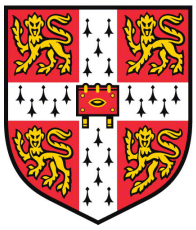
Statistical inference in stochastic/deterministic epidemic models to jointly estimate transmission and severity



Alice Corbella

Magdalene College
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

March 2019

Declaration

I hereby declare that this dissertation is the result of my own work, undertaken between October 2015 and March 2019. All work is original and includes nothing which is the outcome of work done in collaboration except as declared in the Acknowledgements and specified in the text.

This dissertation is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Alice Corbella
March 2019

Statistical inference in stochastic/deterministic epidemic models to jointly estimate transmission and severity

Alice Corbella

Abstract

This thesis explores the joint estimation of transmission and severity of infectious diseases, focussing on the specific case of influenza. Transmission governs the speed and magnitude of viral spread in a population, while severity determines morbidity and mortality and the resulting effect on health care facilities. Their quantification is crucial to inform public health policies, motivating the routine collection of data on influenza cases.

The estimation of severity is compromised by the high degree of censoring affecting the data early during the epidemic. The challenge of estimating transmission is that each influenza data source is often affected by noise and selection bias and individually provides only partial information on the underlying process.

To address severity estimation with high censored data, new methods, inspired by demographic models and by parametric survival analysis, are formulated. A comprehensive review of these methods and existing methods is also carried out.

To jointly estimate transmission and severity, an initial Bayesian epidemic model is fitted to historical data on severe cases, assuming a deterministic severity process and using a single data source. This model is then extended to describe a more stochastic and hence more realistic process of severe events, with the data generating process governed by hidden random variables in a state-space framework. Such increased realism necessitates the use of multiple data sources to enhance parameter identifiability, in a Bayesian evidence synthesis context. In contrast to the literature in the field, the model introduced accounts for dependencies between datasets. The added stochasticity and unmeasured dependencies result in an intractable likelihood. Inference therefore requires a new approach based on Monte Carlo methods.

The method proposed proves its potential and usefulness in the concluding application to real data from the latest (2017/18) epidemic of influenza in England.

There is nothing like looking,
if you want to find something [...].
You certainly usually find something,
if you look, but it is not always
quite the something you were after.

J. R. R. Tolkien, *The Hobbit*

This thesis is dedicated to my families in England, who opened their homes and shared this adventure with me.

Acknowledgements

This thesis would never have even started without the help and support of the people acknowledged below and of many others that I cannot list here, otherwise I would have to add another chapter.

I am deeply thankful to my supervisors Dr Anne Presanis and Dr Daniela De Angelis. The patience and care that Anne put into revising and correcting this work is only the tip of the iceberg of her continuous support over these 3 and a half years. Anne persistently lifted me up, always helping me to find a positive starting point, even in the most confused moments of my research. I am grateful to Dani for teaching me continuously that an apparent problem is in reality an even more interesting aspect of research and for passing onto me her passion for research. Despite not being my official supervisor, Dr Paul Birrell has invested so much time and effort in discussing my Poisson models and SMC samplers, with his contribution improving this thesis so substantially, for which I am truly grateful. Thank you Anne, Dani and Paul for having made these years so memorable and research so enjoyable.

I shall also thank my advisers Dr Lorenz Wernish and Dr Christopher Jackson for their help and advice during these years, along with Dr Mikhail Shubin for the stimulating discussion around Chapter 5 of this thesis.

I thank my examiners Dr Paul Kirk and Dr TJ McKinley for the engaging discussion and the fruitful observations. Their suggestions improved the quality of this work to a large extent.

This thesis has been substantially enriched by collaboration with Pubic Health England. Hence, I would like to thank Dr Richard Pebody, Dr Xu-Sheng Zhang and Dr Andre Charlett. I also acknowledge the financial support of the Medical Research Council for this project.

Many other people have made my time at the MRC Biostatistics Unit enjoyable and fruitful. First of all, my fellow PhD students: those who took care of me when I arrived, particularly Fran; and those who were my companions in this adventure, particularly Brieuc, Alessandra and Chiara. Further thanks go to Colin for all his help with the high-performance computer and to the postdocs and members of the Unit that contribute to create a stimulating environment. I am grateful to master Nicholas Town for providing a warm welcome to Warwick and for the support and useful discussion during the final period of my PhD.

I am full of gratitude towards all the families that have unconditionally opened their homes to me and to which this thesis is dedicated. Thank you, Irene, Maddi, Cha e Luci for letting me be part of your life; thank you Bossi for welcoming me in your family; thank you Mazzi, Tazzari, Larcoms, Rognins, Youngs, Piffer, Rosotti and Gerosa. Many thanks also to the little ones from the Yellow House who keep teaching me the importance of simplicity and to all my friends who passed by Cambridge, entering my life unexpectedly.

I have been receiving continuous support from my friends at home and from around the world during these years. Thank you, Silvi and Jack, Agne and Eli, don Ezio, Meri, Bleis, Rich, Teddi and Bella.

Much love to my family that have seen the worst of me over these years and have always been present and supportive. Thank you, mamma Maria e papà Alfonso; you keep showing me and teaching me that I am loved so much that I shall not fear anything. I am full of gratitude towards my siblings Pietro, Caterina, Martina, Giuditta and Vova because you face your lives with bravery and gladness, and this is a marvel to watch.

A final enormous thank goes to my husband Francesco who is always at my side. The road in front of us is still long and my passion to walk it with you keeps increasing.

Abstract

This thesis explores the joint estimation of transmission and severity of infectious diseases, focussing on the specific case of influenza. Transmission governs the speed and magnitude of viral spread in a population, while severity determines morbidity and mortality and the resulting effect on health care facilities. Their quantification is crucial to inform public health policies, motivating the routine collection of data on influenza cases.

The estimation of severity is compromised by the high degree of censoring affecting the data early during the epidemic. The challenge of estimating transmission is that each influenza data source is often affected by noise and selection bias and individually provides only partial information on the underlying process.

To address severity estimation with high censored data, new methods, inspired by demographic models and by parametric survival analysis, are formulated. A comprehensive review of these methods and existing methods is also carried out.

To jointly estimate transmission and severity, an initial Bayesian epidemic model is fitted to historical data on severe cases, assuming a deterministic severity process and using a single data source. This model is then extended to describe a more stochastic and hence more realistic process of severe events, with the data generating process governed by hidden random variables in a state-space framework. Such increased realism necessitates the use of multiple data sources to enhance parameter identifiability, in a Bayesian evidence synthesis context. In contrast to the literature in the field, the model introduced accounts for dependencies between datasets. The added stochasticity and unmeasured dependencies result in an intractable likelihood. Inference therefore requires a new approach based on Monte Carlo methods.

The method proposed proves its potential and usefulness in the concluding application to real data from the latest (2017/18) epidemic of influenza in England.

Preface

Most of the work presented in this thesis will be published or has already been published in scientific journals. Specifically:

- The case study employed to illustrate transmission models in Section 3.3 has been published in BMC Public Health in 2018 (Corbella et al., 2018).
- The work contained in Section 4.4, where an extension of standard sequential Monte Carlo (SMC) methods is proposed, will be further extended with derivation of the results presented, proof of the validity of the algorithm and, possibly, further applications and examples. It will then be submitted to a statistics journal.
- The models, methods and the simulation study presented in Chapter 5 will be further extended with a comparison on real data and submitted to a statistics journal.
- The work presented in Chapter 6 is self contained, providing a complete analysis of influenza severity and transmission during the 2017/18 epidemic. The content of this chapter will be submitted to an epidemiology journal or a science dissemination journal.

Contents

List of Figures

List of Tables

List of Abbreviations

HA Haemagglutinin

NA Neuraminidase

CFR case fatality risk

hCFR hospitalised case fatality risk

iCFR intensive-care case fatality risk

sCFR symptomatic case fatality risk

tvFR time-varying case fatality risk

IFR infection fatality risk

APC age-period-cohort

iid independent and identically distributed

SMC sequential Monte Carlo

SSM state space model

MSM multi-state model

KM Kaplan-Meier

ML maximum likelihood

MISE mean integrated squared error

HMM hidden Markov model

POMP partially observed Markov process

r.v. random variable

MH Metropolis Hastings

MCMC Markov chain Monte Carlo

MC Monte Carlo

USISS UK Severe Influenza Surveillance System

PHE Public Health England

HPA Health Protection Agency

ILI influenza-like illness

GP General Practitioner

NHS National Health Service

ICU Intensive Care Unit

HDU High Dependence Unit

WHO World Health Organization

ONS Office of National Statistics

ABC Approximate Bayesian Computation

CrI credible interval

DAG directed acyclic graph

AIC Akaike information criterion

RCGP Royal College of General Practitioners

PCR Polymerase chain reaction

BPF bootstrap particle filter

GIMH grouped independence Metropolis Hastings

MCWM Monte Carlo within Metropolis

CI confidence interval

HMC Hamiltonian Monte Carlo

Basic nomenclature

Find below some common notation used over the chapters, when exceptions are made, this is specified explicitly in the text.

$\sim$	distributes according to
$\propto$	it is proportional to
$\vert$	conditional on/given
X, Y, Z	capital latin letters are used to represent random variable
x, y, z	small latin letters are used to represent a given value of the respective random variable
$X_{a:b}$	sequence of random variables indexed from a to b, short form of $(X_a, X_{a+1}, \ldots, X_{b-1}, X_b)$
$x_{a:b}$	sequence of values indexed from a to b, short form of $(x_a, x_{a+1}, \ldots, x_{b-1}, x_b)$
α, β	greek letters are used to represent parameters, parameters can also be random variables
$\boldsymbol{\alpha}, \boldsymbol{\beta}$	bold greek letters are used to represent vectors of parameters
$P(A)$	probability that the event A takes place
$f(x)/p(x)$	probability density function/probability mass function of the continuous/discrete random variable X evaluated in x
$F(x)$	distribution function of the random variable X evaluated in x, short form of $P(X \leq x)$
$\mathbb{I}_{\{\texttt{bool}\}}$	indicator function: $\mathbb{I}_{\{\texttt{bool}\}} = \begin{cases} 1, & \text{if } \texttt{bool} \text{ is true} \\ 0, & \text{if } \texttt{bool} \text{ is false} \end{cases}$

Chapter 1

Introduction

The importance of the analysis of infectious diseases from a public health perspective is summarised in the devastating consequences of epidemics.

Precisely 100 years ago the epidemic of Spanish Flu emerged and 50 to 100 million people are estimated to have died (Patterson and Pyle, 1991); the AIDS epidemic is estimated to have caused 35.4 million deaths until 2017 (UNAIDS, 2018) and the more recent Ebola epidemic in west Africa in 2014 resulted in more than 10 thousand deaths (WHO, 2016). These are just few examples that motivate the effort to estimate and predict infectious diseases' transmission, with an ultimate aim of informing public health decisions to mitigate and/or prevent epidemics.

Besides, there are many reasons for which the investigation of epidemics has gathered attention of statistician and mathematical modellers. Niels G. Becker, in his 1989 book summarised them saying that:

"Firstly, infectious disease data are not the result of planned experiments, but arise from naturally occurring epidemics. Secondly, infectious disease data are highly dependent because infected cases are the cause of further infected cases. Thirdly, the infection process is generally only partially observable" (Becker, 1989)

Motivated by these stimuli and by the public health need, the literature on epidemic models and their inference has proliferated in the last two centuries (Heesterbeek et al., 2015).

1.1 Influenza virus and public health response

1.1.1 A public health threat

This year marks the centenary of the Spanish influenza pandemic. Since the emergence of this virus in the human population, pandemic events and seasonal epidemics have threatened world health almost continuously.

Brief history of influenza pandemics

Influenza outbreaks can be broadly classified into seasonal epidemics and pandemics; the latter being defined as epidemics of worldwide spread. Often a pandemic event takes place when the majority of the population does not have immune defences against the circulating virus.

There have been four human pandemics in the last century, in correspondence with reassortment events and contacts across species, as illustrated in Figure 1.1.

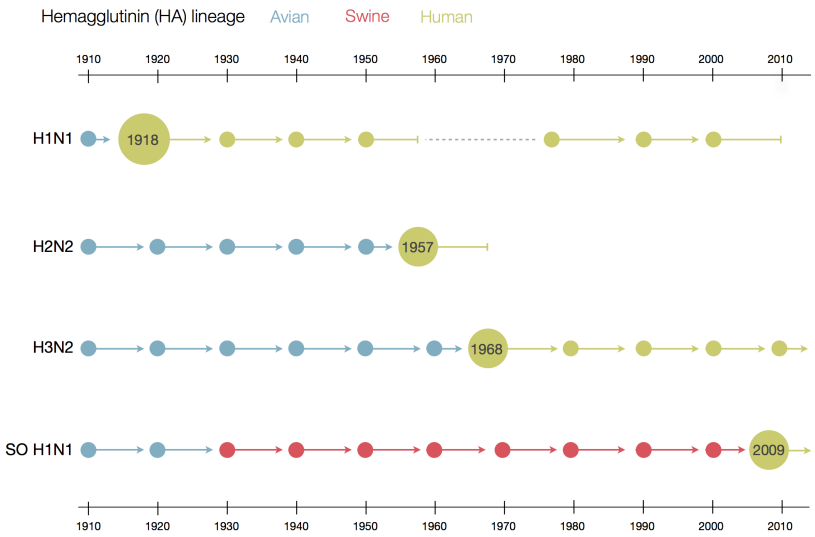


Figure 1.1: Representation of the pathogen evolution in the population of influenza A virus and its sub-types currently present in the population (image from Trevor Bedford, Course in *Pathogen evolution, selection and immunity*, Seattle 2016).

A hundred years ago the Spanish 'flu invaded all nations worldwide. This pandemic is estimated to have infected 1/3 of the world population, causing over 50 million deaths, and is thought to have emerged straight from bird to human contact (Taubenberger and Morens, 2006). The sub-type of the Spanish 'flu was H1N1 and this strain disappeared in 1957 when a shifted virus (H2N2) was transmitted from birds to humans. The same event happened a decade later and led to the current circulating H3N2 virus (Palese, 2004). Ten years ago a new variant of the H1N1 virus emerged in Mexico. Despite being of the same sub-type as the Spanish 'flu, this particular strain was the result of years of between and within species evolution, towards which only few individuals had immunity (Neumann, Noda, and Kawaoka, 2009).

Influenza A/H1N1/Pdm9 and influenza A/H3N2 are currently circulating seasonally, together with influenza B that is present only in humans. This characteristic prevents antigenic shift, practically granting no influenza B pandemic.

Current burden of the disease

Influenza might resolve in an asymptomatic infection or result in mild symptoms such as fever or feverishness, headache, muscle pain, general feeling of ill-health, runny nose, sore throat and non-productive cough. Moreover, in the most at risk groups, severe symptoms might arise, potentially leading to hospitalization, Intensive Care Unit (ICU) admission or death (ECDC, 2018).

Seasonal influenza has been recently shown to be the infectious disease with the highest burden on health systems in Europe (Cassini et al., 2018). This result is due both to the high incidence of influenza in the population and to the high mortality within cases, as can be observed in Figure 1.2 from Cassini et al. (2018).

[1] "One DALY can be thought of as one lost year of *healthy* life. The sum of these DALYs across the population, or the burden of disease, can be thought of as a measurement of the gap between current health status and an ideal health situation where the entire population lives to an advanced age, free of disease and disability." (Definition from WHO (2014))

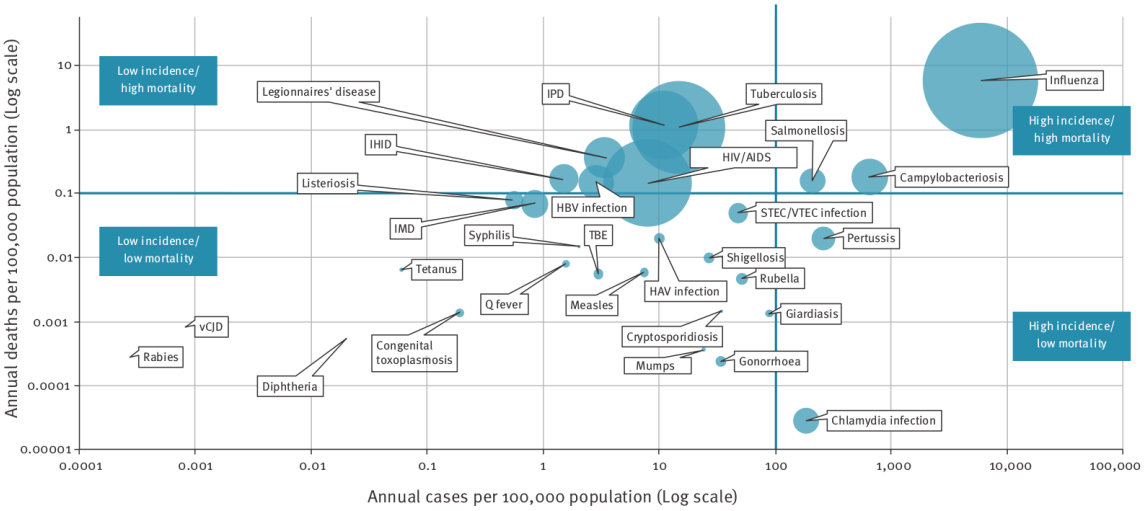


Figure 1.2: Bubble chart of the burden of selected infectious diseases in terms of mortality and incidence, EU/EEA countries, 2009-2013. The diameter of the bubble reflects the number of DALYs [1]per 100,000 population per year. Figure taken from Cassini et al. (2018).

Disparate interventions exists to lighten the load of influenza on public health systems, including: vaccination policies (Baguelin et al., 2010); school closures to reduce transmission in a pandemic (Te Beest et al., 2015; Vynnycky and Edmunds, 2008); reinforced use of antiviral drugs (Ferguson et al., 2006); or changes in hospital management policies.

The magnitude of influenza burden and the need to assess the potential effectiveness of these public health policies motivated the quantification and prediction of the number of infected people at different levels of severity. Indeed many health facilities are put under pressure during influenza epidemics: the flow of influenza-like illness (ILI) patients into General Practitioner (GP) clinics increases, the beds in hospital, particularly in ICUs cases would be highly occupied by 'flu severe cases etc. Inference and prediction of the transmission and severity of each influenza epidemic is therefore key to support preparedness and response so that hospitals may free beds to admit a high number of cases, GP clinics might be ready for the increasing demand etc.

The reason why influenza is still present in humans, after so many years of active fight by public health systems, is to be found in its biologic characteristics and extremely rapid evolution (Palese, 2004) which are illustrated below.

1.1.2 Biology and classification of the influenza virus

The influenza virus is a single-strand RNA virus of the family of Orthomyxoviridae (Taubenberger and Morens, 2008). Influenza viruses are divided into four types (A, B, C and D) according to their genetic and biological composition. Influenza A and B are the types currently causing most epidemics and are very similar in their virion structure (WHO, 2018b). The RNA (i.e. the genetic material to be replicated during infection) is enclosed in a lipid envelope covered with Haemagglutinin (HA) and Neuraminidase (NA) proteins (Figure 1.3). There exist several different types of these proteins.

These external proteins are key for the infection process: once a virion enters the organism, HA proteins allow it to lock to the external membrane of the cell, to enter it via endocitosis

and to move towards the cell nucleus for replication. Once the virus RNA has multiplied, child viruses invade the cell and eventually, thanks to the action of NA proteins, are released to infect other cells of the organism (Webster et al., 2013).

