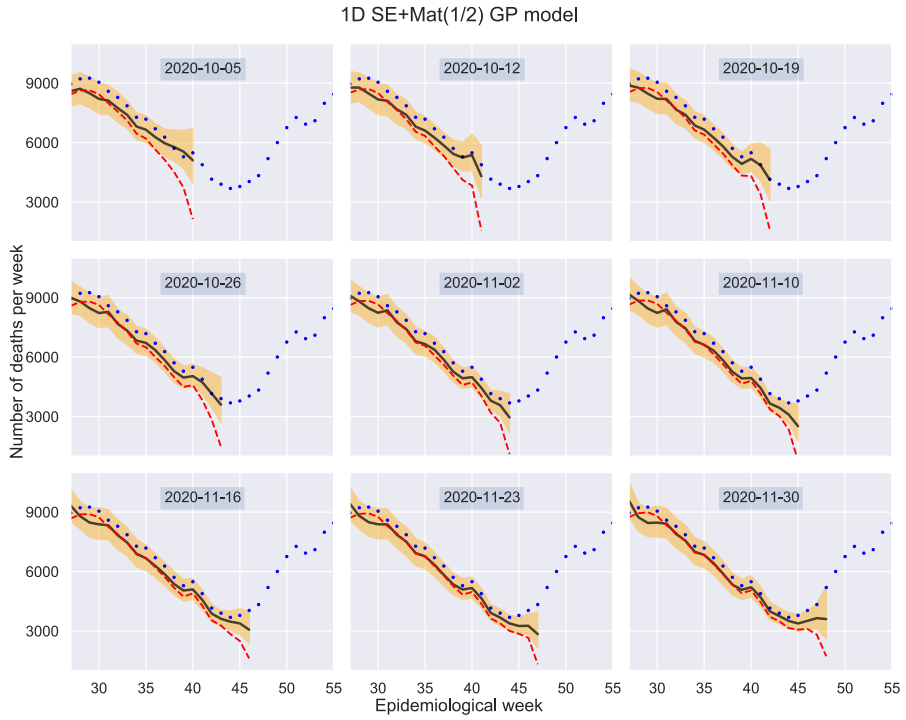


Figure A.4: Retrospective tests for the 1D SE+Mat(1/2) GP model.

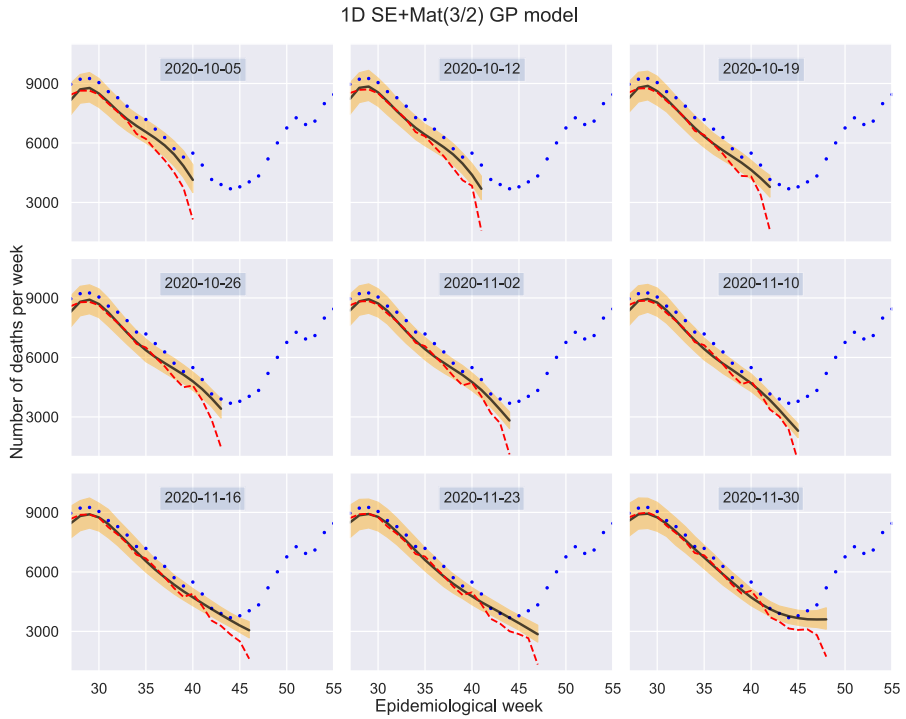


Figure A.5: Retrospective tests for the 1D SE+Mat(3/2) GP model.