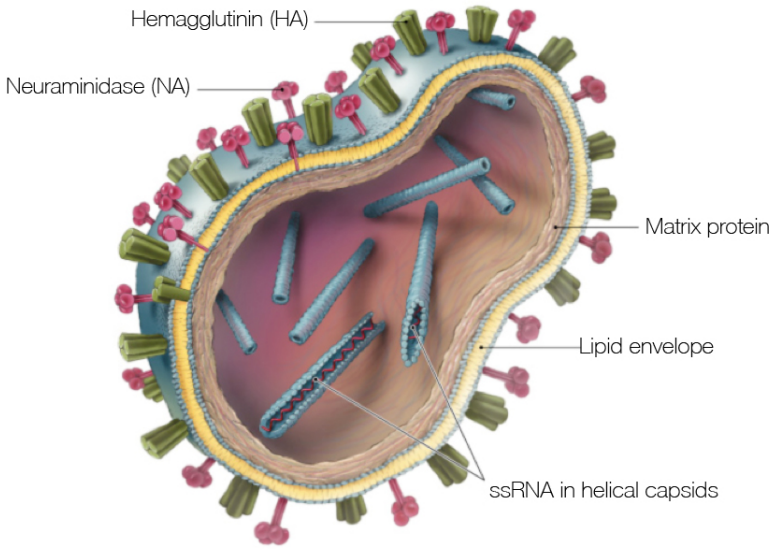


Figure 1.3: Representation of an influenza A virion (image from Trevor Bedford, Course in *Pathogen evolution, selection and immunity*, Seattle 2016).

The immune system responds to infection by producing antibodies that operate on the HA proteins. Once the virus is identified, antibodies bind with HA proteins, preventing the virion from entering the host cell or from releasing genetic material (NIAID, 2012).

The antibody response is compromised by the continuous evolution of the influenza virus so that antibodies are not able to bind to its HA proteins any more. There are three ways in which a virus can evolve. An *antigenic shift* takes place when the HA or NA type changes. This event usually corresponds to an interaction of different species and the consequent adaptation of the virus to a new specie. Antigenic shifts happen only for influenza A and lead to its classification into sub-types according to the type of HA and NA protein, which is indicated by the ordered number of their discovery; for example, influenza A/H1N1, which was the first influenza virion to be observed, or influenza A/H3N2. An *antigenic drift* is more common and corresponds to a punctual mutation in one or more of the amino acids of the HA or NA proteins and can take place every time the virus is replicated within the organism. These mutations give rise to different lineages of influenza within the same sub-type and they are identified by the location where the virus was first observed or by the respective pandemic event (e.g. influenza B/Victoria, influenza A/H3N2/Panama, influenza A/H1N1/pdm09) (Treanor, 2004). A last key mutation is *genetic reassortment*, that takes place when two different viruses are jointly present within the same organism. If this is the case the genetic material is merged and this leads to a new strain of influenza.

Once an organism is fully infected he might infect other organisms. Transmission takes place via mainly three routes: droplets, aerosol and direct contact. When an infected individual sneezes or coughs he produces expiratory sprays made of small particles that might be inhaled and reach the upper respiratory tract (droplets), and of even smaller particles that might reach the lower respiratory tract (aerosol). In direct contact instead, particles are transferred to

another organism directly or via a contaminated object/person, that is, indirectly (Killingley and Nguyen-Van-Tam, 2013).

1.2 Measuring an influenza epidemic

The incidence and mortality of an infectious disease are determined by its transmission and severity. Quantities defining these two features are proposed here, for which the goal is inference.

1.2.1 Characterizing severity

An individual infected with the influenza virus might experience more or less severe symptoms and events, which are often one subsumed in the other.

The severity process can be thought as a pyramid, such as the one represented in Figure 1.4 (De Angelis and Presanis, 2018). The quantities of interests are then the probability of specific severe events (e.g. the probability of death given hospitalization or the probability of hospitalization given infection etc.). However, the probability of death conditional on the individual's status is one of the most relevant and studied severity measures.

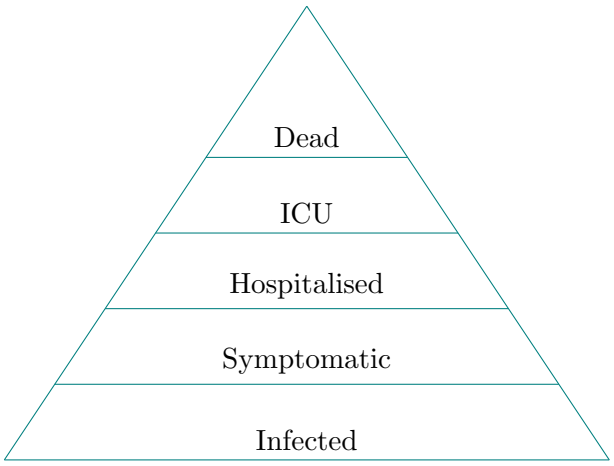


Figure 1.4: Severity of influenza as a pyramid.

This quantity is often called the case fatality risk (CFR), which is defined as the probability that a person dies from the infection given that he is a case. The literature has not agreed yet on the computation of this quantity as it depends on which individuals are categorized as "cases" (Lipsitch et al., 2015). Several studies focus on estimating the infection fatality risk (IFR), i.e, the probability of death given infection. However, this quantity is difficult to determine because, on the one hand, the number of infected people is not available (mainly due to asymptomatic cases and under reporting), and on the other hand, registers may misclassify infection-related death. Other measures of severity are the symptomatic CFR (i.e. the probability of death among patients with symptoms) and the hospitalized CFR (i.e. the probability of death among hospitalized patients).

This thesis approaches the general estimation of the probability of a severe event conditionally on a less severe condition. Nevertheless, the main application of the methods to estimate severity will be the quantification of the hospitalization/intensive-care CFR.

1.2.2 Characterizing transmission

The influenza virus, similarly to many other infectious diseases, makes the infected host undergo a latent period, when the virus spreads within the host body, and an infectious period, when the host might spread the virus to other individuals, before recovering and becoming immune to the specific strain of influenza that infected him. Consecutive events of this type (i.e. infections) make an epidemic take off; quantifying the transmission means measuring the magnitude and speed of the epidemic spread within a population.

Mechanistic models assume the infection process described above, with factors that determine transmission of a virus being: (i) the rate of effective contacts within the population, where an effective contact is a contact close enough to potentially lead to infection; (ii) the probability of infection given contact; and (iii) the length of the infectious period. While the first factor is a characteristic of the population, the latter two are determined by biological properties of the circulating virus (Keeling and Rohani, 2011). A key quantity to measure transmission that depends on these three factors is the basic reproduction number R_0, defined as the average number of new infections generated by an infectious individual when introduced in a population fully susceptible to the virus. The level of immunity in the population also affects the spread of the disease and is accounted for in another index targeted by this thesis: $R_e(0)$, defined as R_0 in a population with a specific level of immunity. An illustration of the spread of the disease for specific values of these parameters is reported in Figure 1.5.

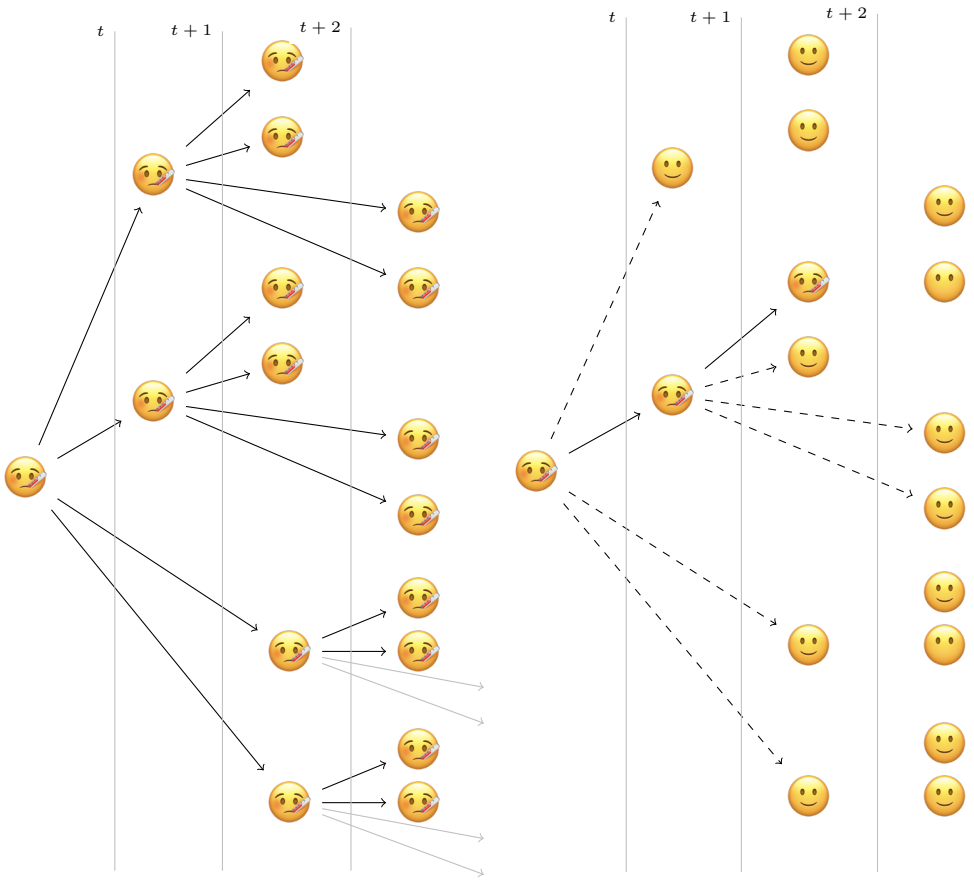


Figure 1.5: Cartoon of the spread of a disease with $R_0 = 4$ and average infectious period of two days. An infectious individual (🤧) is introduced in a fully susceptible population (left panel) and in a population where 75 % of the individuals are immune to the virus (🙂) and 25% are susceptible (😶) (right panel). Figure freely adapted from Vynnycky and White (2010), Chapter 1.

1.3 Statistical methods for modelling and inferring severity and transmission

To perform inference of severity and transmission some choices must be made of the model to approximate the data-generating process and on the inferential approach. The branches of statistics that are used to model transmission and severity and to infer the quantities summarising each, are briefly reviewed here.

1.3.1 Modelling approaches

In each specific context and depending on the type of data available, disparate methods can be used.

Mechanistic epidemic models are a useful tool for modelling the spread of a virus in a population. There is a substantial breadth of literature regarding these models alone that will be reviewed in Chapter 3. Mechanistic epidemic models are specific cases of stochastic processes (Chiang, 1975), where there is dependence across states. In particular, the number of new infections at the next time point will depend on the current number of infectious and susceptible individuals. They can also be approached within a state space models (SSMs) perspective, where

SSMs are themselves special cases of stochastic processes (Schön and Lindsten, 2017).

Non-mechanistic models can be also used to analyse infectious disease data. The *HHH model* (Held, Höhle, and Hofmann, 2005) is based on decomposing the variability of the data into an epidemic component, which is modelled through auto-regression terms, and an endemic component, which is usually modelled via trends and seasonality of the time series. Other types of phenomenological models are based on flexible Gaussian process regressions (Johnson et al., 2018) which can be fitted to the data and then used to predict the future evolution of the epidemic. An ancestor of the phenomenological model is the *method of analogues* whose central aim is prediction and is founded on a comparison of the current observations with past epidemics (Viboud et al., 2003). These non-mechanistic methods are not covered in this thesis because, despite having shown their predictive power (Pell et al., 2018), they rarely provide clear estimates of the quantities of interests such as R_0 and $R_e(0)$ and of other epidemic parameters, nor do they attempt to quantify the number of underlying infections generating the data.

Regarding severity, the choice of modelling approach heavily depends on the type of data available. When time-to-event data are available, multi-state models (MSMs) can be adopted to describe the data in terms of event rates and event probabilities (Andersen and Keiding, 2002). Several sub-fields of MSMs are also considered in the context of the estimation of severity, including competing risks models (Andersen and Keiding, 2002) and models for dynamic prediction in survival analysis (Van Houwelingen and Putter, 2011). Alternatively, if count data on events at different levels of severity are available, counting-processes methods can be used (Chang and Hsiung, 2005).

1.3.2 Inferential approaches

As for the modelling choices, there is not a unique inferential approach adopted. In cases where only one data source is available and its likelihood provides straightforward and informative inference, a frequentist approach is adopted throughout the thesis.

More often the Bayesian paradigm is crucial to tackle intrinsic problems such as: the joint use of multiple datasets, availability of prior information from external sources, model uncertainty and the introduction of hidden stochastic quantities that guide the data.

In these cases Bayesian evidence synthesis (Ades et al., 2008; De Angelis and Presanis, 2018) helps in formulating a full model that exploits all available information. Moreover, Bayesian computation (Gelman et al., 2013) provides straightforward settings to account for the missing data, and hidden stochastic quantities of a comprehensive data-generating model.

1.4 Thesis structure

The following two chapters are focused on estimation of severity and transmission respectively. Chapter 2 illustrates several estimators for the severity of an epidemic. The chapter begins with a broad introduction to MSMs and their respective counting processes, bridging MSMs with SSMs. The background introduced in this section is repeatedly used throughout this thesis. Several estimators proposed in the literature are listed; they are broadly classified according to the data that they address, either time-to-event data or count data. Two extensions to the time-to-event estimators are proposed: one adopts parametric survival and the other applies

a class of model developed for demography, age-period-cohort models, to severity data. These extensions, together with an investigation of the applicability of the different methods to real data, are the innovation of this chapter. The chapter ends with some open challenges introduced by specific data collection schemes.

Chapter 3 covers epidemic models, the most common tool for the inference of transmission. Elements for specifying an epidemic model are introduced, together with criteria to classify epidemic models according to assumptions and approximations undertaken. Estimation methods for transmission and their application to the inference of influenza transmission are briefly reviewed. To illustrate, a case study of the analysis of severe cases arising from three waves of seasonal influenza in England is presented. The results reveal that epidemic models can be usefully applied to severe-case data consisting of small counts. This chapter concludes with a discussion on the inference of transmission: as shown by the illustrative analysis, a single data source is often affected by noise and selection bias, providing only partial information on transmission, and therefore more data sources, informing different aspects of severity and transmission, should be used jointly.

The challenges discussed in Chapters 2 and 3, call for more advanced inferential methods. The process of individuals becoming infected and moving across different levels of severity is in fact a complex process where hidden stochastic quantities govern the data behaviour. For this reason more advanced methods that allow for this hidden stochasticity in the inference, e.g. sequential Monte Carlo (SMC) methods, are introduced. Chapter 4 opens with a concise summary of the methods for state inference and parameter inference in state-space models. SMC methods are then applied to the estimation of severity to data affected by high censoring. The application of SMC methods to such complicated data is not straightforward: path degeneracy, one of the most important obstacles of SMC methods, here comes to a head, with full particle demise. To solve this problem an extension of the standard sequential importance sampling algorithm is proposed solving the particle demise problem. Moreover, an epidemic model with stochastic transmission and stochastic severity is specified in the SSM framework and a standard SMC method is proposed for approximating its likelihood.

Chapter 5 explores more in detail the layers of hidden stochastic states that can be used to model the severity process and the consequent inference of transmission and severity parameters. In this context, the need to jointly exploit multiple data sources is even more meaningful than when more simplistic epidemic models are used. A challenge to the inference with multiple data sources is disentangling dependencies between data. In fact, datasets often overlap, leading to repeated information; the magnitude of the overlap, however, is often unknown. A motivating example is represented by the severe-case data collected in England. Within this context, a simulation method is proposed to account for the dependence. A simulation study is carried out to show the difference between the estimates obtained with the proposed method and the estimates obtained assuming independence.

Lastly, given all the tools developed in the thesis, transmission and severity of influenza in England during the 2017/18 season are inferred from multiple sources in Chapter 6.

The final chapter conveys the main discoveries of this thesis and outlines possible future research and extensions.

Chapter 2

Estimation of severity

This chapter contains a brief review and some developments of the methods to estimate severity at the early stages of an epidemic. Section 2.1 reports the statistical models used to describe the data-generating process. The notation is set generally for the context of time to event processes and counting processes and is used throughout this thesis. Section 2.2 briefly reviews the most important severity-estimation methods found in the literature, two of which are analysed in more depth in Sections 2.3 and 2.5. Section 2.4 instead, extends the method exposed in 2.3, filling a current gap in the literature. Lastly, section 2.6 draws conclusions on the applicability of the described and proposed methods.

2.1 Counting processes and multi-state processes

The occurrence of an event can often be interpreted as a transition from one state to another: an individual dying is a transition from *life* to *death*, an individual falling ill with an infectious disease is a transition from *susceptible* to *infectious* and from *healthy* to *at risk* of severe events, etc. Compartmental models (Andersen and Keiding, 2002) (also known as compartment models in system kinetics (Matis, Wehrly, and Kiffe, 2005)) represent the evolution over time of a population that can be categorised into disjoint compartments; events are transitions from one compartment to another.

Data on these transitions might be available at different levels of aggregation. Counting-process models are used to analyse data on the number of people in a compartment or moving between compartments over time. On the other hand, survival analysis and multi-state models (MSMs) for event history analysis deal mainly with individual-level data consisting of times to events and types of events (Putter, Fiocco, and Geskus, 2007). According to the available data, the statistician decides which method to use and, by assuming a population or survival model, he decides on an approximation of the underlying phenomenon. Counting processes for population dynamics are described in Subsection 2.1.2, and Subsection 2.1.3 contains an introduction to multi-state processes for individual-level data.

Whether the model refers to a single individual changing states over his time at risk, or to the change over time in the composition of a population, the model is approximating the evolution of a system over time. Therefore almost all the models used in this thesis might be interpreted as state space models (SSMs) which are introduced in Section 2.1.1 below, and whose inference

will be extensively treated in Chapter 4.

2.1.1 A general framework: state-space models

SSMs are one of the most widely used models for dynamical systems (Commandeur, Koopman, and Ooms, 2011; Lindsten, 2013), i.e. systems that evolve over a specific domain (e.g. space, calendar time, survival time). This section focuses on the specific case of a system evolving over time, and often observed in the form of a time series, which is the context of our research question. Hence, the domain considered is discrete time, composed of equally spaced intervals indexed by $t = 0, 1, 2, \ldots$.

Many definitions of SSMs are given across the different fields in which these models are used (weather forecasts, signal processes, statistics, etc.). One of the most general is:

Definition 1. *A SSM is a stochastic process that makes use of a latent variable representation to describe dynamical phenomena (Schön and Lindsten, 2017).*

Definition 1 provides the essence of these models: a multivariate stochastic process that comprises a latent, unobserved, process $\{X_t\}_{t\geq 0}$ representing the underlying dynamics, and a process of observable components $\{Y_t\}_{t\geq 1}$.

The properties of this model change according to assumptions on the domain and the dependence of the stochastic processes. The unobserved process is often defined on a more dense domain than the observed process (constrained by data collection schemes); moreover the initial state of the latent process is often fully unobserved (i.e. X_0 exists but there is no observation Y_0). Regarding the dependence across variables of the SSM, the state process is often assumed Markovian over time. Then the SSMs considered can also be called partially observed Markov processs (POMPs) (King, Nguyen, and Ionides, 2016) or hidden Markov models (HMMs) (Churchill, 2005).

A parameter-driven Markovian SSM,can be defined through the state equation (Equation 2.1) and the observation equation (Equation 2.2)

$$X_t|(X_{t-1}, \boldsymbol{\Theta}) \sim p(x_t|x_{t-1}, \boldsymbol{\theta}) \tag{2.1}$$

$$Y_t|(X_t, \boldsymbol{\Theta}) \sim p(y_t|x_t, \boldsymbol{\theta}) \tag{2.2}$$

both characterised by a vector of parameters $\boldsymbol{\theta}$ (Birrell, De Angelis, and Presanis, 2018; Brockwell and Davis, 2016). Such a model is illustrated in Figure 2.1.

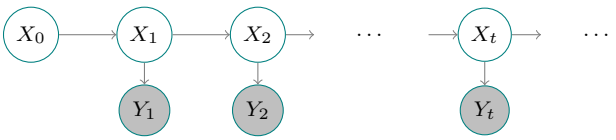


Figure 2.1: Graphical model for the SSM in (2.1), (2.2). Each r.v. is represented as a node, where grey nodes correspond to variables that are observed and white nodes are latent variables. The arrows express the dependence among the variables.