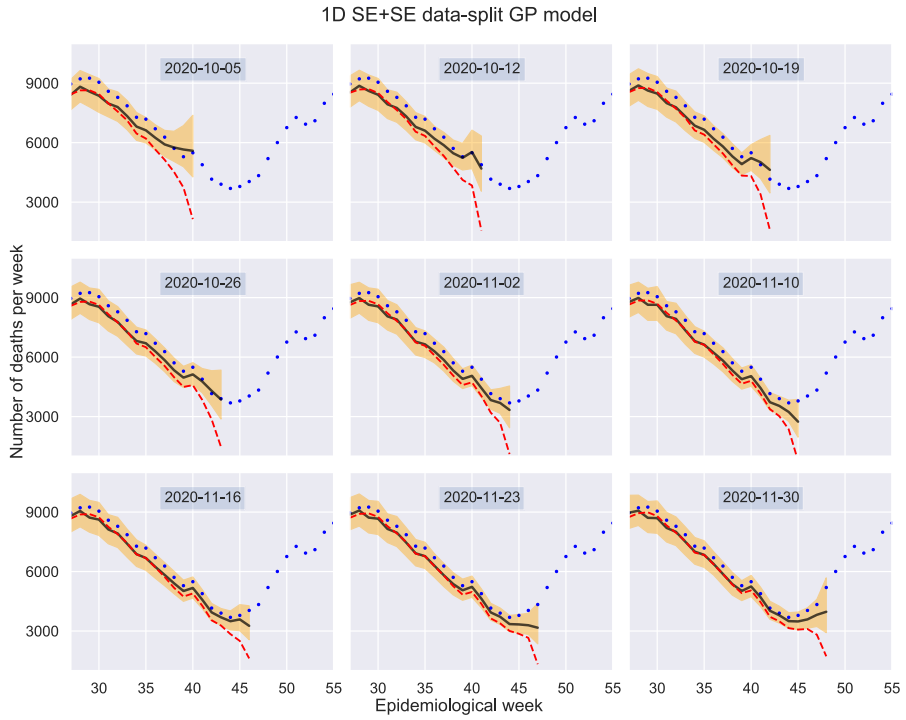


Figure A.6: Retrospective tests for the 1D SE+SE data-split GP model.

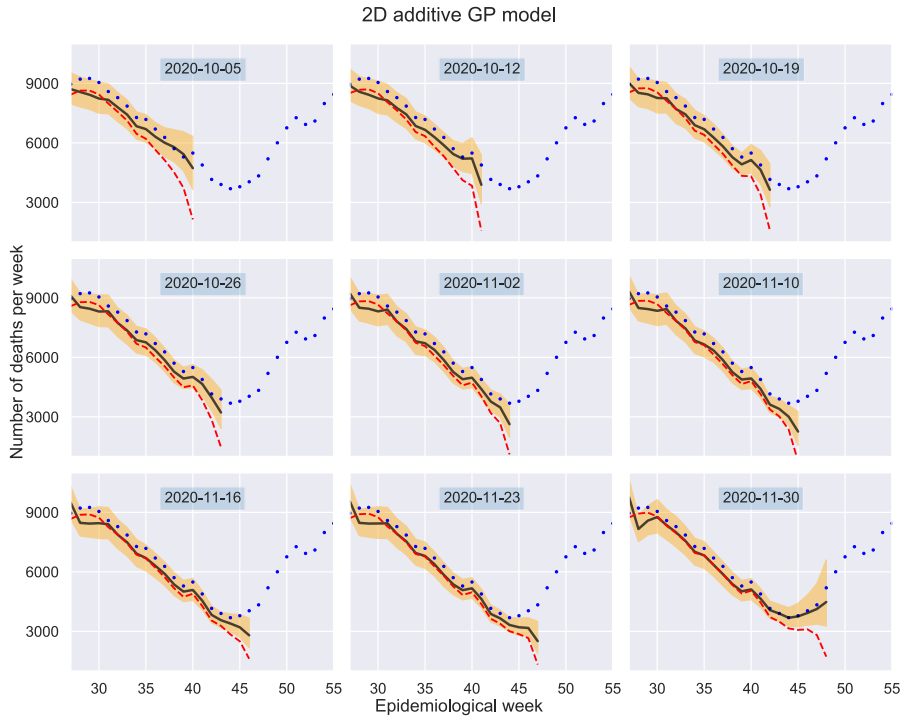


Figure A.7: Retrospective tests for the 2D additive kernel GP model.

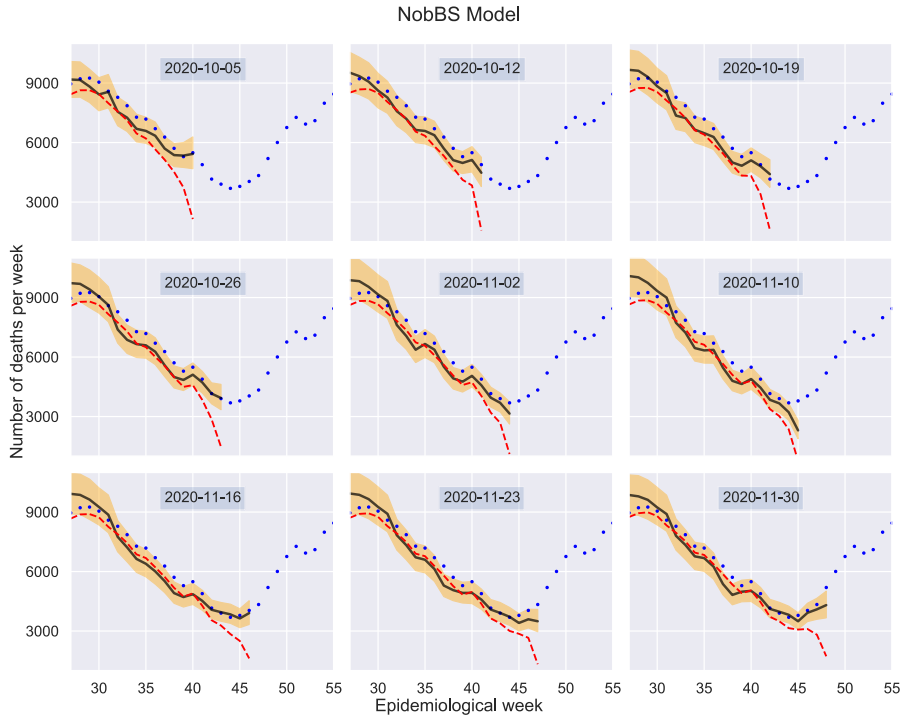


Figure A.8: Retrospective tests for the NobBS model.

A.2 Model diagnostics

For each of the model runs shown in this section, 4 chains were run for 1000 iterations, with 500 iterations used for burn-in. All fits presented are done for nowcasting using all data available up to 11$^{\text{th}}$ Jan 2021.

Table A.1: Diagnostics for the 1D SE+SE data-split GP model.

	mean	sd	hdi 3%	hdi 97%	mcse mean	mcse sd	ess bulk	ess tail	$\hat{R}$
$\rho_{1,\text{long}}$	28.004	0.101	27.820	28.194	0.002	0.001	2794.0	1241.0	1.00
$\rho_{2,\text{long}}$	29.009	0.102	28.826	29.209	0.002	0.002	2205.0	1377.0	1.01
$\rho_{1,\text{short}}$	1.003	0.010	0.984	1.021	<0.001	<0.001	2985.0	1388.0	1.00
$\rho_{2,\text{short}}$	1.013	0.009	0.996	1.032	<0.001	<0.001	2096.0	1691.0	1.00
$\alpha_{1,\text{long}}$	19.165	1.570	16.248	22.085	0.074	0.053	468.0	482.0	1.01
$\alpha_{2,\text{long}}$	23.366	1.234	21.145	25.692	0.084	0.060	219.0	288.0	1.02
$\alpha_{1,\text{short}}$	4.958	0.481	4.048	5.830	0.024	0.017	412.0	460.0	1.01
$\alpha_{2,\text{short}}$	5.941	0.279	5.447	6.463	0.014	0.010	378.0	886.0	1.01
δ_1	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	1448.0	795.0	1.00
δ_2	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	583.0	1188.0	1.01
r	246.499	11.073	225.669	267.226	0.217	0.154	2632.0	1399.0	1.00