A simple example of a SSM is the Linear Gaussian SSM (Commandeur, Koopman, and Ooms, 2011; Durbin and Koopman, 2012), where the state process and the observational process are

generated by linear transformations (A and C) and Gaussian noises (V_t and E_t). Consider finite discrete time indexed by $t = 0, 1, 2 \ldots$, where 0 indexes the first (unobserved) interval, denote by σ and τ the variance of the unobserved and observed processes, respectively, and by α the mean of the state process at time 0; the model is defined by:

$$\begin{aligned} X_t &= AX_{t-1} + V_t \\ Y_t &= CX_t + E_t \end{aligned} \qquad \begin{aligned} V_t &\sim N(0, \sigma) \\ E_t &\sim N(0, \tau) \\ X_0 &\sim N(\alpha, \sigma) \end{aligned} \qquad \forall t = 1, 2, \ldots$$

Several phenomena can be modelled with an SSM. The MSMs introduced in Section 2.1.3 with specific distributions, might be interpreted as SSMs.

Within this thesis, SSM notation is used only when referring to counting processes and the data arising from their observation. In our examples the latent state X_t is often composed only of the counts of individuals in a specific compartment (prevalence counts) or moving from a compartment to another in a specific interval (incidence counts). Y_t is an often noisy and biased observation of the incidence or prevalence counts at time t.

2.1.2 General counting processes

In this section counting processes, a special case of stochastic processes, are briefly introduced.

The results reported below make intensive use of properties of r.v.s and stochastic processes. The aims of this section are: to introduce some elements of counting-process theory, to set a general notation and to report properties that will be used further on. The proofs of the following statements can be found in the major books on stochastic processes (e.g. Kingman (1992), Chiang (1980)).

A *stochastic process* $\{X(t); t \in [0, \infty)\}$ is a family of r.v.s describing an empirical process, whose development is governed by probability laws.

A *counting process* is a stochastic process where $X(t)$ takes values in the natural set $X(t) \in 0, 1, 2, \ldots$ (Chiang, 2007).

An *arrival process* is a sequence of increasing r.v.s $0 < S_1 < S_2 < \ldots$ which represent the times at which some repeating phenomenon occurs.

The S_n are called *arrival epochs*. We can also equivalently specify the arrival process by specifying either the *interarrival intervals*

$$X_n = S_n - S_{n-1} \qquad \forall n = 1, 2, \ldots$$

(and therefore $S_n = \sum_{i=1}^{n} X_i$), or the counting process

$$\{N(t); t > 0\}$$

that represents the *number of arrivals* in the interval $(0, t]$. This process is related to the epochs by:

$$\{S_n \leq t\} = \{N(t) \geq n\}$$

A particular case of an arrival process is the renewal process.

A *renewal process* is an arrival process for which the sequence of interarrival times $X_1, X_2, \ldots$ is a sequence of independent and identically distributed (iid) r.v..

A special counting process: the Poisson process

Definition 2. *A homogeneous Poisson process is a renewal process in which the interarrival intervals have an Exponential distribution function; i.e., for some rate $\lambda > 0$, each X_i has the density $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.*

The main properties of this process are:

P1 homogeneous Poisson processes have stationary increments;

P2 homogeneous Poisson processes have independent increments;

P3 the arrival epochs $S_1, S_2, \ldots$ for a homogeneous Poisson process are distributed according to an Erlang distribution

$$f_{S_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$$

P4 a r.v. $N(t)$, in the counting process $\{N(t); t > 0\}$, denoting the number of arrivals in $(0, t]$, is a Poisson r.v. with probability mass function

$$f_{N(t)}(n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

These properties follow from the definition of a homogeneous Poisson process. We report below two alternative definitions of the Poisson process.

Definition 3. *A homogeneous Poisson counting process $\{N(t); t > 0\}$ is a counting process that satisfies property **P4** (i.e., has the Poisson probability mass function) and has the independent and stationary increment properties.*

Following Definition 3, consider the number of arrivals in a very small interval $(t, t + \delta t]$. Given property **P1**, $N(t, t + \delta t)$ has the same distribution as $N(\delta t)$ and therefore:

$$\begin{aligned} P\{N(t, t+\delta t) = 0\} &= e^{-\lambda\delta} \approx 1 - \lambda\delta + o(\delta) \\ P\{N(t, t+\delta t) = 1\} &= \lambda\delta e^{-\lambda\delta} \approx \lambda\delta + o(\delta) \\ P\{N(t, t+\delta t) \geq 2\} &\approx o(\delta) \end{aligned} \tag{2.3}$$

From which Definition 4 follows:

Definition 4. *A homogeneous Poisson counting process is a counting process that satisfies Equation 2.3 (i.e. the probability of having more than one event in a small interval approaches 0 as δ approaches 0) and has the stationary and independent increment properties*

The Poisson process can be extended in a number of ways.

A *thinned Poisson process* can be constructed as follows. Let $\{N(t); t > 0\}$ be a Poisson counting process of rate λ and let $\{N_1(t); t > 0\}$ and $\{N_2(t); t > 0\}$ be two counting processes constructed as follows. Suppose that each arrival in $\{N(t); t > 0\}$ is sent to the first process $N_1(t)$ with probability p and to the second process $N_2(t)$ with probability $1-p$. We are therefore combining a Poisson(λ) process with a Bernoulli(p) process. The resulting processes $\{N_1(t); t > 0\}$ and $\{N_2(t); t > 0\}$ are also two Poisson processes with rates $p\lambda$ and $(1-p)\lambda$ respectively. The two Poisson processes can be proved to be independent.

A *non-homogeneous Poisson process* is a Poisson process whose rate λ is non constant over time. A non-homogeneous Poisson process with time varying arrival rate $\lambda(t)$ is defined as a counting process $\{N(t); t > 0\}$ that for all $t > 0$, $\lambda(t) > 0$ also satisfies:

$$P\{N(t, t+\delta t) = 0\} = e^{-\int_t^{t+\delta} \lambda(u)\mathrm{d}u} \approx 1 - \lambda(t)\delta + o(\delta)$$
$$P\{N(t, t+\delta t) = 1\} = \left(\int_t^{t+\delta} \lambda(u)\mathrm{d}u\right) e^{-\int_t^{t+\delta} \lambda(u)\mathrm{d}u} \approx \lambda(t)\delta + o(\delta)$$
$$P\{N(t, t+\delta t) \geq 2\} \approx o(\delta)$$

Which means that over intervals of a small length (where the variation of $\lambda(t)$ is small), the number of arrivals in an interval of length δ is still a Poisson r.v. with parameter $\lambda(t)\delta$, since it is assumed that $\int_t^{t+\delta} \lambda(u)\mathrm{d}u = \lambda(t)\delta$. This result is very useful for simulating Poisson processes. Let us note that for the non-homogeneous Poisson process, all the properties listed above hold, except for the stationarity of the increments. However, using the Mapping theorem reported in Kingman (1992) (page 18), it can be proven that a transformation of the domain of the process allows to rewrite a non-homogeneous Poisson process as a homogeneous Poisson process, with both stationary and independent increments.

Poisson processes are intensively used in *queue theory*, i.e. the study of waiting lines. Let $\{N(t); t > 0\}$ be a Poisson process with rate λ. Imagine that upon arrival, every individual is independently assigned an Exponential distribution with rate μ (i.e. the mean lifetime is $1/\mu$) after which he is served. This is known as an *immigration-death process* or, in queuing theory, the $M/M/\infty$ queue. Let us now consider $\{R(t), t > 0\}$ to be the prevalent cases, i.e. the people who have arrived and are waiting to be served. Given the arrival Poisson process, $R(t)$ increases at a rate λ and, given the Exponential distribution of the waiting times, $R(t)$ decreases at a rate $\nu(t) = r(t)\mu$, where $r(t)$ is the realization of the arrival process $R(t)$. The service process $S(t)$ inherits the Poisson properties of the arrival process, since, in a small interval of length δ, it can be seen as a sum (convolution) of thinned Poisson processes (those that describe the individuals that arrived at t and waited a time τ before service). It follows that $\{S(t); t > 0\}$ is a non-homogeneous Poisson process with rate $\nu(t) = r(t)\mu$.

2.1.3 Multi-state processes

A multi-state process is a stochastic process $\{X(t), t \in [0, \tau]\}$ with a finite state space $\mathcal{S} = 1, \ldots, p$, used for event history analysis (Andersen and Keiding, 2002).

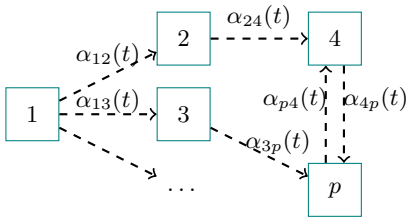
An individual multi-state process is defined by the initial distribution of the states $\pi_h(0) = P(X(0) = h)$, with $h \in \mathcal{S}$ indexing the possible states, and by the transition probabilities, i.e. the probability of moving from state h to state j ($h, j \in \mathcal{S}$) during the interval from s to t ($s, t \in [0, \tau]$ and $s \leq t$):

$$q_{hj}(s, t) = P(X(t) = j | X(s) = h, \mathcal{X}_{s^-})$$

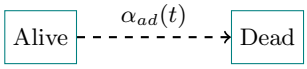
where $\mathcal{X}_{s^-}$ denotes the history of the process $X(\cdot)$. We assume the existence of transition intensities that can be obtained by derivatives:

$$\alpha_{hj}(t) = \lim_{\Delta t \to 0} \frac{q_{hj}(t, t+\Delta t)}{\Delta t}$$

Multi-state models represent these processes and can be visualized through boxes, each one identifying one state in $\mathcal{S}$, and arrows representing the possible events, as shown in Figure 2.2.



(a) Graphical representation of a MSM with p states. Every arrow represents a possible transition and is labelled with its intensity.



(b) Graphical representation of a survival (two-state) model: individuals can only make one transition $a \to d$. The intensity $\alpha_{ad}(t)$ is called the *hazard function* in the survival literature.

Figure 2.2: Graphical representation of (a) a generic MSM and (b) of a two-state model for survival data.

A simple example: survival model

It is useful to introduce one of the most employed models in this class: the survival model represented in Figure 2.2 (b). It has two states (*Alive* and *Dead*) and the only possible transition is from the former to the latter.

It is useful to name the specific functions that are used to characterise this model and that are just transformations of the general transition probabilities and intensities defined for the MSM. Let T be the r.v. that defines the time elapsed from a defined origin 0 (e.g. birth, study-entry) to death with distribution function $F(t)$. The r.v. T is often characterised by the *survival function* $S(t) = P(T > t) = 1 - F(t)$.

In the MSM notation the transition can be defined as:

$$q_{ad}(s,t) = P(T < t | T \geq s)$$

and the transition intensity, named in the survival literature the *hazard rate* is defined by:

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{P(T \leq t + \Delta t | T \geq t)}{\Delta t}$$

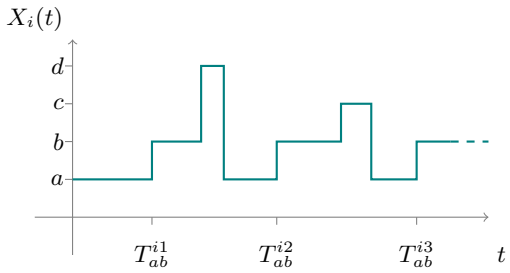
which is, by definition of the survival function, equal to $-\frac{\mathrm{d} \log S(t)}{\mathrm{d}t}$ and leads to the parametrization of the survival function according to the hazard rate:

$$S(t) = \exp\left(-\int_0^t \alpha(u)\mathrm{d}u\right)$$

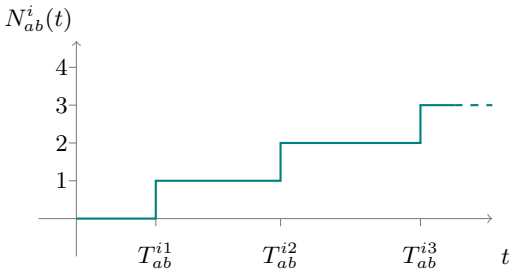
2.1.4 Likelihood of multi-state processes

We can build a counting process on this multi-state process. Assume $X_i(t)$ is a multi-state process such as the ones just described and it is observed over the interval $[0, \tau_i]$, with $i = 1, \ldots, n$ labelling the individual to whom the process refers. Denote by $N_{hj}^i(t)$ with $h \neq j$ the number of direct transitions $h \to j$ during the interval time $[0, t]$ for individual i, and by T_{hj}^{ik} the respective ordered transition times, where $k = 1, \ldots, N_{hj}^i(\tau_i)$ counts the ordered transitions $h \to j$ for individual i.

A graphical representation of an instance of these r.v.s for a univariate MSM with space $\mathcal{S} = a, b, c, d$ is drawn in Figure 2.3; here all the transitions among states are possible.



(a) Time spent in the states by individual i and transitions.



(b) Count of the transitions $a \to b$ for individual i (y axis) and associated transitions times (x axis).

Figure 2.3: A realization of (a) the multi-state process of individual i and (b) its associated counting and transition-time processes.

Denote with $Y_h(t)$ the number of individuals in state h at time $t-$:

$$Y_h(t) = \sum_{i=1}^{n} Y_h^i(t)$$

where $Y_h^i(t) = \mathbb{I}_{\{X_i(t-)=h\}}$ denotes whether individual i is at risk of moving out from state h.

Ideally, a complete dataset would consist of: the initial states $X_i(0)$, the transition times T_{hj}^{ik}, the individuals at risk $Y_h^i(t)$ (and the number of direct transitions $N_{hj}^i(t)$ that can be derived from T_{hj}^{ik} and $Y_h^i(t)$) for events $h \neq j$, for individuals $i = 1, \dots n$ and for transition count k. The static parameters of the model consist of: the intensities $\alpha_{hj}^i(t)$ and the initial probabilities $\pi_h^i(0)$, for events $h \neq j$, for individuals $i = 1, \dots n$.

The likelihood of the full data conditionally on the model parameters is defined in Equation 2.4:

$$\mathcal{L}(\mathbf{Y}^{1:n}, \mathbf{T}^{1:n}, \mathbf{N}^{1:n} | \boldsymbol{\alpha}^{1:n}, \boldsymbol{\pi}^{1:n}) = \prod_{i=1}^{n} \pi_{X_i(t)}^i(0) \prod_{h \neq j} \prod_{k=1}^{N_{hj}^i(\tau_i)} \alpha_{hj}^i(T_{hj}^{ik}) e^{-\int_0^{\tau_i} \alpha_{hj}^i(t) Y_h^i(t) dt} \tag{2.4}$$

where for $i = 1, \dots, n$, the following sets can be defined: $\mathbf{Y}^i = \left\{ Y_h^i; h \in \mathcal{S} \right\}$, $\mathbf{T}^i = \left\{ T_{h,j}^{i1}, \dots T_{h,j}^{iN_{h,j}^i}; h, j \in \mathcal{S}, h \neq j \right\}$, $\mathbf{N}^i = \left\{ N_{h,j}^i; h, j \in \mathcal{S}, h \neq j \right\}$, $\alpha^i = \left\{ \alpha_{h,j}^i; h, j \in \mathcal{S}, h \neq j \right\}$ and $\pi^i = \left\{ \pi_h^i; h \in \mathcal{S} \right\}$.

However this model is never adopted in practice, both due to the unavailability of the full data and to the high number of parameters. More likely some assumptions are made. For example, the intensities are often assumed to be not individual-specific: $\alpha_{hj}^i(t) = \alpha_{hj}(t), \forall i = 1, \dots n$; or constant over time, i.e. $\alpha_{hj}(t) = \alpha_{hj}$. Moreover the model is often assumed to be Markovian, i.e. $\alpha_{hj}(t)$ depends only on the current state $X(t)$. These assumptions allow a significant simplification of the likelihood 2.4 and will be often adopted when using MSMs in this thesis.

2.2 Literature review

Accurate estimates of the risk of severe outcomes are necessary for policy planning in epidemic responses (Van Kerkhove et al., 2010). Measures of the severity of an epidemic have been

introduced in Chapter 1; the proportion of fatal cases given a specified condition has been mentioned as the quantity of interest that goes under the name of Case Fatality Rate, Case Fatality Ratio or Case Fatality Risk (Nishiura et al., 2009; Porta, 2008). It is usually expressed as a percentage and, albeit its name might refer to a person-time measure (rate/ratio), it expresses a probability; therefore the notation of Wong et al. (2013a) is adopted in this thesis, naming this quantity the case fatality risk (CFR).

2.2.1 Methods for severity

Multiple indexes of severity exist, according to the aspect that is to be described. This chapter focuses on the hospitalised case fatality risk (hCFR), that is the probability of death given hospital admission. However, this literature review covers also other measures of severity, such as the infection fatality risk (IFR) or the symptomatic case fatality risk (sCFR), which are the fatality within infected and symptomatic cases respectively, since they have been often estimated together with the hCFR while describing the entire severity process.

The IFR is one of the most interesting severity measures because it represents the actual mortality in the whole population of infected individuals. The IFR is rarely computed via direct estimation of the number of infections and number of deaths in the population. The UK Department of Health, together with the Health Protection Agency (HPA) (now Public Health England (PHE)) achieved the goal of obtaining estimates of both the numerator and the denominator during the 2009 influenza A/H1N1 pandemic (Donaldson et al., 2009). This was possible due to the combination of data from several surveillance schemes rapidly activated during the pandemic. Other studies have estimated the IFR during the 2009 H1N1 pandemic. Presanis et al. (2009) used Bayesian evidence synthesis to obtain estimates of the IFR from data of two cities of the USA. They adopted a pyramid approach, estimating the IFR by combining estimates of the probabilities of reaching different stages of severity (e.g. probability of having symptoms given infection, probability of hospitalization given symptoms, ...). Moreover their model accounted for specific testing and reporting probabilities for the different stages of severity. During the following years the group applied the same methodology to estimate the severity in the UK (Presanis et al., 2011; Presanis et al., 2014). Similar analyses were performed in New Zealand (Baker et al., 2009), in Finland (Shubin et al., 2014) and in the whole southern hemisphere (Baker, Kelly, and Wilson, 2009) where, thanks to a wide range of surveillance schemes already active from 2008, the infected population and the sCFR could be estimated. This was achieved by combining sentinel data on symptomatic cases from General Practitioner (GP) consultations with population data on GP consultations and with experimental studies.

Another approach to estimate the IFR is to infer the excess mortality due to an infection. A simple model is applied by Murray et al. (2006) to analyse the influenza pandemic of 1918-1920. The authors calculated the average mortality rate in 1915-17 and 1921-23, and subtracted this average from mortality in 1918-20. This simple computation gave an estimate of the mortality due to influenza under the assumption that all excess deaths are truly associated with influenza. A more complex model is formulated in Wong et al. (2013b) to estimate the IFR of the 2009 influenza A/H1N1 pandemic Virus in Hong Kong. They derived a proxy of the total influenza activity (in terms of weekly incidence rates of pH1N1 infections) from weekly influenza-like illness (ILI) data and weekly proportions of specimens that tested positive for influenza. Statistical

models (namely, linear regression, time-series regression and Poisson regression) are used to model mortality from 2003 to 2009. Mortality is regressed on the proxy of 2009 influenza A/H1N1 pandemic activity and other covariates including other seasonal influenza proxies and weather variables such as temperature and humidity.

Serfling models (Serfling, 1963) are a sub-family of excess mortality methods that describe the excess of deaths (in counts) by comparing time series and including trigonometric functions to model seasonality. However all these excess mortality models rely on the assumption that the excess mortality is due to the virus we are analysing. Therefore, in the case of a mildly severe epidemic, or when other causes can increase mortality (other diseases, wars, ...) the estimates of the IFR are biased.

This consideration, together with the fact that data on both the numerator and the denominator of the IFR are rarely available, motivates the focus on the hCFR, hereby denoted only by CFR, in the following analysis of severity.