Table A.2: Diagnostics for the 2D additive GP model.

	mean	sd	hdi 3%	hdi 97%	mcse mean	mcse sd	ess bulk	ess tail	$\hat{R}$
$\alpha_{1,t}$	29.169	1.015	27.467	31.164	0.018	0.013	3288.0	1544.0	1.01
$\alpha_{2,t}$	5.067	0.942	3.360	6.824	0.043	0.031	469.0	764.0	1.01
$\alpha_{1,d}$	0.548	0.119	0.346	0.778	0.005	0.004	471.0	680.0	1.01
$\alpha_{2,d}$	0.616	0.096	0.460	0.805	0.004	0.003	654.0	914.0	1.01
δ_1	<0.001	<0.001	<0.001	<0.001	0.000	<0.001	1776.0	854.0	1.00
δ_2	<0.001	<0.001	<0.001	<0.001	<0.0010	<0.001	1672.0	1110.0	1.01
r	89.733	6.478	77.809	102.541	0.176	0.126	1390.0	1365.0	1.00

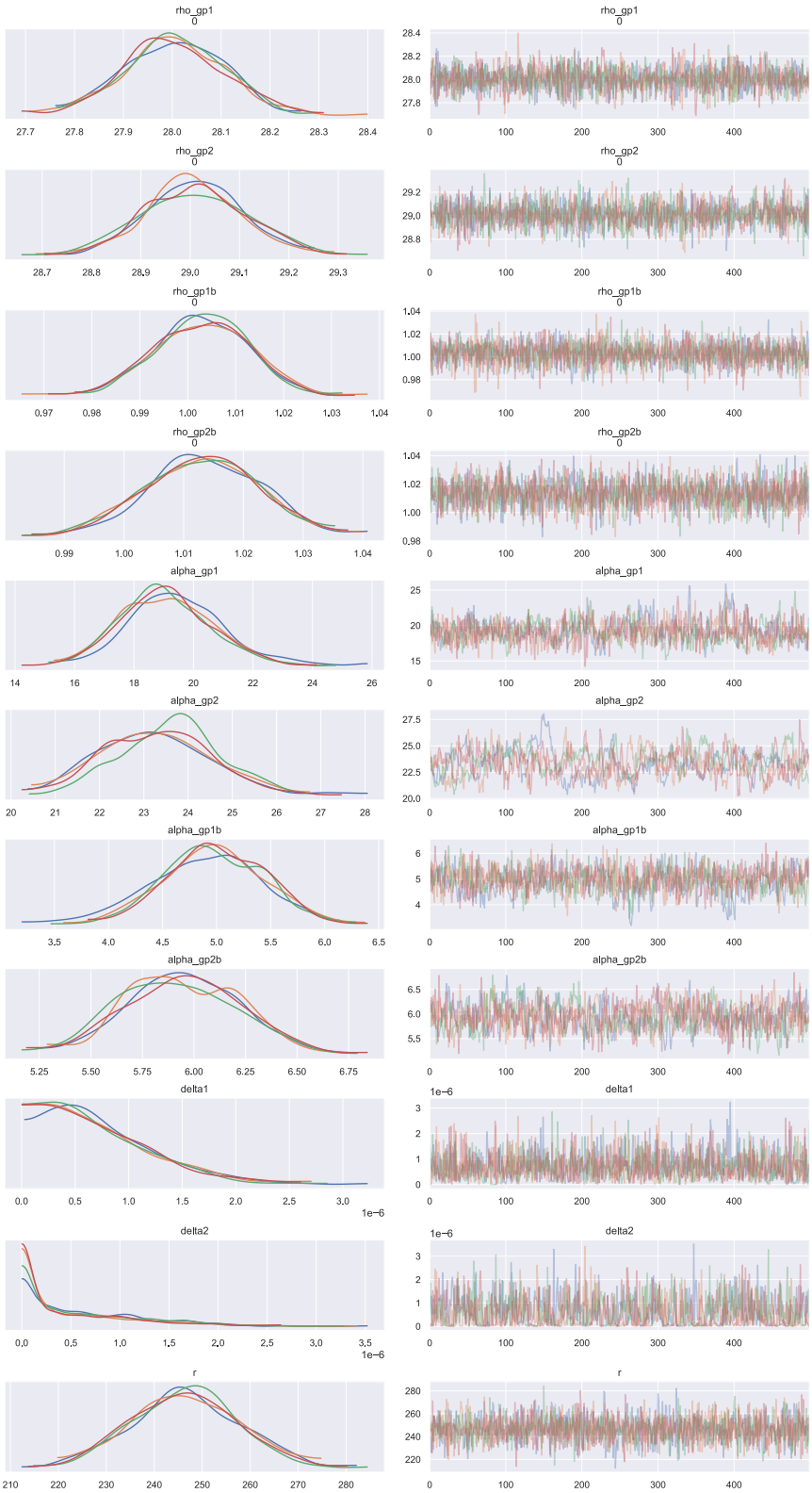


Figure A.9: Traceplots for the 1D SE+SE data-split GP model.

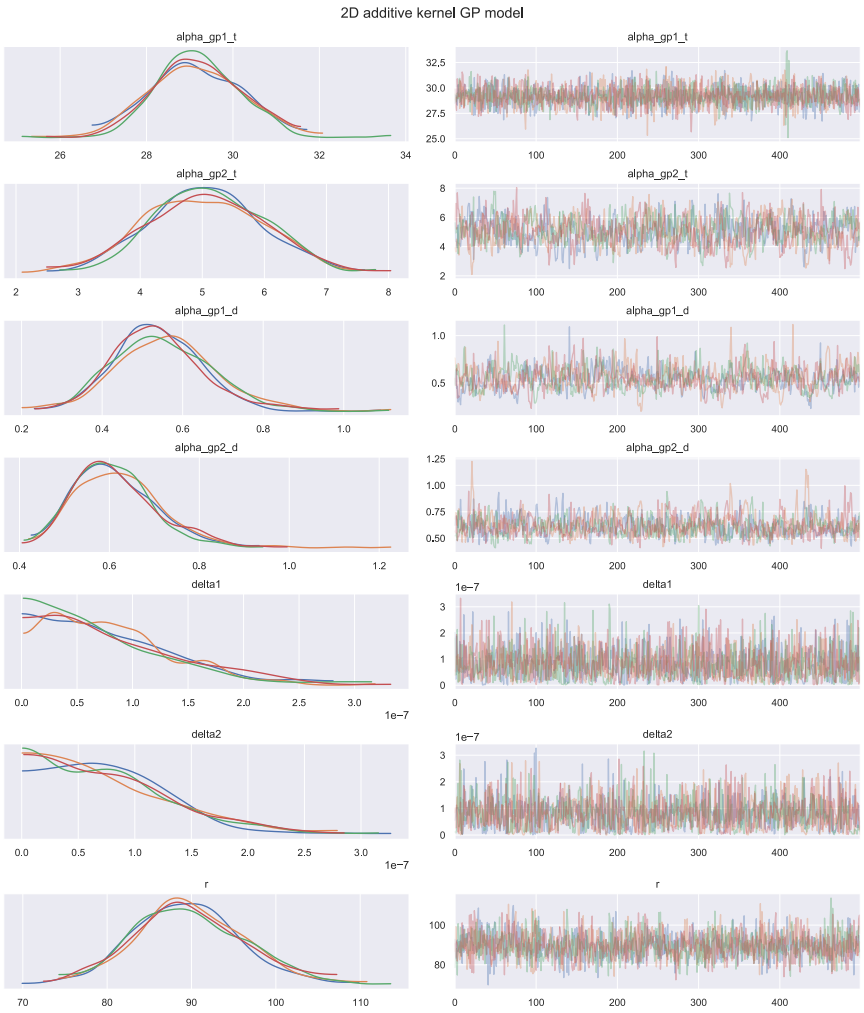


Figure A.10: Traceplots for the 2D additive GP model.