2.2.2 Methods for estimating the Case Fatality Risk

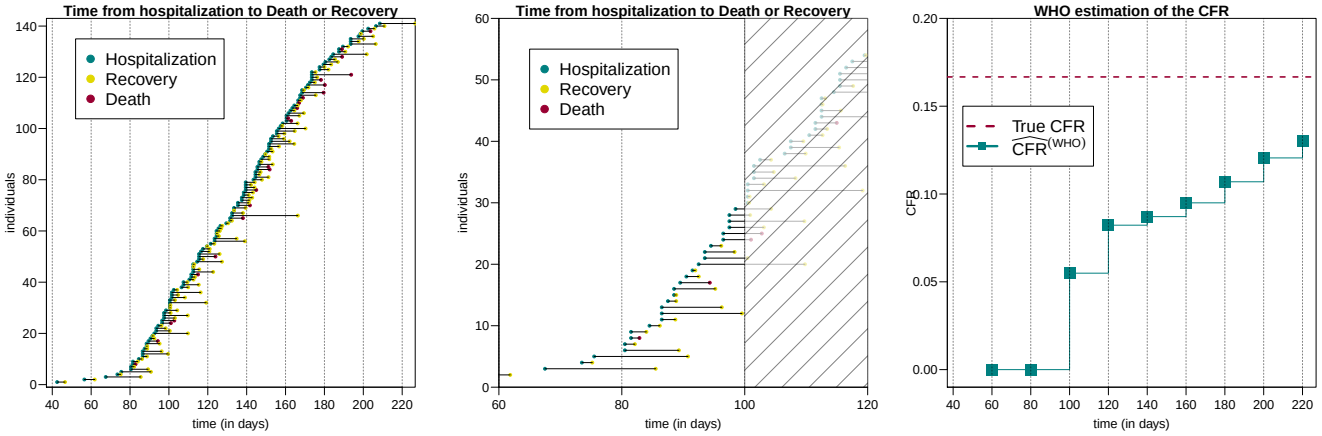
The World Health Organization (WHO) proposed the following estimator of the CFR in the case of an epidemic (WHO, 2015):

$$\widehat{\text{CFR}}^{(\text{WHO})}(s) = \frac{\text{cumulative number of deaths}(s)}{\text{cumulative number of (hospitalised) cases}(s)} \tag{2.5}$$

with $s \in (0, S)$ being the time of analysis. Estimator 2.5 assumes constant CFR and it is well known to be biased until the end of the epidemic, here denoted by S (Lipsitch et al., 2015). This bias is due to right censoring that happens when the analysis is carried out at time $s < S$, when some patients at risk have not experienced any event such as death or recovery yet. To understand the effect of right censoring on the estimator 2.5 a simulated dataset has been plotted in Figure 2.4. This dataset contains the time from hospitalization to death and recovery generated using a parametric survival model for death and recovery from simulated hospitalization counts mirroring the counts of cases during 2012/13 epidemic. When analysed early, for example on the $100th$ day from the beginning of the epidemic, the number of hospitalizations is increasing according to the epidemic dynamics, and many individuals have not experienced the final event yet.

Several papers have addressed this problem and most of them have used survival analysis approaches, both parametric and non-parametric.

The problem of estimation of the CFR from survival data has been addressed under two perspectives. The first one assumes that the data-generating process is a mixture model for survival data (Farewell, 1982): the individuals belong to the group of people that die with probability CFR and to the group of people that survive with probability 1-CFR. Their time-to-event is then defined conditionally on the group to which the individuals belong. This approach has been adopted by Donnelly et al. (2003) within a parametric-survival context. The other approach takes a prospective perspective and assumes that the data-generating parameter is a competing risk process (which is a special case of MSMs introduced in Section 2.3.1). Ghani et al. (2005) and Jewell et al. (2007) proposed an estimator for the CFR in this context. Garske et al. (2009) briefly reviewed the underestimation error of the CFR estimates and the solutions proposed by Donnelly et al. (2003) Ghani et al. (2005) and Jewell et al. (2007).



(a) Simulated survival data during an epidemic.

(b) Same data observed until day 100.

(c) Estimated CFR at several time points during the epidemic.

Figure 2.4: A simulated dataset with daily flu-related hospitalizations, deaths and recovery. In Panel (a) and (b) the x axis is the calendar time from the beginning of the epidemic and the y axis is the ordered (by hospital admission date) number of individuals. Panel (b) is the bottom left corner of Panel (a), when data are only observed until day 100. Panel (c) reports in green the estimator 2.5 calculated at times $s = 60, 80, \ldots, 220$ from the beginning of the epidemic and in red the true value used to generate the data.

Estimators of the CFR starting from count data have been proposed by Yip et al. (2005b) and Yip et al. (2005a), using a counting process approach and relaxing the assumption of constant CFR on which survival data estimators are based. In the same context, Lam et al. (2008) proposed a test for constant CFR in the case of an emerging epidemic.

The papers listed above differ mainly in the data they analyse (individual time-to event data vs population count data) but also in their assumptions on the CFR (constant versus time-varying). However, a point in common is that they all attempt to overcome the problem of the biased estimators due to right censoring.

2.3 Analysis of survival data

Donnelly et al. (2003), Ghani et al. (2005) and Jewell et al. (2007) all address the estimation of the CFR from time-to-event data, which are often called "survival data". An initial introduction to these data and the models to describe them, the MSMs, has been given in Section 2.1.3. To correct for the bias that affects Estimator 2.5, MSMs are applied under the hypothesis of competing risks.

MSMs with competing risks are introduced in Subsection 2.3.1 below; they are then applied to the estimation of the CFR in Subsection 2.3.2, essentially recalling the same estimator proposed in Ghani et al., 2005. Lastly in Subsection 2.3.3 these methods are extended to the parametric-survival setting, to increase the precision of the estimator.

2.3.1 Analysis of survival times with competing events

Competing-risks models are a subset of MSMs, developed in the context of the investigation of cause-specific mortality and therefore this example is used to illustrate the model. However,

many other situations can be described by competing-risks models, including death within hospitalised flu cases, as shown below. The process $\{X(t), t \in [0, \tau]\}$ with space $\mathcal{S} = 0, \ldots, k$ has one transient state (0: alive) and k absorbing states, each state $h = 1, \ldots, k$ corresponding to death from cause h (Andersen and Keiding, 2002), and it is represented in Figure 2.5. If death for cause h happens, none of the other events is possible.

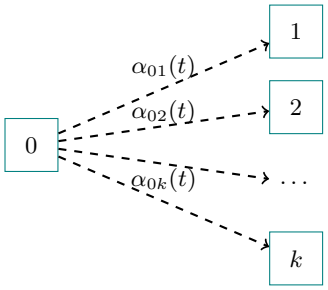


Figure 2.5: Graphical representation of a competing-risks MSM representing death for multiple (k) causes. Every arrow represents a possible transition and it is labelled with its intensity.

Since all the transitions span from state 0 this index is dropped and the transition probabilities and intensities are simply indexed by the cause of death h:

$$q_h(s,t) = P(X(t) = h | X(s) = 0, \mathcal{X}_{s^-})$$

$$\alpha_h(t) = \lim_{\Delta t \to 0} \frac{q_h(t, t+\Delta t)}{\Delta t}$$

As shown in Subsection 2.1.3, the process can be described by the time to event (also called "survival time" or "time to failure") T and by the cause of failure D. The intensity function $\alpha_h(t)$ takes the name of *cause-specific hazard function* in this context and can be interpreted as the hazard of failing from a given cause in the presence of competing events (Putter, Fiocco, and Geskus, 2007), i.e.:

$$\alpha_h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t, D = h | T > t)}{\Delta t}$$

The competing-risks model has been considered in the past as a multivariate failure time model, where individuals were assumed to have potential survival times for each cause of death but only the shortest was observed. This modelling approach was discarded, early on, both for the unrealistic interpretation and for identifiability issues (Prentice et al., 1978). The approach used here instead exploits only the cause-specific hazard without defining a cause-specific survival.

Denote by $A(t)$ the cumulative cause-specific hazard defined by

$$A_h(t) = \int_0^t \alpha_h(s) \mathrm{d}s$$

for $h = 1, \ldots, k$. The *overall survival*, i.e. the probability of not having failed for any of the k causes of death, is defined as follows

$$S(t) = e^{-\sum_{h=1}^{k} A_h(t)} \tag{2.6}$$

The *cumulative intensity function* of cause h, I_h is the probability of failing from cause h at or before time t and it depends on all the other causes of death through the survival function,

hence:

$$P(T \leq t; D = h) = I_h(t) = \int_0^t \alpha_h(s)S(s)\mathrm{d}s$$

The cumulative intensity function is not a cumulative probability function as it does not converge to 1. Its limit for the time going to infinity is however very useful since it denotes the overall probability of death for cause h, i.e.:

$$\lim_{t\to\infty} I_h(t) = P(D = h) \tag{2.7}$$

Assuming that the only possible causes of death are the k causes considered,
$\sum_{h=1}^{k} \lim_{t\to\infty} I_h(t) = 1$.

Methods for estimating cause-specific hazards, overall survival functions and cumulative intensity functions can be found in Putter, Fiocco, and Geskus (2007) and are derived simply by using the Kaplan-Meier (KM) estimator (Kaplan and Meier, 1958).

2.3.2 Estimation of the Case Fatality Risk using survival analysis with competing events

Ghani et al., 2005 exploit competing risks models to formulate an estimator for the CFR. Set $k = 2$ and define a competing-risks process for the time from hospitalization to death or recovery $(h = D, R)$, as illustrated in Figure 2.6.

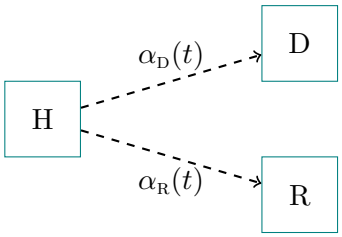


Figure 2.6: Competing-risks model for death and recovery: H, D and R represent respectively, the hospitalized, death and recovered states. $\alpha_D(t)$ is the hazard of death at time t since hospital admission and $\alpha_R(t)$ is the hazard of recovery at time t since hospital admission.

Let calendar time be indexed by s ($s \in [0, S]$, with S the time of the end of the epidemic) and let the time since hospital admission (i.e. the time-to-event) be indexed by $t \in [0, T]$. Assume that the CFR is constant over calendar time.

Let:

$\alpha_D(t) = \lim_{\Delta t\to 0} \frac{P(t<T\leq t+\Delta t, h=D|T\geq t)}{\Delta t}$ be the cause-specific hazard function of death;

$\alpha_R(t) = \lim_{\Delta t\to 0} \frac{P(t<T\leq t+\Delta t, h=R|T\geq t)}{\Delta t}$ be the cause-specific hazard function of recovery;

$t_{max}(s)$ be the maximum observed time from hospital admission to death or recovery that has occurred by time s;

$S(t)$ be the overall survival function, i.e. the survival if both endpoints are treated as a single composite endpoint;

$I_D(t) = \int_0^t S(u)\alpha_D(u)du$ be the cumulative intensity function of death, i.e. the probability of death at or before time t.

$I_{\mathrm{R}}(t) = \int_0^t S(u)\alpha_{\mathrm{R}}(u)du$ be the cumulative intensity function of recovery, i.e. the probability of recovery at or before time t.

Then the overall probability of death before or at calendar time s can be estimated by the cumulative intensity function computed at the maximum observed time:

$$\theta_{\mathrm{D}}(s) = I_{\mathrm{D}}(t_{max}(s)) = \int_0^{t_{max}(s)} S(t)\alpha_{\mathrm{D}}(t)dt \tag{2.8}$$

and the probability of recovery before or at calendar time s can be estimated by:

$$\theta_R(s) = I_{\mathrm{R}}(t_{max}(s)) = \int_0^{t_{max}(s)} S(t)\alpha_{\mathrm{R}}(t)dt \tag{2.9}$$

Note that for $t \to +\infty$, $I_{\mathrm{D}}(t)$ represents the overall probability of death, following from Equation 2.7

$$\mathrm{CFR} = \lim_{t \to +\infty} I_{\mathrm{D}}(t)$$

and therefore if data are available for a $t_{max}(s)$ large enough, $\theta_{\mathrm{D}}(s)$ approximates the overall probability of death. This event happens certainly if the epidemic is complete ($s = S$), because there cannot be any $t > t_{max}(S)$ and everyone has either died or recovered. In this setting, the only alternative events are death and recovery; at and after the largest time-to-event S no individual is at risk, everyone had an event and, since death and recovery form a partition,

$$\theta_D(S) + \theta_R(S) = 1$$

from which follows that

$$\mathrm{CFR} = \theta_{\mathrm{D}}(S)$$

Therefore to derive an estimate of the CFR, estimates of the cumulative intensity functions for both events must be obtained from the data on the whole epidemic (until its end S) and evaluated at the maximum observed survival time $t_{max}(S)$:

$$\widehat{\mathrm{CFR}} = \widehat{\theta}_{\mathrm{D}}(S)$$

To illustrate the use of the cumulative intensity function to estimate the CFR, estimates of the cumulative intensity functions obtained from the analysis of the whole survival datasets of Figure 2.4 are plotted below.

Approximating the CFR before the end of the epidemic

During the epidemic $\theta_{\mathrm{D}}(s) + \theta_{\mathrm{R}}(s) < 1$ because individuals can have recovered, died , but also be in the hospital without having had any event. Before the end of the epidemic, when $s \leq S$, $t_{max}(s) \leq t_{max}(S)$, from which follows that the probability of death at or before time s is smaller or equal than the probability of death at or before S, which is the CFR. The same reasoning can be made for its inverse, the probability of recovery, so that:

$$\theta_{\mathrm{D}}(s) \leq \theta_{\mathrm{D}}(S) \leq 1 - \theta_{\mathrm{R}}(s)$$

This inequality can be observed in Figure 2.7, where the empirical cumulative intensity function of death at S, drawn in yellow, is shown to lie always between the empirical cumulative intensity function of death at s and 1 minus the empirical cumulative intensity function of recovery at s.

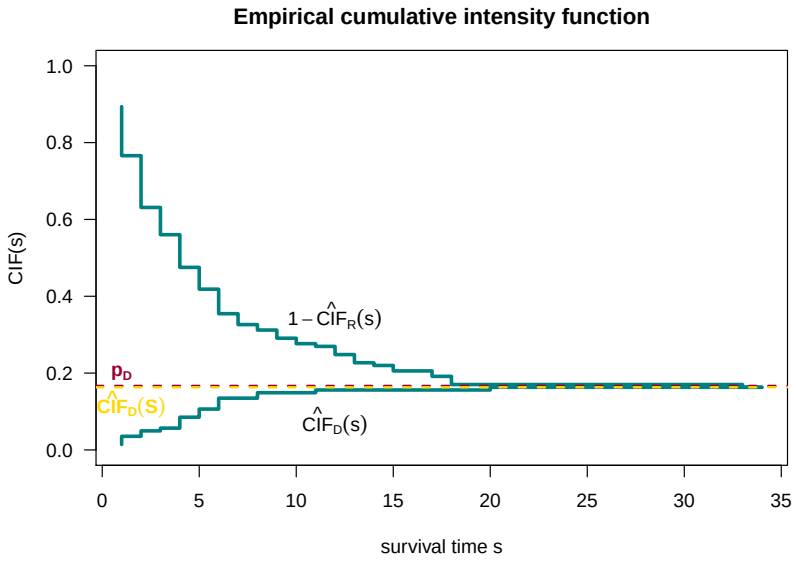


Figure 2.7: Results from the competing-risks analysis of the time to death and recovery. The x axis is the time to event, the green curves are the empirical cumulative intensity function of death and 1 minus the cumulative intensity function of recovery. The limit of the cumulative intensity function of death for $s \to \infty$ is approximated by the cumulative intensity function of death at the highest survival time S, which is drawn in yellow. The value of CFR used to generate the data is drawn in red and denoted by p_{D}.

Under the assumption that patients who remain in the hospital between s, the observation time, and S, the end of the epidemic, experience a CFR equal to those who had an event up to time s:

$$\widehat{\theta_{\mathrm{D}}}(S) = \frac{\theta_{\mathrm{D}}(s)}{\theta_{\mathrm{D}}(s) + \theta_{\mathrm{R}}(s)} \tag{2.10}$$

Estimates of $\theta_{\mathrm{D}}(s)$ and $\theta_{\mathrm{R}}(s)$ can be obtained using parametric survival to estimate the cumulative intensity functions. They can then be plugged into Equation 2.10 which allows the estimation of $\theta_{\mathrm{D}}(S)$, the probability of dying over all the epidemic (i.e. during $[0, S]$) which approximates the CFR.

Estimating the CFR

The steps to obtain estimates of the cumulative intensity function using standard KM methods are summarised below. Consider the discrete time from hospitalization indexed by $j = 1, 2, \ldots J$, for example days. Given an analysis time s, denote by:

$d_{\mathrm{D}j}(s)$: number of deaths on day j from admission to hospital;

$d_{\mathrm{R}j}(s)$: number of recoveries on day j from admission to hospital;

$n_j(s)$: number remaining at risk j days after admission to hospital;

$J(s)$ the maximum observed number of days from hospital admission to death or recovery that has occurred by time s (i.e. $J(s)$ is a discrete version of $t_{max}(s)$).

The overall survival probability is computed with the KM formula considering both endpoints.

$$\widehat{S}_j(s) = \prod_{r=1}^{j} \left(1 - \frac{d_{\mathrm{D}}(s) + d_{\mathrm{R}}(s)}{n_r(s)}\right)$$

The discretised version of the hazard of dying is:

$$\alpha_{\mathrm{D}j} = P(J = j, i = D | J > j - 1)$$

where i indexes the event and at analysis time s, it can be estimated by:

$$\widehat{\alpha}_{\mathrm{D}j}(s) = \frac{d_{\mathrm{D}j}(s)}{n_j(s)}$$

Then the overall probability of death before or at calendar time s can be approximated by the cumulative intensity function for death computed at the maximum observed time $t = J(s)$:

$$\widehat{\theta_{\mathrm{D}}}(s) = \sum_{j=1}^{J(s)} \widehat{S}_{j-1}(s)\widehat{\alpha}_{\mathrm{D}j}(s) \tag{2.11}$$

Similarly, the overall probability of recovery before or at calendar time s is

$$\widehat{\theta_{\mathrm{R}}}(s) = \sum_{j=1}^{J(s)} \widehat{S}_{j-1}(s)\widehat{\alpha}_{\mathrm{R}j}(s) \tag{2.12}$$

An estimator for the CFR at an early stage of the epidemic can be obtained by plugging 2.11 and 2.12 in 2.10:

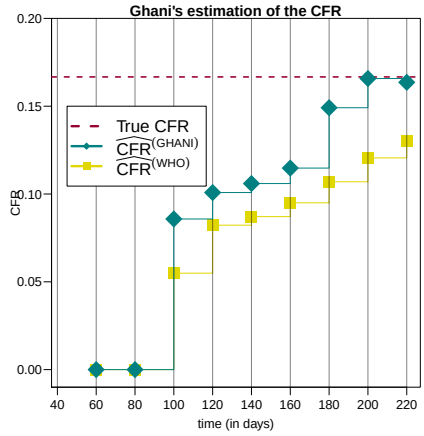
$$\widehat{\mathrm{CFR}}^{(\mathrm{GHANI})} = \widehat{\theta}_D(S) = \frac{\widehat{\theta}_D(s)}{\widehat{\theta}_D(s) + \widehat{\theta}_R(s)} \tag{2.13}$$

This estimator was computed at several time points during the simulated epidemic reported in Figure 2.4. On this dataset alone the correction of Ghani's estimator is not particularly evident (Figure 2.8 (a)). However, if a set of simulations is carried out and the median behaviour of the estimator is analysed, the estimator based on competing-risks converges to the true parameter much earlier than WHO's estimator (Figure 2.8 (b)).

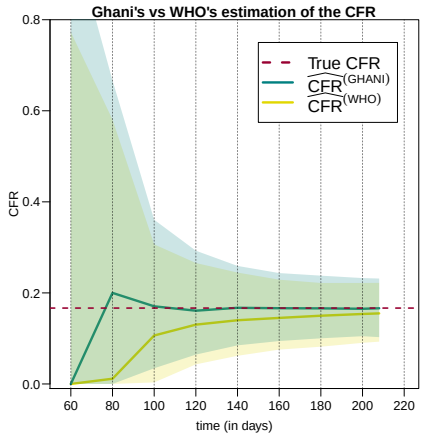
The estimator proposed by Ghani et al. (2005) is very appealing, as it takes a fully non-parametric approach and it allows correction for right censoring. However, this methods calls for two improvements. The first edit follows from the fact that, instead of computing the cumulative intensity function of death at $+\infty$, $I_{\mathrm{D}}(\infty)$ is approximated by $I_{\mathrm{D}}(t_{max})$ and therefore hazard and survival functions are only defined on the observed survival times. This is due to the fact that a KM estimator is used for the survival function. It seems natural therefore to extend this estimation method to a parametric setting where, in some cases, $\lim_{t\to+\infty} I_{\mathrm{D}}(t)$ can be solved mathematically, given the estimated cause-specific hazards and the assumed survival distribution. Moreover, the main disadvantage of this estimator is the assumption of constant CFR over calendar time s which is not realistic. To relax this hypothesis a time-varying version of the estimator is also proposed below.

2.3.3 Extension to parametric survival analysis

The parametric version of the CFR estimator proposed is here derived in the most simple case, when the cause-specific hazards are constant over time, leading to the overall survival time being Exponentially-distributed.



(a) $\widehat{\text{CFR}}^{(\text{GHANI})}$ and $\widehat{\text{CFR}}^{(\text{WHO})}$ computed over time.



(b) Median (line) and 95% simulation CI (shades) of $\widehat{\text{CFR}}^{(\text{GHANI})}$ and $\widehat{\text{CFR}}^{(\text{WHO})}$.

Figure 2.8: In Panel (a) and (b) the x axis is the calendar time from the beginning of the epidemic and the y axis is the CFR. Panel (a) reports in green the estimator 2.13 and in yellow the estimator 2.5 calculated at times $s = 60, 80, \dots 220$ from the beginning of the epidemic and in red the true value used to generate the data. Panel (b) reports median over 500 simulated datasets, obtained with the simulation scheme of Figure 2.4 and consisting of ≈ 150 cases, of the same estimators. The shaded area represent the limits where 95% of the 500 estimates lie.

Denote with $\alpha(t)$ the hazard function to any event, which can be shown from 2.6 to be equal to the sum of the cause-specific hazard functions:

$$S(t) = e^{-\sum_{h=1}^{k} A_h(t)}$$

$$A(t) = \sum_{h=1}^{k} \int_0^t \alpha_h(u)\mathrm{d}u$$

$$\alpha(t) = \alpha_{\text{D}}(t) + \alpha_{\text{R}}(t)$$

In this context data are composed again of time to event and type of event for individual i: (t_i, d_i).

If the overall survival time follows an Exponential r.v., than the overall hazard is assumed constant over time. The cause-specific hazards are also assumed to be constant over time such that

$$\alpha(t) = \alpha = \alpha_{\text{D}} + \alpha_{\text{R}}$$

$$T \sim \exp(\alpha)$$

The probability density function of the time to any event is:

$$f(T = t_i) = \alpha e^{-\alpha t_i}$$

Given the time to event t_i, D is a Bernoulli r.v. (Beyersmann et al., 2009), so the conditional probability of an event is:

$$g(D = d_i | T = t_i) = \left(\frac{\alpha_{\text{D}}}{\alpha_{\text{D}} + \alpha_{\text{R}}}\right)^{d_i} \left(\frac{\alpha_{\text{R}}}{\alpha_{\text{D}} + \alpha_{\text{R}}}\right)^{1-d_i}$$

that does not depend on t_i thanks to the assumption of constant hazard.

Every observation (t_i, d_i) determines the following element of the likelihood as follows:

$$\begin{aligned}\mathcal{L}(t_i, d_i|\alpha_{\text{D}}, \alpha_{\text{R}}) &= P(D = d_i, T = t_i) \\ &= g(D = d_i|T = t_i) \cdot f(T = t_i) \\ &= \left(\frac{\alpha_{\text{D}}}{\alpha_{\text{D}} + \alpha_{\text{R}}}\right)^{d_i} \left(\frac{\alpha_{\text{R}}}{\alpha_{\text{D}} + \alpha_{\text{R}}}\right)^{1-d_i} \cdot (\alpha_{\text{D}} + \alpha_{\text{R}}) e^{-(\alpha_{\text{D}}+\alpha_{\text{R}})t_i}\end{aligned}$$

The likelihood of the whole sample $(\boldsymbol{t}, \boldsymbol{d})$ of size n is:

$$\mathcal{L}(\boldsymbol{t}, \boldsymbol{d}|\alpha_{\text{D}}, \alpha_{\text{R}}) = \prod_{i=1}^{n} \left(\frac{\alpha_{\text{D}}}{\alpha_{\text{D}} + \alpha_{\text{R}}}\right)^{d_i} \left(\frac{\alpha_{\text{R}}}{\alpha_{\text{D}} + \alpha_{\text{R}}}\right)^{1-d_i} \cdot (\alpha_{\text{D}} + \alpha_{\text{R}}) e^{-(\alpha_{\text{D}}+\alpha_{\text{R}})t_i}$$

with corresponding log likelihood

$$\begin{aligned}\ell(\boldsymbol{t}, \boldsymbol{d}|\alpha_{\text{D}}, \alpha_{\text{R}}) = \log\left(\frac{\alpha_{\text{D}}}{\alpha_{\text{D}} + \alpha_{\text{R}}}\right) \sum_{i=1}^{n} d_i + \\ \log\left(\frac{\alpha_{\text{R}}}{\alpha_{\text{D}} + \alpha_{\text{R}}}\right) \sum_{i=1}^{n} (1 - d_i) + n \log(\alpha_{\text{D}} + \alpha_{\text{R}}) - (\alpha_{\text{D}} + \alpha_{\text{R}}) \sum_{i=1}^{n} t_i\end{aligned}$$

Via equating the Score matrix to 0, maximum likelihood (ML) estimates $\widehat{\alpha}_{\text{D}}$ and $\widehat{\alpha}_{\text{R}}$ can be derived. Given the distributional assumptions on the model, the survival function is:

$$\widehat{S}(t) = P(T > t) = e^{-\widehat{\alpha}t}$$

where $\widehat{\alpha}$ is the estimator of the hazard of any event, obtained as the sum of the two estimators : $\widehat{\alpha} = \widehat{\alpha}_{\text{D}} + \widehat{\alpha}_{\text{R}}$ due to invariance property of ML estimators. Similarly, the cumulative intensity functions can be estimated, as follows:

$$\begin{aligned}I_{\text{D}}(t) &= \int_0^t \widehat{S}(u)\widehat{\alpha}_{\text{D}} du \\ &= \widehat{\alpha}_{\text{D}} \int_0^t e^{-\widehat{\alpha}u} du \\ &= \frac{\widehat{\alpha}_{\text{D}}}{\widehat{\alpha}} - \frac{\widehat{\alpha}_{\text{D}} e^{-\widehat{\alpha}t}}{\widehat{\alpha}}\end{aligned} \tag{2.14}$$

The CFR can be estimated as in the previous section by

$$\widehat{\theta}_{\text{D}} = \lim_{t\to\infty} I_{\text{D}}(t) = \frac{\widehat{\alpha}_{\text{D}}}{\widehat{\alpha}} = \frac{\widehat{\alpha}_{\text{D}}}{\widehat{\alpha}_{\text{D}} + \widehat{\alpha}_{\text{R}}} \tag{2.15}$$

Equivalent results can be obtained for the competing event

$$I_{\text{R}}(t) = \frac{\widehat{\alpha}_{\text{R}}}{\widehat{\alpha}} - \frac{\widehat{\alpha}_{\text{R}} e^{-\widehat{\alpha}t}}{\widehat{\alpha}}$$

$$\widehat{\theta}_{\text{R}} = \lim_{t\to\infty} I_{\text{R}}(t) = \frac{\widehat{\alpha}_{\text{R}}}{\widehat{\alpha}_{\text{D}} + \widehat{\alpha}_{\text{R}}} \tag{2.16}$$

The estimators 2.15 and 2.16 can be computed at every calendar time s. Then they can be plugged in 2.10 leading to a parametric version of the estimator proposed by Ghani et al., 2005:

$$\widehat{\text{CFR}}^{(\text{PARA})} = \widehat{\theta}_D(S) = \frac{\widehat{\theta}_D(s)}{\widehat{\theta}_D(s) + \widehat{\theta}_R(s)}$$

Figure 2.9 reports the cumulative intensity functions of the form reported in Equation 2.14, plugging-in the maximum likelihood estimators of the rates. The limiting value of $I_{\mathrm{D}}(t)$ (Equation 2.15) is also plotted.

If a small simulation study is carried out, the parametric estimator slightly outperforms the estimator proposed by the WHO and, given the assumption of parametric survival, shows a higher confidence in the results (Figure 2.10 (b))

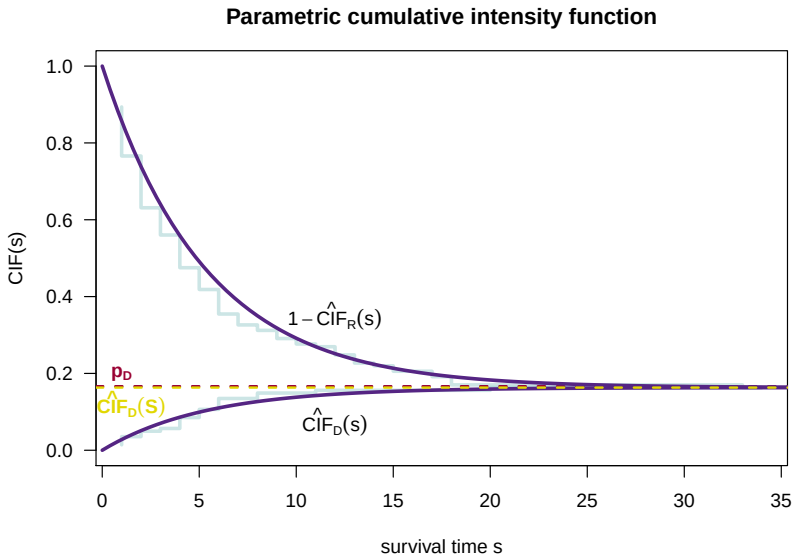
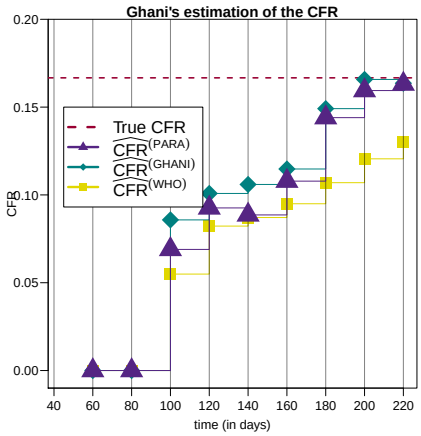
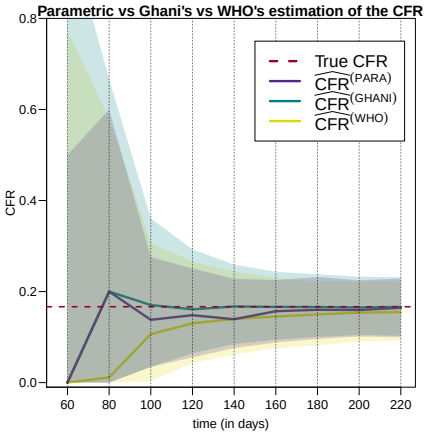


Figure 2.9: Results from the parametric competing-risks analysis of the time to death and recovery. The x-axis is the time to event, the purple curves are the parametric cumulative intensity function of death and 1 minus the cumulative intensity function of recovery computed with the plug in estimator. The limit of the cumulative intensity function of death for $s \to \infty$ is drawn in yellow. The value of CFR used to generate the data is drawn in red and denoted by p_{D}. The pale blue lines in the background represent the non-parametric estimated cumulative intensity function from Figure 2.7 and are plotted for comparison.



(a) $\widehat{\text{CFR}}^{(\text{PARA})}$, $\widehat{\text{CFR}}^{(\text{GHANI})}$ and $\widehat{\text{CFR}}^{(\text{WHO})}$ computed over time.



(b) Median (line) and 95% simulation CIs (shades) of $\widehat{\text{CFR}}^{(\text{PARA})}$ $\widehat{\text{CFR}}^{(\text{GHANI})}$ and $\widehat{\text{CFR}}^{(\text{WHO})}$.

Figure 2.10: In Panel (a) and (b) the x axis is the calendar time from the beginning of the epidemic and the y axis is the estimated CFR. Panel (a) reports in green the estimator 2.13 and in yellow the estimator 2.5 calculated at times $s = 60, 80, \ldots 220$ from the beginning of the epidemic and in red the true value used to generate the data. Panel (b) reports median and confidence intervals of the same estimators computed over 500 simulated datasets.

The parametric formulation considered here is simplistic since the cause-specific hazard functions and the overall hazard function are assumed to be constant over the survival time. The derivations above show that the properties of the Exponential r.v. allow to derive, not only the overall survival, but also the two marginal survival distributions. This is not necessarily true for other distributions.

If other parametric distributions are used to describe survival in a competing-risks setting, such a neat estimation of the CFR is impossible. However, there exists some parametric-survival-analysis packages in R (e.g. Jackson, 2015) that, applied to this context, allow the inference of cause-specific survival functions. Nevertheless, as reported in Appendix B the CFR cannot be derived.

2.4 Analysis of survival data with time-varying rate

While it might be reasonable to assume constant transition intensities over survival time (i.e. Exponentially-distributed times to event), it is less reasonable to assume that the death- and recovery-specific rate would remain constant over calendar time, leading to a constant CFR. More likely, the mortality risk will change during the course of an epidemic as knowledge is acquired on the diagnosis and cure of the disease or, for example, due to different age groups being affected at different times (Yip et al., 2005b).

To further extend the CFR estimator to a time-varying setting, a model developed within demography, the age-period-cohort (APC) model is considered.

2.4.1 The age-period cohort model

The APC model is a descriptive tool for observations in a Lexis diagram, which represents the survival time of an individual with respect to his *age*, the calendar time at which the event of interest happens (*period*) and his date of birth (*cohort*) (Carstensen, 2007). An illustrative Lexis diagram is reported in Figure 2.11, drawing survival times of the Danish prime ministers since the war (Carstensen, 2016). The calendar time is the scale of the x axis, which measures the cohort (i.e. entry, which in this case is the date of birth) and the period (i.e. the time of event, in this case death); the scale of the y axis is instead the survival time which measures the age.

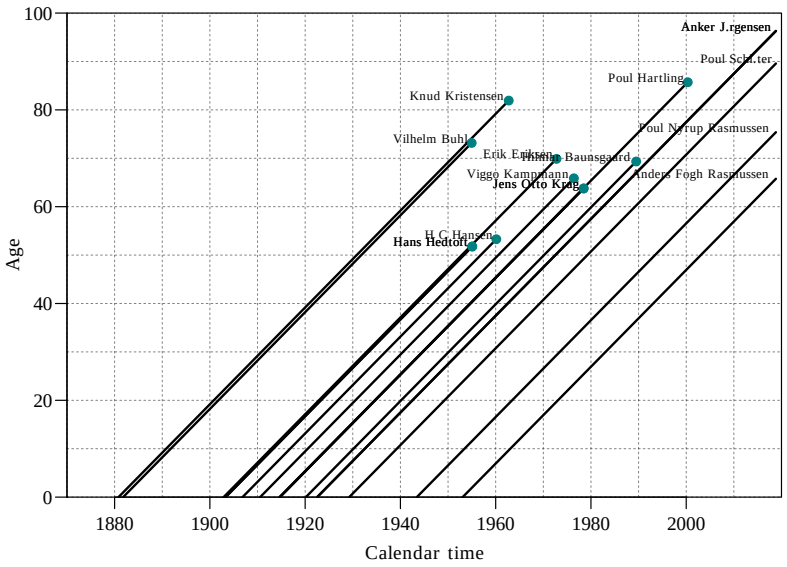


Figure 2.11: Lexis diagram of the time to death of the Danish prime ministers.

Clayton and Schifflers (1987a) and Clayton and Schifflers (1987b) firstly introduced the APC model to investigate whether the rate of the event of interest (in the example above, death) was related to the three quantities measured in the Lexis diagram. Age, period and cohort are mathematically related by the equation

$$period = cohort + age$$

The APC model can be seen as an evolution of the age-cohort model and the age-period model. The three models describe the relationship between log-rates of the event of interest and the respective time-scale(s).

The model can be formulated in the perspective of the Lexis diagram. Data are tabulated by first setting a grid on the age ($a = 1, 2, \ldots, A$ where a is the index of equally spaced intervals of age) and on the cohort time scale ($c = 1, 2, \ldots, C$ where c is the index of an equally spaced intervals of cohort). The second grid uniquely determines also a grid for the period denoted by $p = 1, 2, \ldots, P$. Secondly, for each (a, p) cell, the set of data (d_i, y_i) is recorded, where d_i is the indicator function of an event (i.e. $d_i = 1$ if individual i has the event of interest, e.g. death, within the cell (a, p), 0 otherwise) and y_i is the time at risk for the individual i in cell (a, p). i is the index of the patient in cell (a, p) and therefore takes values $i = 1, \ldots, I_{(a,p)}$ where $I_{(a,p)}$ is the total number of individuals within cell (a, p), usually called the *risk set*. Denote by $\lambda(a, p)$ the rate of the event of interest (e.g. death) in age interval a and period interval p.

Given some functions (e.g. polynomial or splines) $f(\cdot), g(\cdot)$ and $h(\cdot)$, the general form of the APC model for the rates $\lambda(a,p)$ (or equally $\lambda(a,c)$) is:

$$\log[\lambda(a,p,c)] = f(a) + g(p) + h(c)$$
$$\log[\lambda(a,p)] = f(a) + g(p) + h(p-a)$$

Over intervals sufficiently small to assume constant $\lambda(a,p)$, the likelihood contribution of each (d_i, t_i) is given by:

$$L(\lambda(a,p)|d_i,y_i) = \mathrm{e}^{-\lambda(a,p)y_i} \times \lambda(a,p)^{d_i}$$

with log-likelihood

$$\ell(\lambda(a,p)|d_i,y_i) = d_i \log(\lambda(a,p)) - \lambda(a,p)y_i \tag{2.17}$$

which is similar to the Poisson distribution with parameter $\mu_i = y_i\lambda$, i.e. $D_i \sim \text{Pois}(y_i\lambda)$:

$$\begin{aligned}\ell(\lambda y_i|d_i) &= d_i \log(\lambda y_i) - \lambda y_i \\ &= d_i \log(\lambda) - \lambda y_i + d_i \log(y_i)\end{aligned}$$

The Poisson log-density is equal to Equation 2.17 except for the term $d_i \log(y_i)$ which is constant w.r.t. λ.

Hence, a Generalized Linear Model with Poisson distribution, logarithmic link function and offset $\log(y_i)$ can be used to model the count data d_i. The likelihood under this model can be maximized leading to ML estimates of the parameters of the functions $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$. Within the Poisson machinery, several shapes of $f(a)$, $g(p)$ and $h(c)$ can be considered. Age, period and cohort can be thought of as categorical variables, where in each category, rates are assumed to be piecewise constant and one estimate per cell is obtained. Alternatively, linear, polynomial or smooth (spline) relations can be assumed.

The maximization of the Poisson likelihood does not have a unique solution: different combinations of coefficients for period and cohort lead to the same fitted values, and therefore one set of parameters must be fixed to identify the other. This compromises the interpretation of the parameters and the simpler models formulated in Clayton and Schifflers (1987a) might be practically more useful. The age-period model investigates the effect of the survival time and of the current calendar time on the event rate

$$\log[\lambda(a,p)] = f(a) + g(p)$$

while the age-cohort model investigates the effect of the survival time and the calendar time of the entry in the cohort on the event rate

$$\log[\lambda(a,c)] = f(a) + h(c)$$

The methods to set the APC, age-period and age-cohort models are implemented in the `R` package `Epi` (Carstensen et al., 2018) and generalized linear models can be fitted to obtain estimates of the model-parameters.