A.3 Sensitivity analysis

For all sensitivity analyses, the 1D SE+SE data-split GP model has been used. Each time we run the models with 4 chains for 1000 iterations, with 400 iterations used for burn-in. All fits presented are done for nowcasting using all data available up to 11$^{\text{th}}$ Jan 2021.

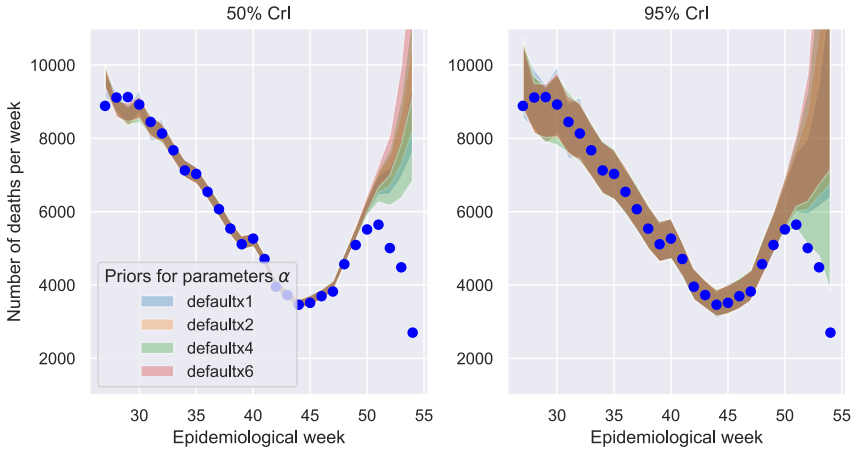


Figure A.11: Model fits with different $\alpha_{\text{long},1}$, $\alpha_{\text{long},2}$, $\alpha_{\text{short},1}$ and $\alpha_{\text{short},2}$ prior density variance.

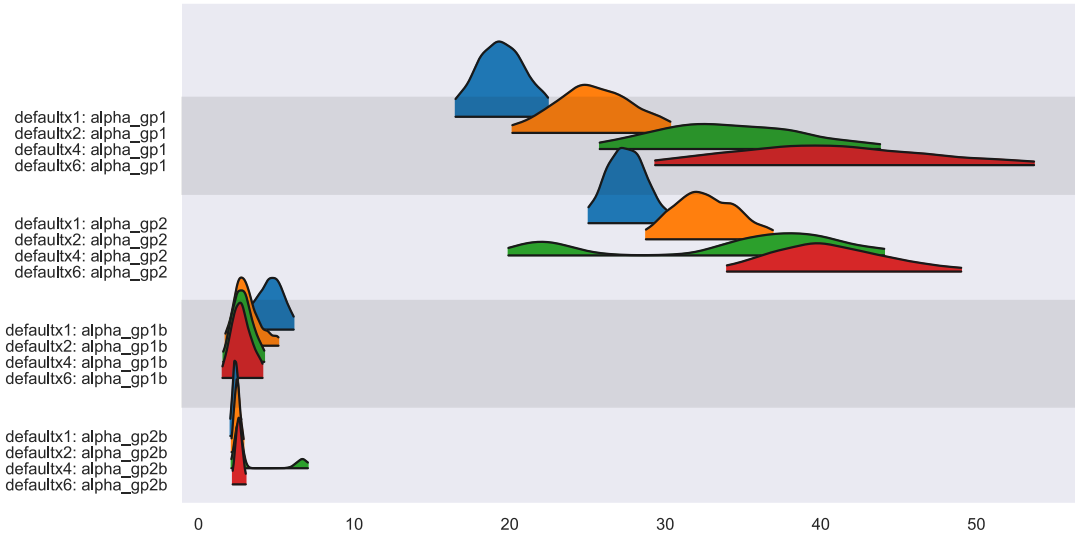


Figure A.12: Model fits with different $\alpha_{\text{long},1}$, $\alpha_{\text{long},2}$, $\alpha_{\text{short},1}$ and $\alpha_{\text{short},2}$ prior density variance.