2.4.2 Estimation of the time-varying Case Fatality Risk with the age-cohort model

Time-to-event data recording dates of admission to hospital or Intensive Care Units (ICUs) and date of death or recovery of each individual can be represented by means of a Lexis diagram. A

dataset of this kind is displayed in Figure 2.12, where some instances of simulated data of daily admissions to ICU, deaths and recoveries over a seasonal influenza epidemic are plotted against calendar and survival time. In the simulated data, the rate of death rapidly decreases from soon after the middle of the epidemic (around calendar day 130).

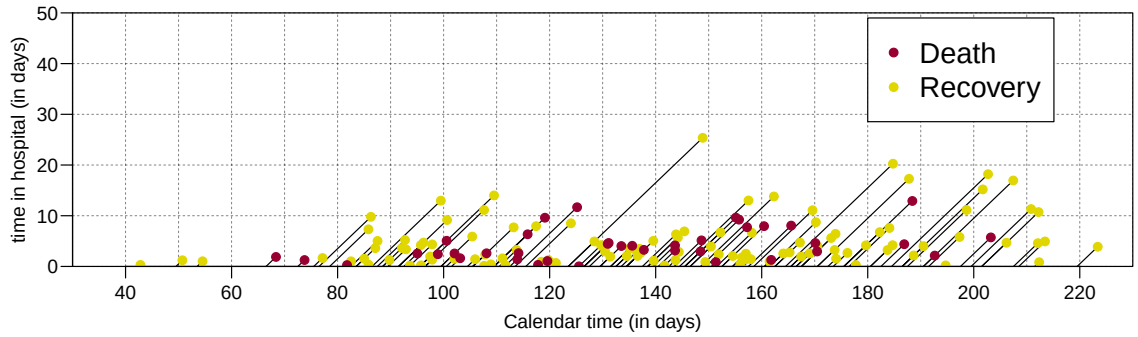


Figure 2.12: Lexis diagram of the time to death (red) and recovery (yellow) of influenza cases admitted to ICU.

In this context, entry into the cohort takes place when an individual is admitted to ICU; consequently, the *age* variable is the time since ICU admission and the *cohort* is the calendar day of ICU admission, where calendar time is counted from $c = 0$, the beginning of the epidemic.

The following age-cohort model can be fitted to the data (y_i, d_i) where d_i is the indicator function of death and y_i is the time of death:

$$\log[\lambda_{\mathrm{D}}(a, c)] = f_{\mathrm{D}}(a) + h_{\mathrm{D}}(c) \tag{2.18}$$

from which the death-specific hazard rates are estimated. Likewise, a model can be fitted to estimate the recovery-specific hazard rates.

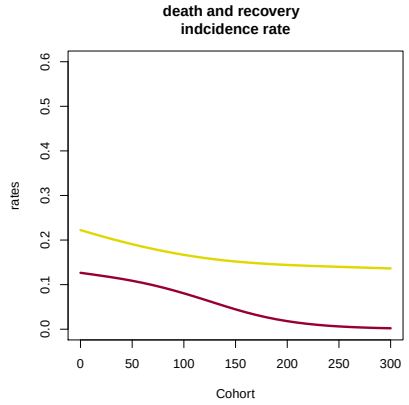
$$\log[\lambda_{\mathrm{R}}(a, c)] = f_{\mathrm{R}}(a) + h_{\mathrm{R}}(c) \tag{2.19}$$

The two models were fitted separately, leading to the estimation of age and cohort effect on cause-specific (i.e. death and recovery) hazard. The shape of the functions $f_{\mathrm{D}}(\cdot), h_{\mathrm{D}}(\cdot), f_{\mathrm{R}}(\cdot)$ and $h_{\mathrm{R}}(\cdot)$, can be chosen according to the data observed, but usually flexible functions, such as splines, describe well temporal dynamics.

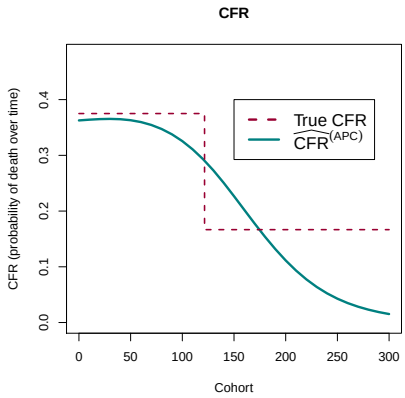
In this context, the estimation of CFR when the epidemic has finished consists, again, of the evaluation of the cumulative intensity function while the survival time (in the APC notation, the age) approaches ∞. The estimate of the cumulative intensity function of death as $a \to +\infty$ for individuals admitted to hospital in cohort c is

$$\widehat{\mathrm{CFR}}^{(\mathrm{APC1})}(c) = \widehat{I}_{\mathrm{D}}(c) = \sum_{a=1}^{A} \widehat{\lambda}_{\mathrm{D}}(a, c) \exp(\widehat{\lambda}_{\mathrm{D}}(a, c) + \widehat{\lambda}_{\mathrm{R}}(a, c)) \tag{2.20}$$

where $\widehat{\lambda}_{\mathrm{D}}(a, c)$ and $\widehat{\lambda}_{\mathrm{R}}(a, c)$ are the predicted death- and recovery-specific hazards obtained by fitting the age-cohort models (2.18) and (2.19), respectively. $\widehat{I}_{\mathrm{R}}(c)$ can be obtained similarly. Estimates of the death and recovery rates from a simulated dataset, as well as the estimated time-varying CFR are plotted below in Figure 2.13. Note that the dataset used here is 5 times bigger than the dataset used in Sections 2.3 and 2.4: when this method is attempted on smaller



(a) $\widehat{\alpha}_{\mathrm{D}}$ (red) and , $\widehat{\alpha}_{\mathrm{R}}$ (yellow) at constant age and varying cohort.



(b) Estimated $\widehat{\mathrm{CFR}}^{(\mathrm{APC1})}$ over time

Figure 2.13: In Panel (a) and (b) the x axis is the calendar time from the beginning of the epidemic. Panel (a) reports the death and recovery rates, in this analysis they have been assumed constant over age. Panel (b) reports in green the estimated CFR with this method and in red the true value.

sample sizes the model performs poorly, not identifying the time-variation of the CFR and giving very variable estimates.

When the epidemic is still ongoing and data are affected by right censoring, approximation 2.10 can be applied to estimate the CFR at an early stage

$$\widehat{\mathrm{CFR}}^{(\mathrm{APC2})}(c) = \frac{\widehat{I}_{\mathrm{D}}(c)}{\widehat{I}_{\mathrm{D}}(c) + \widehat{I}_{\mathrm{R}}(c)} \tag{2.21}$$

A comparison with the methods reported in the previous section seems unfair since under a time-varying-CFR scenario they would be clearly outperformed by the APC method proposed since they would average out the effect over time.

By contrast, when these methods are used early in the epidemic, when data are sparse, they perform poorly, since a decent amount of data is required to infer, not only the average CFR but also its variation over time.

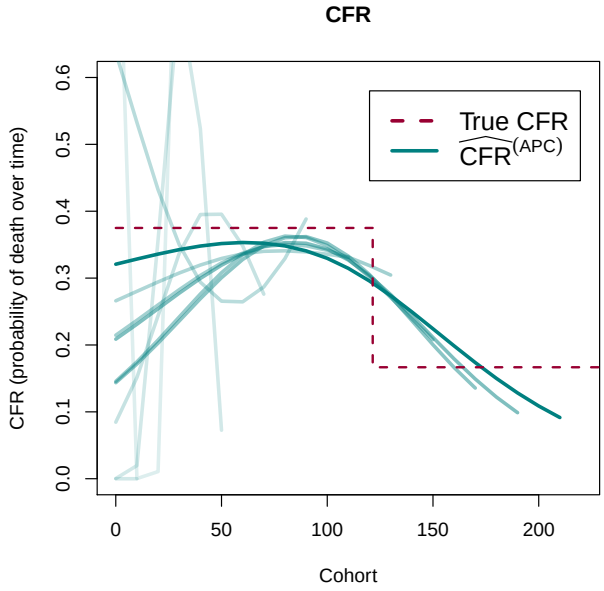


Figure 2.14: The estimated cohort-varying CFR is here plotted at various truncation times. The lighter the lines, the more the dataset was truncated as if the epidemic was at its beginning.

2.5 Analysis of count data

If survival data exist, they often inform only a very small sample of the population. This is the case for the survival data from the UK Severe Influenza Surveillance System (USISS) sentinel scheme: they describe only the people admitted to ICU in a few sentinel trusts (before 2013) or only the children admitted in these trusts (after 2013). Models such as the age-cohort model might lead to misleading results (see Section 2.6 below for an example).

By contrast, count data on a larger population might be available. These data are easier to collect but they do not carry any information on the time elapsing between events.

2.5.1 Estimation of the Case Fatality Risk using counting processes

The method reviewed below (Yip et al., 2005a,b) assumes that count data arise from a chain-Multinomial model over discrete time (e.g. days) that is indexed by s on the calendar scale, $s = 0, 1, 2, ...S$, with $s = 0$ and $s = S$ being the beginning and the end of the epidemic respectively. Assume that at each day s, every hospitalized patient can either die, recover or stay in the hospital according to a Multinomial r.v.. The transitions are determined by the daily probability of dying and of recovering, that depend only on calendar time s. The CFR is assumed to vary over time.

Denote by:

p_{Ds} the probability of dying during day s given that a person is in the hospital at the beginning of day s;

p_{Rs} the probability of recovering during day s given that a person is in the hospital at the beginning of day s;

N_{Ds} the number of patients that die during day s;

$N_{\text{R}s}$ the number of patients that recover during day s;

H_{s-1} the number of hospital inpatients at the beginning of s that includes all the new admissions during day $s-1$ and all those that did not die during day $s-1$.

Then the chain-Multinomial model assumption is

$$N_{\text{D}s}, N_{\text{R}s}, N_{\text{H}s} | H_{s-1} \sim \text{Multinomial}(H_{s-1}; p_{\text{D}s}, p_{\text{R}s}, 1 - p_{\text{D}s} - p_{\text{R}s}) \tag{2.22}$$

A graphical representation of what happens at each day s is reported in Figure 2.15.

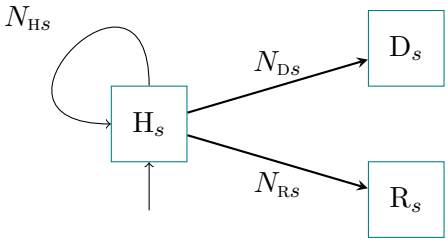


Figure 2.15: Chain-Multinomial model for daily inpatients, deaths and recoveries: H_s, D_s and R_s represent the number of people in the hospital at the beginning of interval s and $N_{\text{H}s}$, $N_{\text{R}s}$, $N_{\text{D}s}$ represent the number people who stay in the hospital, recover or die during day t according to a Multinomial r.v..

Following the assumption of time varying CFR,the time-varying case fatality risk (tvFR) can be defined via Equation 2.23, as it is the conditional probability of death given either recovery or death, denoted by π_s:

$$\pi_s = \frac{p_{\text{D}s}}{p_{\text{D}s} + p_{\text{R}s}} \tag{2.23}$$

Consider one realization of the Multinomial process at a given time s: patients die with probability $p_{\text{D}s}$ and recover with probability $p_{\text{R}s}$. The remaining proportion will eventually experience one of the two events. If we assume that this proportion of people still in hospital will experience the same probabilities $p_{\text{D}s}$ and $p_{\text{R}s}$ as those who have already had one event, then π_s would describe exactly the overall probability of death, i.e. the CFR at day s.

To obtain a naive, instantaneous estimator of the tvFR, ML estimators of the daily probability of each event in the Multinomial distribution are exploited, these are:

$$\widehat{p}_{\text{D}s} = \frac{N_{\text{D}s}}{H_{s-1}}$$
$$\widehat{p}_{\text{R}s} = \frac{N_{\text{R}s}}{H_{s-1}}$$

which, plugged into Equation 2.23, leads to the naive estimator:

$$\widehat{\text{tvFR}}^{(\text{NAIVE})} = \widehat{\pi}_t = \frac{\widehat{p}_{\text{D}t}}{\widehat{p}_{\text{D}t} + \widehat{p}_{\text{R}t}}$$

2.5.2 Smoothing the Case Fatality Risk

The resulting estimator is very variable over time, above all at the start and end of the epidemic where few events take place.

To smooth $\widehat{\text{tvFR}}^{(\text{NAIVE})}$, kernel smoothed estimates, $\widetilde{p}_{\text{D}s}$ and $\widetilde{p}_{\text{R}s}$ can computed as a weighted average of the daily naive $\widehat{p}_{\text{D}s}$s and $\widehat{p}_{\text{R}s}$s. This can be done by choosing a kernel bandwidth b and a kernel function, for example the Epanechnikov kernel function,

$$K(x) = [3/4]\,(1 - x^2)I(|x| < 1)$$

and the bandwidth kernel function:

$$K_b(x) = [1/b]\,K(x/b)$$

For each s, a vector of weights $\boldsymbol{w_s}$ can be computed, where each element w_{qs} is

$$w_{qs} = [1/b]\,K\left(\frac{s-q}{b}\right),$$

which represent the contribution of the observation at time s to the estimate at time q. For each s the smoothed estimate $\widetilde{p}_{\text{D}s}$ as a weighted average of the $\widehat{p}_{\text{D}s}$ with weights w_{qs} is:

$$\widetilde{p}_{\text{D}s} = \frac{\sum_{q=1}^{S} w_{qs}\widehat{p}_{\text{D}s}}{\sum_{q=1}^{T} w_{qs}}$$

and similarly

$$\widetilde{p}_{\text{R}s} = \frac{\sum_{q=1}^{S} w_{qs}\widehat{p}_{\text{R}s}}{\sum_{q=1}^{T} w_{qs}}$$

Plugging both the estimators $\widetilde{p}_{\text{D}s}$ and $\widetilde{p}_{\text{R}s}$ in 2.23, a smoothed version of the tvFR can be obtained:

$$\widehat{\text{tvFR}}^{(\text{SMOOTH})} = \widetilde{\pi}_s = \frac{\widetilde{p}_{\text{D}s}}{\widetilde{p}_{\text{D}s} + \widetilde{p}_{\text{R}s}}$$

The optimal level of smoothing of an observed dataset is hard to define, ideally a random oscillation around a mean should be smoothed, but long-term trend, as well as sudden significant changes in CFR should be captured.

2.6 Applicability of the methods to the available data

Many methods have been reviewed and this section addresses the possible use of these methods in the inference of the probability of death within hospitalised or ICU influenza cases in England.

Data on severe influenza cases are collected via the USISS, which comprises a sentinel scheme (Health Protection Agency, 2011b), that involves a stratified sample of English trusts, and a mandatory scheme, that collects data from ICUs in all trusts in England (Health Protection Agency, 2011a).

The former contains individual-level data of confirmed-influenza ICU admissions such as day of symptom onset, day of hospitalization, day of ICU admission and final outcome (e.g. death, discharge). Moreover it contains count data on the weekly number of hospitalisations and deaths. The mandatory scheme instead, contains weekly counts of confirmed-influenza ICU admissions and deaths.

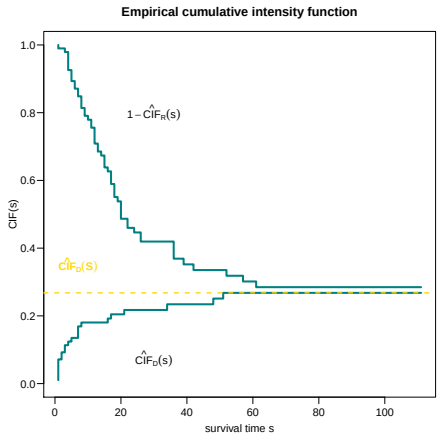
For the estimator proposed above in Section 2.3 and 2.4, the only useful data are the individual-level timings obtained from the sentinel collection scheme. They are survival data that fit in the competing-risks setting of Section 2.3 and 2.4.

Data collected from the 2013/14 influenza season onwards cannot be used to get a full picture of the population since they monitor paediatric cases only.

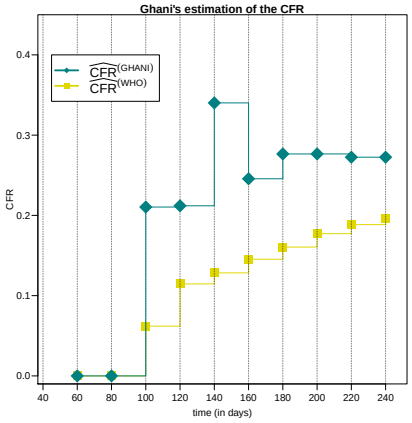
A preliminary exploration of the data from seasons 2010/11, 2011/12 and 2012/13 showed that the reporting process was not rigorous and a lot of information is missing. The dataset that was affected by this problem the least was the one collected during the 2012/13 season, where almost 100 individuals had the variables of interest (ICU admission date, final outcome and final outcome date) non- missing.

These survival data were analysed using the WHO estimator, Ghani's estimator, and its parametric version under the assumption of Exponentially-distributed times to event. When the epidemic is finished, the WHO estimator led to a CFR of about 20%, and the empirical cumulative intensity functions points at a value closer to 30% (see Figure 2.16 (a)).

When the estimate is computed over time, Ghani's estimate oscillates around the final value as soon as the first deaths are recorded, way earlier than the WHO estimator.



(a) Empirical cumulative intensity functions.



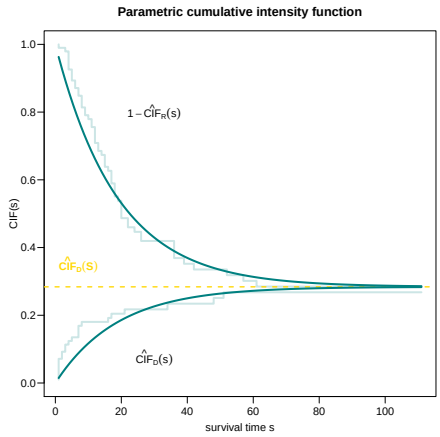
(b) $\widehat{\text{CFR}}^{(\text{WHO})}$ and $\widehat{\text{CFR}}^{(\text{GHANI})}$ at different truncation time

Figure 2.16: Panel (a) plots the result of the non-parametric competing-risks analysis of the data: the x axis represent survival time and the y axis the cumulative intensity functions. Panel (b) reports Ghani's and the WHO estimates of the CFR computed at different times from the beginning of the epidemic (x axis).

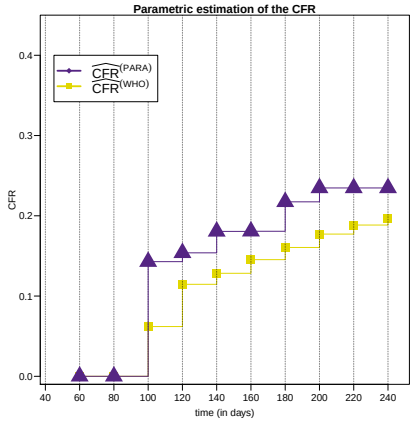
The parametric version of Ghani's estimator, at the end of the epidemic, leads to a value slightly lower than its non-parametric counterpart but it takes slightly more time to arrive at convergence (see Figure 2.17 (b)), probably due to the poor fitting of the Exponential distribution for shorter survival times (see Figure 2.17 (a)).

Despite the data being so sparse, an attempt to fit the APC-based estimator was made. A spline transformation was used and the estimated CFR is reported in Figure 2.18. As expected, the estimates at the end and at the beginning of the epidemic were extremely variable. Moreover a steeply increasing trend in severity is detected around the epidemic peak, this might be due to better reporting of severe events, rather than to an actual increase in the probability of death.

None of the estimators reviewed or proposed in this section is suitable for the analysis of



(a) Parametric cumulative intensity functions.



(b) $\widehat{\text{CFR}}^{(\text{WHO})}$ and $\widehat{\text{CFR}}^{(\text{PARA})}$ at different truncation time

Figure 2.17: Panel (a) plots the result of the parametric competing-risks analysis of the data: the x axis represent survival time and the y axis the cumulative intensity functions. Panel (b) reports the parametric and WHO estimates of the CFR computed at different times from the beginning of the epidemic (x axis).

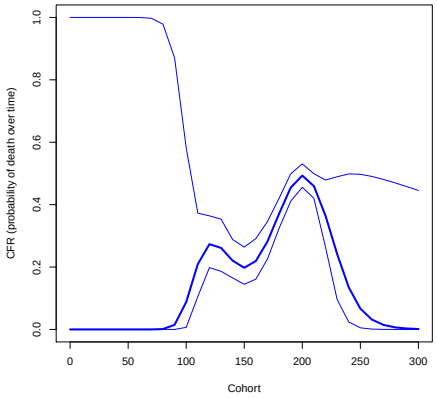


Figure 2.18: Plot of the estimated CFR and its CI from the analysis of real data.

count data collected through USISS. These data are collected on a large population (all the trusts in England) but they collect only data on ICU admissions and deaths. Information about the recovery process in completely absent and is required for all the estimators presented above. A new estimator is needed to exploit the information collected via this scheme correcting for the underestimation of the WHO estimate. This issue is addressed in Chapter 4.

2.7 Conclusions

This Chapter serves as a review of the currently available methods aimed at estimating severity. The motivating problem is the fact that the estimator used in practice and suggested by the WHO underestimates the probability of death (measured by the CFR) while the epidemic is taking off.

The methods proposed in the literature comprise, among others: (i) one application of competing-risks theory to the severity setting for the analysis of time-to-event data under the

assumption of a constant CFR and (ii) another estimator aimed at the analysis count data relaxing the constant CFR assumption.

Section 2.4 fills the gap in the literature, proposing an estimator that exploits time-to event data without necessarily assuming a constant CFR.

All the methods proposed and reviewed assume the availability of data on the recovery process, de facto introducing more information than the one used by the WHO estimator. Survival data on influenza patients from a sample of ICUs in England are analysed in Section 2.6. This dataset however, is highly affected by missingness and comprises very few patients.

By contrast, count data on the number of ICU admissions and deaths in all the trusts in England are collected routinely by PHE providing a representative source of information. Nevertheless, the only estimator that relies on these data alone is still the one proposed by the WHO and therefore a good alternative is still to be found. A new method to estimate the CFR with only admission and death counts is proposed in Chapter 4.

A wider look into this problem indicates that extra information might be accounted for while modelling the severity of influenza. Flu' cases, both severe and not, will follow over time an epidemic pattern, increasing when the outbreak is spreading and decreasing while it is fading away. Hence, it would be useful to approach the analysis of data on ICU/hospital admissions and deaths within a model that accounts for epidemic dynamics. In this model, other sources of information might be taken into account to further inform the process that generates case-data at different levels of severity. The rest of this thesis approaches the investigation of the dynamics of influenza severe cases within this perspective.

Chapter 3

Estimation of transmission

This chapter contains an introduction to epidemic models (Section 3.1), to their parameter estimation (Section 3.2), and an example of their use within the analysis of infectious disease data (Section 3.3).

The literature on epidemic models is vast and varied, as is reported in Subsection 3.1.1, and a comprehensive review and classification of these models is beyond the scope of this thesis. Instead this chapter reports the most common approaches in the field and lists some of their pitfalls and advantages. In light of the latter, model choices are made with respect to a real-case study, whose analysis is reported in Section 3.3. The two sections are linked by a part dedicated to estimation methods for transmission and their application to the case of influenza (Section 3.2).

The chapter ends with a discussion of the models mentioned and the results obtained in the analysis performed (Section 3.4).

3.1 Epidemic models

There are many reasons for which the investigation of epidemics is intriguing. Niels G. Becker, in his 1989 book (Becker, 1989) highlights the following three points:

- *infectious disease data are not the result of planned experiments, but arise from naturally occurring epidemics*
  this aspect discriminates epidemic models from many other statistical models, which rely on the possibilities of having replicates of the same experiments;
- *infectious disease data are highly dependent because infected cases are the cause of further infected cases*
  epidemic models need to account for high dependence within data, which lead to each data-point being informative on the whole process but reduces the amount of independent information from multiple observations on the same epidemic;
- *the infection process is generally only partially observable*
  this latter feature, on the one hand, encourages the development of new methods to approximate the likelihood for incomplete/partial data, and, on the other hand, favours the use of Bayesian methods that exploit other sources of information, where available.

Motivated by these stimuli, the literature on epidemic models and their inference has proliferated in the last two centuries (Heesterbeek et al., 2015).

A brief summary of the main historical steps of epidemic theory, together with the books that represent its cornerstones, are contained in Subsection 3.1.1. The following subsections instead define more formally the theory behind epidemic models and their classification.

3.1.1 Historical background on the analysis of epidemics

The need for a formal analysis of infectious disease data was documented as early as 1840 in a letter by William Farr addressed to the Registrar General for England and Wales (Farr, 1840). In the letter, he indicates the seasonal oscillation of some diseases and encourages their biological and mathematical analysis (Serfling, 1963). Since then an increasing number of analyses of the temporal pattern of infectious disease cases and deaths have been carried out. Of particular relevance are the studies by Brownlee (1907), where the temporal pattern was compared to the Normal distribution, and by Ross (1911) where a first deterministic epidemic model was formulated, linking the future number of infections to the current number of infected people. Later in the 20^{th} century the most relevant advances in the field were made. Kermack and McKendrick wrote a series of papers on the mathematical theory of infectious diseases (Kermack and McKendrick, 1927, 1932, 1933). In these papers, not only did they define models already similar to those currently used, but they also studied their behaviour, giving the first definition of an epidemic threshold, *i.e.* the composition and size of the population necessary for an epidemic to take off.

Following in their footsteps, Greenwood and Reed and Frost provided two formulations of a stochastic epidemic model that allows fluctuations due to chance. The former is documented in a paper of 1931 (Greenwood, 1931), whereas the latter can only be found in later citations (Abbey, 1952) or in recorded lectures. A previous use of stochastic epidemic models is attributed to P. D. En'ko as early as the end of the 19^{th} century (Dietz, 1988).

Closer to the end of the century, books such as Bailey (1975) and Becker (1989) started the process of collecting the work in this field, creating solid references for the literature that followed. Work on these topics proliferated, boosted by the occurrence of severe outbreaks such as the HIV/AIDS epidemic.

Several recently-published books on infectious disease modelling provide a comprehensive explanation of these models and their application. The language and the complexity of their content varies according to the targeted audience. Vynnycky and White (2010) give an intuitive introduction to the models, directed to a broad audience, including public health providers and managers, whereas Diekmann and Heesterbeek (2000) is aimed at mathematical modellers and computational biologists, providing insights on the models adopted and derivations of theoretical results. Keeling and Rohani (2011) can be seen as an hybrid between these two extremes, adopting a pragmatic and practical approach and justifying some relevant theoretical arguments.

The content of this section is based on the three mentioned books and on some foundation papers (e.g Britton (2010)). Unfortunately, the notation is not consistent across these books and within the epidemic models literature (e.g. one should not take for granted that Vynnycky's β has the same meaning as Keeling's β). For this reason all the elements are re-defined below, ensuring consistency across this thesis.

Lastly, there are many dimensions on which models could differ: for example time might be measured continuously or at different discrete times, or the individuals involved in the epidemic might be aggregated in different ways. The following subsections approach the formulation of epidemic models from many points of view, proposing multiple classifications.

3.1.2 Model definition

The spread of an infectious disease in a population is characterised by specific processes that are assumed in many epidemic models, as described by Becker (1979):

"A susceptible may become infected as a result of an adequately close contact with an infective. The newly infected individual passes through a stage during which he is latent and then through a stage during which he is infectious, before being removed by isolation, by death or by naturally losing his infectiousness and becoming immune for the remaining duration of the epidemic."

More specifically, at an individual level, an infection takes place when a pathogenic micro-organism invades a host (WHO, 2018a). Then the disease progresses and the amount of pathogen increases within the host, stimulating an immune response (Keeling and Rohani, 2011). There are therefore two different classifications of the host: the first one, according to his infection status (e.g he might be susceptible, infectious, immune) and a second one according to his clinical status (e.g. he might be asymptomatic or diseased). An illustration of the growth of the pathogen within a host is reported in Figure 3.1.

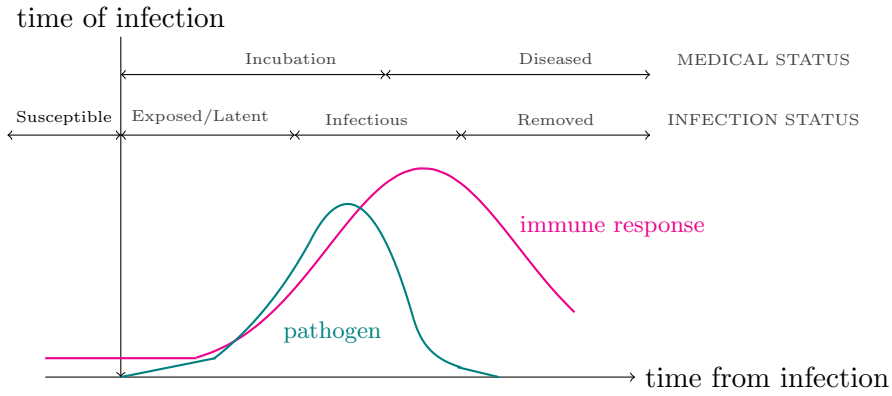


Figure 3.1: A representative caricature of the infection process within a host. The dynamics of the pathogen (green) are associated with infection status (e.g. Susceptible, Exposed, Infectious and Removed); the immune response instead (magenta) is associated with medical statuses (e.g. asymptomatic, mildly ill, severely ill). Note that these dynamics are not quantified here (the y-axis is not labelled), this figure only depicts the temporal dynamics of the pathogen. Figure freely adapted from Keeling and Rohani (2011).

Following the notation of Section 2.1, the infection status of an individual i over his ageing time s can be described by a multi-state process. At initial time $s = 0$, this individual is in state $X_i(0) = x_i(0)$ with $x_i(0) \in \{S, E, I, R\}$, these letters representing susceptible (S), exposed (E, *i.e.* infected but not yet infectious), infectious (I) and removed (R). As the ageing time s progresses he might move to another state a with a rate $\alpha^i_{x_i(0)a}(s)$, with $a \in \{S, E, I, R\}$. The multi-state process of the SEIR model is represented in Figure 3.2.

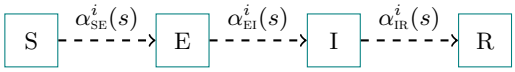


Figure 3.2: Graphical representation of the SEIR infection process for an individual i.

The SEIR model is only one of the many types of models used to describe different infectious diseases. For example, if the infected individual becomes infectious as soon as the pathogen enters his body, a SIR model is used to describe the epidemic since there is no latent state. In the case of no (or ineffective) immune response the R state is discarded, leading to SIS and SEIS models. If instead, the host becomes immune to the virus, but this immunity fades away with time, making him susceptible again, the models are called SIRS and SEIRS. Figure 3.3 displays the multi-state models associated with this assortment of epidemic models.

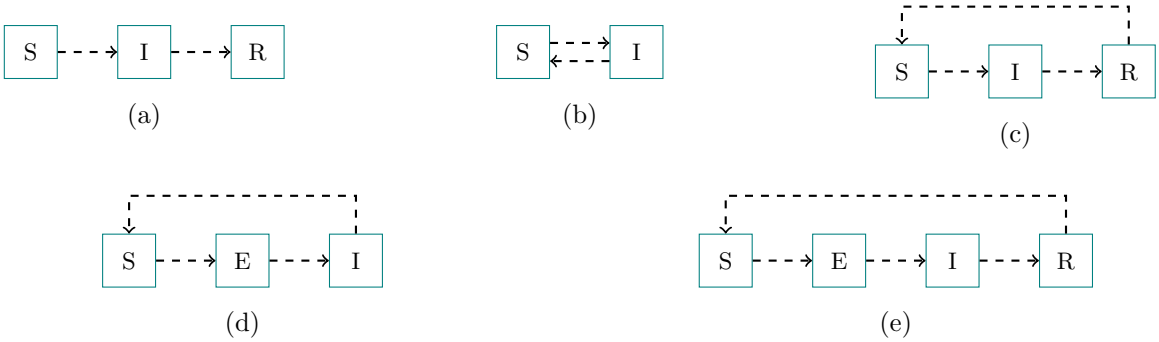


Figure 3.3: Graphical representation of the (a) SIR, (b) SIS, (c) SIRS, (d) SEIS and (e) SEIRS multi-state models.

The structure of the multi-state model (*i.e.* which compartments to use) is imposed by the virus under analysis. Since influenza virus development exhibits a latent period and grants immunity (Carrat et al., 2008), a SEIR model is used throughout this chapter. Extension of the results presented in Sections 3.1.3-3.1.7 to other epidemic models is straightforward.

3.1.3 From individual to population level: model set-up

The dynamics of an epidemic within a population can be approximated at different levels of granularity. What happens in reality is that each single individual in the population has his own history across the epidemic states and the mathematical model used to describe this process can be more or less realistic.

Agent based models (Ajelli et al., 2010) consider single multi-state processes for each individual (named "agent"). This is the maximum resolution that an epidemic model could reach.

The contacts across individuals could be explicitly modelled by a network. The analysis of the spread of infectious disease on such constructions takes the name of *network epidemic models* (Keeling and Eames, 2005) and these models have recently proliferated (Duan et al., 2015).

A further approximation can be made by grouping individuals for which the frequency of contact, as well as the propensity to infect and be infected can be considered homogeneous. Aggregations of such individuals are commonly named "households" and the models that explicitly account for them are called *meta-population models* (Ball et al., 2014).

Finally the population could be considered completely homogeneous (both in contact and

susceptibility) leading to *single population models*. Here the information on the history of a single individual is lost and only the count of individuals moving across compartments is tracked. The models presented below are single population models (Keeling and Rohani, 2011).

To describe the average dynamics of a population model of, for example, a SEIR epidemic, a set of differential equations can be used to express the outflow and inflow of individuals in each of the four compartments defined. Denote by $t \in [0, T]$ continuous time, with 0 and T denoting the beginning and the end of the epidemic. At time $t \in [0, T]$, denote by $S(t)$ the number of susceptible individuals, by $E(t)$ the number infected, but not yet infectious, by $I(t)$ the number infectious and by $R(t)$ the number of people that are removed either by recovering and becoming immune or dying. Moreover, consider a closed population (*i.e.* with no birth, death, immigration or emigration) of fixed size N so that $N(t) = N = S(t) + E(t) + I(t) + R(t)$, $\forall t \in [0, T]$. Then the system of equations takes the form:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta S(t)\frac{I(t)}{N} \\ \frac{dE(t)}{dt} &= +\beta S(t)\frac{I(t)}{N} - \sigma E(t) \\ \frac{dS(t)}{dt} &= +\sigma E(t) - \gamma I(t) \\ \frac{dS(t)}{dt} &= +\gamma I(t) \end{aligned} \tag{3.1}$$

where β denotes the transmission rate, which accounts both for the probability of transmission given contact (κ) and for the contact rate (ν), so that $\beta = \kappa\nu$. The infection rate $\beta\frac{I}{N}$ also depends on the proportion of infectious people, $\frac{I}{N}$, the infection pressure to which the population is subject. σ and γ denote the rates of exit from the latent and infectious states, respectively (Keeling and Rohani, 2011).

An epidemic model such as (3.1) can be represented by a plot similar to the multi-state-model plot of Figure 3.2, though each box represents the number of individuals present in that compartment at time t, and the arrows represent the people transiting between compartments and are overlaid by the individual-level transition rates.

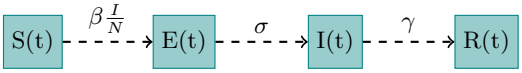


Figure 3.4: Graphical representation of the SEIR population epidemic model.

3.1.4 Time in epidemic models

Equation 3.1 expresses the temporal variation of the average compartment sizes via differential equations. However, often, both for computational convenience and interpretation, it is better to consider discrete time and to use difference equations.

Consider intervals of length $\delta \ll 1$ so that the t-th interval is defined as $[\delta t, \delta t + \delta)$ and the intervals are indexed by $t = 0, 1, 2, \ldots, T$. Denote the number of susceptible individuals at the beginning of interval t by S_t and likewise for the other compartments. Equation 3.1 can be

approximated as:

$$\begin{aligned} S_{t+1} &= S_t - (\beta\delta)S_tI_t/N \\ E_{t+1} &= E_t + (\beta\delta)S_tI_t/N - (\sigma\delta)E_t \\ I_{t+1} &= I_t + (\sigma\delta)E_t - (\gamma\delta)I_t \\ R_{t+1} &= R_t + (\gamma\delta)I_t \end{aligned} \tag{3.2}$$

where two approximations are made:

i The variation of the number of people within an interval is considered irrelevant. This approximation is particularly influential for the formulation of the infection rate: the infection pressure experienced by the population over the interval $[t\delta, t\delta + \delta)$ is considered equal to the infection pressure at the beginning of the interval, even though some people might become infectious or stop being infectious during δ.

ii While a continuous-time system is defined by event rates, *i.e.* instantaneous probabilities of events, a discrete-time system is defined by the probabilities of events within each interval. In Equation 3.2, these probabilities are approximated by the rates thanks to the assumption of Exponentially distributed inter-event times (see the next subsection for more exposition on this). For example, the probability of leaving the E state during an interval of length δ for the ith individual is:

$$P(T_{\mathrm{EI}}^{(i)} \leq \delta) = 1 - e^{-\sigma\delta}$$

whose Taylor expansion, for a small δ, leads to an approximation given by the rate times the interval size:

$$P(T_{\mathrm{EI}}^{(i)} \leq \delta) = 1 - e^{-\sigma\delta} = \sigma\delta + O(\delta^2)$$

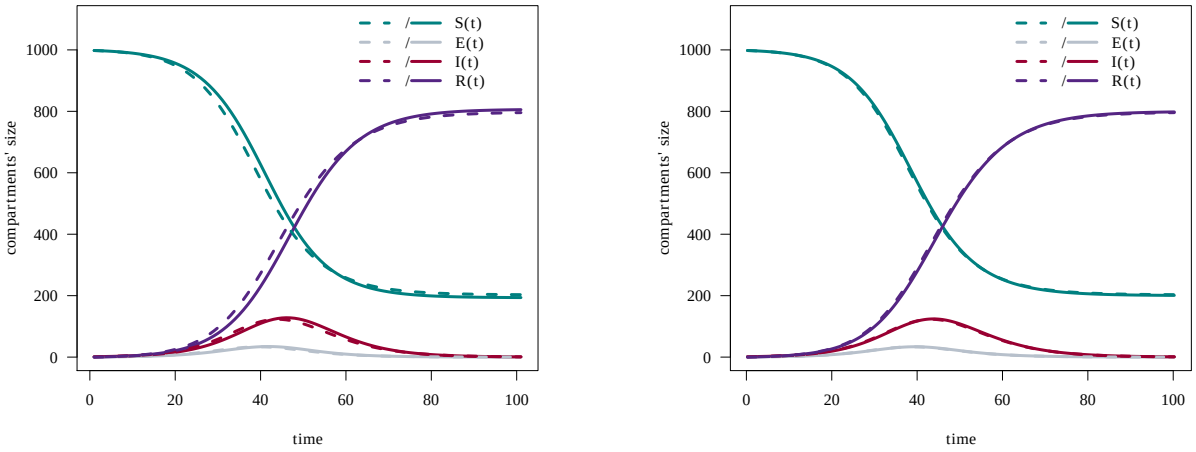
If δ is small enough for these approximations to hold the discrete-time model (3.2) can replace the continuous time model (3.1).