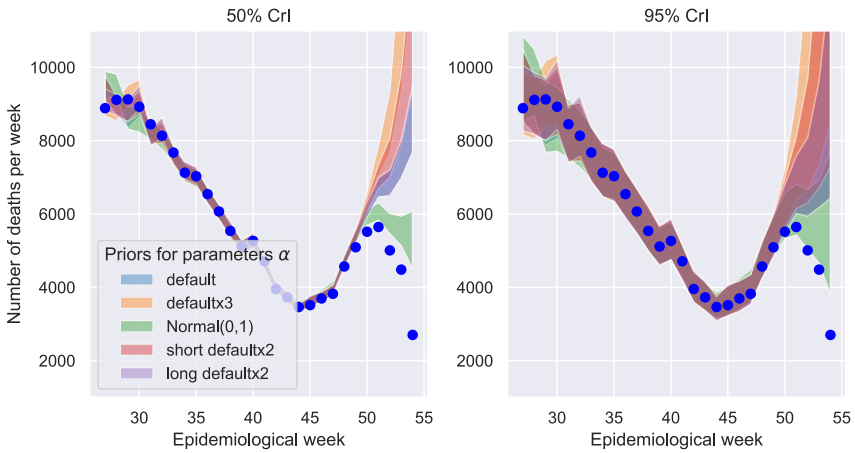


Figure A.13: Model fits with different $\alpha_{\text{long},1}$, $\alpha_{\text{long},2}$, $\alpha_{\text{short},1}$ and $\alpha_{\text{short},2}$ prior density. Default means using the default priors described in Table 4.1, for default x 3 prior we increased the mean in the default priors 3-fold, Normal(0,1) means a standard prior was set to all α-s, and long- and short- default x 2 means increased mean in the default prior long- and short-part respectively.

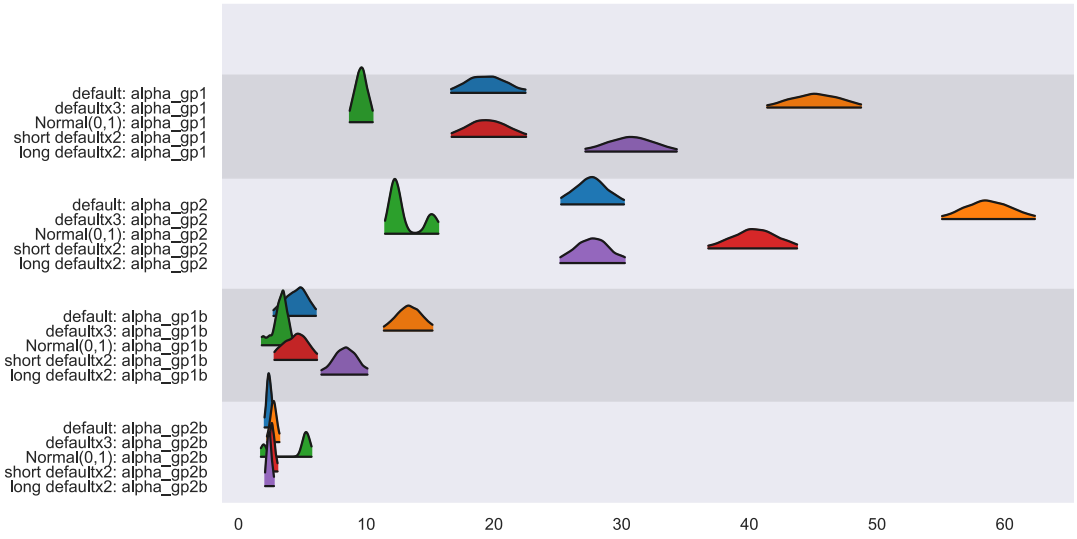


Figure A.14: Model fits with different $\alpha_{\text{long},1}$, $\alpha_{\text{long},2}$, $\alpha_{\text{short},1}$ and $\alpha_{\text{short},2}$ prior density. Default means using the default priors described in Table 4.1, for default x 3 prior we increased the mean in the default priors 3-fold, Normal(0,1) means a standard prior was set to all α-s, and long- and short- default x 2 means increased mean in the default prior long- and short-part respectively.