Figure 3.5 shows how the size of δ might affect the approximation. A SEIR deterministic epidemic on a small population ($N = 1000$) is evaluated and $\delta = 1$ day and $\delta = 1/4$ day are used for the discrete-time approximation. The set of parameters used is: $\beta = 0.40, \sigma = 0.8, \gamma = 0.2$. The colours of the lines identify the compartment (S, E, I and R). The discrete-time approximation of the compartment sizes (solid line) is delayed, compared to the continuous-time solution (dashed line), however this delay decreases as δ approaches 0.

Both in a discrete-time and in a continuous-time epidemic, time is said to run from 0 to T, but there are multiple ways to define these time-points.

In reality the beginning of an epidemic ($t = 0$) coincides with the introduction of an infected person ($I(0) = 1$), nevertheless it is possible to model the epidemic over a later time-window and define an arbitrary $t = 0$, for example the calendar time at which the data collection begins, and therefore allow $I(0) > 1$.

Regarding instead the end of an epidemic ($t = T$), this takes place when the infection rate is 0, *i.e.* when there are either no more infected/infectious or no more susceptible individuals. Yet, for continuous- and discrete-time deterministic epidemics, the number of people in each compartment only approaches 0 asymptotically, never reaching the value. In this case the end



(a) Continuous-time model vs discrete-time model with $\delta = 1$ day

(b) Continuous-time model vs discrete-time model with $\delta = 1/4$ day

Figure 3.5: Average number of susceptible (green), latent (grey), infectious (red) and removed (purple) people at time t. The dashed line is the continuous-time solution and the solid line is the discrete-time approximation, with time step of (a) a day and of (b) a quarter of a day.

of an epidemic can be defined either using thresholds for $S(t), E(t)$ and $I(t)$ or, like the definition of $t = 0$, by fixing an arbitrary calendar time after which the evolution of the epidemic is not of interest or relevant (e.g. usual end of seasonal activity).

3.1.5 Stochastic epidemic models

The models introduced in Subsections 3.1.2 to 3.1.4 only approximate the real dynamics of an epidemic. More realistically, epidemics are probabilistic phenomena formed by a series of discrete events, and, over the years, several different types of stochastic models have been formulated to describe them (Greenwood and Gordillo, 2009).

This subsection contains two formulations of stochastic epidemic models in continuous time and in discrete time, respectively.

Event-driven approach

Consider continuous time $t \in [0, T]$ with $t = 0$ being the beginning of an epidemic, *i.e.* the time at which an infectious individual is introduced, and $t = T$ being the extinction time of an epidemic, *i.e.* the time at which there are no more susceptible, exposed or infectious individuals. A stochastic epidemic model can be defined within the framework of state space models (SSMs) (Section 2.1.1), where the state process is represented by the number of people in each compartment at time t:

$$X(t) = \begin{pmatrix} S(t) \\ E(t) \\ I(t) \end{pmatrix}$$

and, when the population size N is known and fixed, $R(t)$ can be obtained by $R(t) = N - S(t) - E(t) - I(t)$.

Conditionally on $X(t) = x(t)$, and under Markovianity, the next event(s) follows a non-homogeneous Poisson process (Greenwood and Gordillo, 2009). Denote by u the time elapsing between two events. An infection event consists of the state transition to: $X(t+u) = (S(t) - 1, E(t) + 1, I(t))$ and happens at rate $\beta\frac{I(t)}{N}S(t)$; a transition out of the latent period consists of the state transition to: $X(t+u) = (S(t), E(t) - 1, I(t) + 1)$ and happens at rate $\sigma E(t)$; a recovery consists of the state transition to: $X(t+u) = (S(t), E(t), I(t) - 1)$ and happens at rate $\gamma I(t)$.

This model can be better understood in terms of the simulation method used to generate it, which was formulated in Gillespie (1976) for a general compartmental model, and applied to epidemic models in, e.g., Keeling and Rohani (2011) (Section 6.3).

Consider $X(t)$ as defined above, and denote the random variable (r.v.) of the time until the next event by U. Denote by $\alpha(t)$ the rate of any event (infection, end of latent state or recovery), conditionally on the state of the system at t, so that:

$$\alpha(t) = \beta\frac{I(t)}{N}S(t) + \sigma E(t) + \gamma I(t)$$

Given the Poisson-process properties (see Section 2.1.2), U is Exponentially-distributed with rate $\alpha(t)$, hence:

$$U \sim \text{Exp}(\alpha(t))$$

The type of event that happens at $t + s$ has probabilities equal to the contribution of each rate to $\alpha(t)$, *i.e.*:

$$X(t+u) = \begin{cases} S(t) - 1, E(t) + 1, I(t) & \text{w.p.} \quad \beta\frac{I(t)}{N}S(t)/\alpha(t) \\ S(t), E(t) - 1, I(t) + 1 & \text{w.p.} \quad \sigma E(t)/\alpha(t) \\ S(t), E(t), I(t) - 1 & \text{w.p.} \quad \gamma I(t)/\alpha(t) \end{cases}$$

This model can be expressed as a system of stochastic differential equations with the same structure as Equation 3.1, where the time and the compartment sizes are r.v.s.

Chain-Binomial approach

A discretised version of the event-driven model is the chain-Binomial model (Becker, 1989).

Consider again intervals of length $\delta \ll 1$ so that the t-th interval is $[\delta t, \delta t+\delta)$ and the intervals are now indexed by $t = 0, 1, 2 \ldots, T$. Denote by A_t the number of people becoming infected, by B_t the number of people becoming infectious and by C_t the number of people recovering during $[\delta t, \delta t + \delta)$ and define S_t, E_t and I_t and the rates β, σ and γ as in Subsection 3.1.4. Under approximation (ii), A_t, B_t and C_t distribute according to Binomial r.v.s with sizes S_t, E_t and I_t, respectively such that

$$\begin{aligned} A_t &\sim \text{Bin}\left(S_t, \beta\frac{I_t}{N}\delta\right) \\ B_t &\sim \text{Bin}(E_t, \sigma\delta) \\ C_t &\sim \text{Bin}(I_t, \gamma\delta) \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} S_{t+1} &= S_t - A_t \\ E_{t+1} &= E_t + A_t - B_t \\ I_{t+1} &= I_t + B_t - C_t \\ R_{t+1} &= R_t + C_t \end{aligned} \tag{3.4}$$

It is easy to see the Chain Binomial model as a τ-leap approximation of the event-driven model (see Appendix D.2).

This model, compared to the event-driven model, suffers from the same approximation errors as the discrete-time deterministic model compared to the continuous-time version; nevertheless it is still intensively used. As described in Section 3.2, inference in epidemic models is often simulation- or data-augmentation-based; the adoption of chain-Binomial models leads to a more efficient application of these simulation methods than the use of event-driven models and this fact has strongly motivated their use.

Moreover, the first stochastic epidemic models (the Greenwood and the Reed-Frost model) are special cases of the chain-Binomial model, as shown in Becker (1989).

Stochastic vs deterministic models

There are several advantages in replacing a deterministic model with a stochastic model (Vynnycky and White, 2010). Deterministic models are unrealistic as they allow the compartment sizes to take non-integer values. More importantly, they are not able to reproduce the differences in epidemics that can be observed in reality: in a small population and considering the same infection, latency, and recovery parameters, the introduction of an infectious individual may cause the occurrence of a major epidemic or likewise, he can recover affecting none (or almost none).

The outcome of the simulation of a stochastic model can be very different from a deterministic one, as shown in Figure 3.6. Here a deterministic discrete-time epidemic model (dashed) is compared with 1000 simulations from a chain-Binomial model. The left column compares the deterministic discrete-time approximation of the compartment sizes to those of 20 of the 1000 simulations; the right column reports the histograms of the total new infections generated in $[0, T]$ in the 1000 chain-Binomial simulations. The set of parameters used is: $\beta = 0.40, \sigma = 0.8, \gamma = 0.2$. The first row assumes the introduction of 1 infectious and 1 exposed individual in a population of 998 individuals ($N = 1000$). The histogram of the number of individuals infected in $[0, T]$ is bimodal: some epidemics are minor, and the others are distributed around the final size of the deterministic epidemic (Figure 3.6a). In the left-column plots, the minor epidemics can be identified by those trajectories whose number of susceptibles at T is almost equal to the number of susceptibles at 0. When the same parameters are used and the population at time 0 is scaled up by 10 and by 100 ($S_0 = 9980, E_0 = 10, I_0 = 10, N = 10000$ in Figure 3.6b and $S_0 = 99800, E_0 = 100, I_0 = 100, N = 100000$ in Figure 3.6c, respectively), the bi-modality disappears and the final size is distributed around the final size of the deterministic model. The simulations from the chain-Binomial model are better approximated with a higher population scale (Figure 3.6c). Lastly, if only the total population is scaled up ($N = 10000$) and 1 infectious and 1 exposed are introduced, the bi modality arises again.

This highlights that the approximation of a stochastic model by a deterministic model is more reliable after the epidemic has taken off (when the number of infectious individuals is already high) in large populations. In contrast, when one or few infectious individuals enter a small population, a deterministic model does not approximate the dynamics well.

Further results on the final size of stochastic epidemic models can be found in Britton (2010).

Despite these clear advantages, the question of which model to use is case specific and an a-priori decision is never straightforward. Explicit comparison between the fit of a deterministic model and a stochastic model to the data should help in the decision (Rohani, 2016).

3.1.6 Parameters and transmission measures

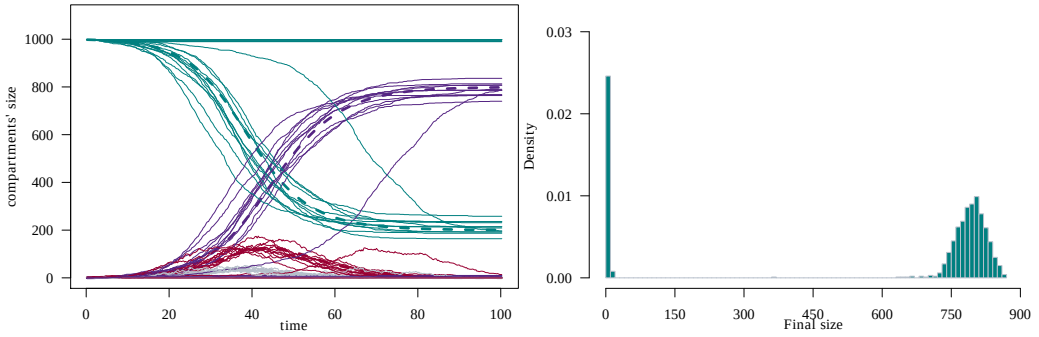
The previous sections showed different ways to express and model the temporal dynamics of an epidemic. In addition to these, an epidemic can be characterized by summary measures that quantify the whole outbreak.

The most common index to quantify transmission is the basic **reproduction number**, denoted by R_0 (Macdonald, 1952). R_0 expresses the average number of new infections generated by an infectious individual when introduced in a totally susceptible population. It is related to the transmission rate β and to the average infectious period $1/\gamma$ via Equation 3.5 (Vynnycky and White, 2010):

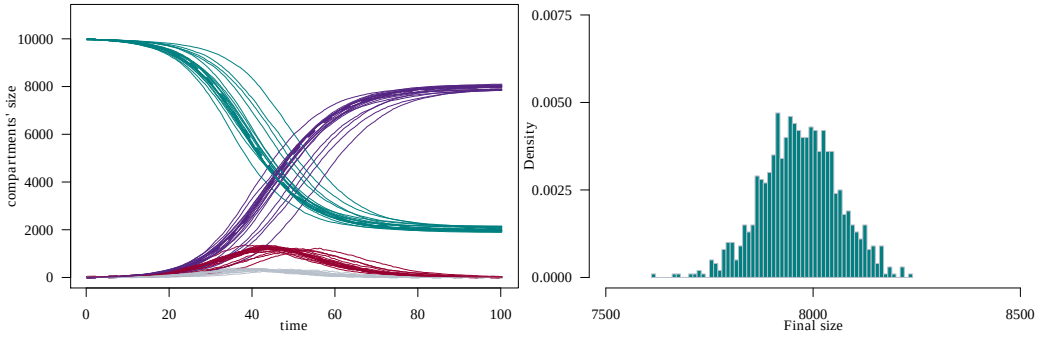
$$R_0 = \beta \frac{1}{\gamma} \tag{3.5}$$

If instead the population is only partially susceptible, due to an initial immune proportion, transmission can be measured by the effective reproduction number at time $t = 0$:

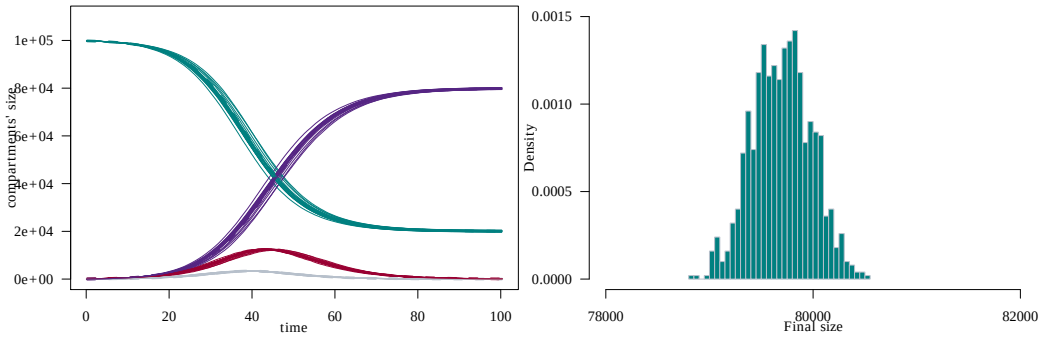
$$R_e(0) = \beta \frac{1}{\gamma} \frac{S(0)}{N} = R_0 \frac{S(0)}{N}$$



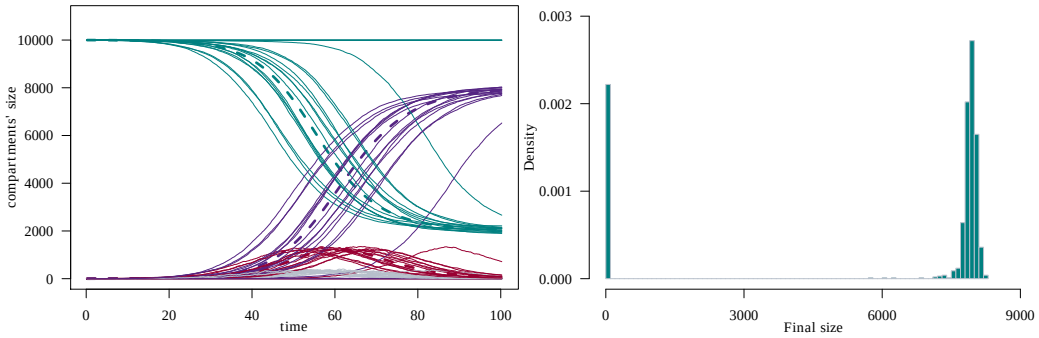
(a) $N = 1,000$ and 0.2% infected individuals: $E_0 = I_0 = 1$



(b) $N = 10,000$ and 0.2% infected individuals: $E_0 = I_0 = 10$



(c) $N = 100,000$ and 0.2% infected individuals: $E_0 = I_0 = 100$



(d) $N = 10,000$ and 0.02% infected individuals: $E_0 = I_0 = 1$

Figure 3.6: Left column: Number of susceptible (green), latent (grey), infectious (red) and removed (purple) people at time t. The dashed line is the discrete-time deterministic model approximation and the solid lines are 20 realizations of a chain-Binomial model with the same parameters. Right column: histogram of the final size of 1000 simulated epidemics.

As an epidemic progresses, fewer people are susceptible and the effective reproduction number changes accordingly:

$$R_e(t) = R_0 \frac{S(t)}{N}$$

$R_e(0)$ is extremely informative because it provides an indication of whether or not a major epidemic is going to take place.

In a deterministic epidemic model, if $R_e(0) < 1$ the epidemic is not going to spread and will fade away soon after its start; the case of $R_e(0) = 1$ (not particularly relevant), consists of the epidemic persisting in the population with a constant number of infectious individuals; if instead $R_e(0) > 1$, the epidemic will take off.

In a stochastic epidemic model instead, if $R_e(0) \leq 1$ the probability of a major epidemic is 0 and if $R_e(0) > 1$ there is a positive probability of major epidemics.

Another quantity of interest is the **final size** of an epidemic, defined as the number of people that have been infected during $[0, T]$, *i.e.* $S(T) - S(0)$. This is useful to compute the **total attack rate**, *i.e.* the proportion of the population that has been infected over the course of the epidemic and denoted by TAR:

$$TAR = \frac{S(T) - S(0)}{N}$$

Other transformations of the parameters of an epidemic models might be useful, not only to characterise the epidemic, but also to better explain particular behaviours of the system or reduce the dimension of the parameter space. This is the case of the **exponential growth rate**, a parameter that might be introduced to model the initial growth of the number of infected and infectious compartments. They are known to increase exponentially (Keeling and Rohani, 2011) with rate ψ:

$$\frac{dE}{dt} = \psi \cdot E$$
$$\frac{dI}{dt} = \psi \cdot I$$

Wearing, Rohani, and Keeling (2005) derived the re-parametrization of the initial state (i.e. $S(0), E(0), I(0)$ and $R(0)$) of the epidemic and of the transmission parameters that results from this assumption. Their results, applied to the model considered in Section 3.3, are reported below.

3.1.7 Observation process

Since an epidemic model can be viewed as a SSM, an appropriate observational model should be defined to link the underlying transmission process $X(t)$ with $t \in [0, T]$ (or X_t with $t = 0, 1, 2, \ldots, T$) to one (or, possibly, multiple) time series of observable quantities $Y_t, t = 0, 1, 2, \ldots T$.

Once an individual is infected he might experience more or less severe symptoms and he might seek health assistance. The process that describes this phenomenon in the infected population is the *severity process* that has been widely discussed in Chapter 2. Severe events might occur after some time from infection, and this can be described via an *observational delay process*. Only a portion of severe/symptomatic individuals might be detected and registered in the data $Y_t, t = 0, 1, 2, \ldots T$; this aspect is accounted for in the *detection process*.

Despite most aspects of the observational process having been treated in Chapter 2, they are recalled here in the context and within the language of epidemic models.

The observation process is a process that happens alongside to the transmission process and there is no one-to-one relation between the two processes. An individual might be asymptomatic but already infectious (as happens, for example in HIV); he can be in the R state as a result of a quarantine policy, while still being at a severe stage of the disease; or, as happens for the Ebola virus, he can reach the most severe state, death, and still be counted in the I compartment due to infectious contacts during burial ceremonies.

Severity process

Similarly to transmission, severity aspects can also be treated either using a deterministic or a stochastic approach and with several levels in between these two extremes. Denote by θ the probability of experiencing a severe event (e.g., the probability of symptoms given infection, or the probability of seeking health assistance at GP clinics, or the case hospitalization risk). Conditionally on the transmission evolution, the number of incident severe cases over time could be described, for example, by a Binomial sample from the incident infections with probability θ. However stochasticity can be ignored and a simple approximation to the average number of severe cases can be made. For example, define ΔI_t^{S}, the number of individuals that are infected during the interval $(t\delta, t\delta + \delta]$ and will eventually become severe (S) cases, and approximate it by:

$$\Delta I_t^{\text{S}} = (S(t\delta) - S(t\delta + \delta)) \cdot \theta$$

Systematic discussion on the level of stochasticity introduced in a severity process and its approximation are given in Chapter 5.

Observational delays

The delay between infection and severe events can also be modelled with more or less stochasticity. For individuals that develop severe symptoms, the r.v. D denotes the time elapsing between infection and the occurrence of severe symptoms and detection. D might be approximated by a parametric distribution with density $f(d; \vartheta)$. For routinely collected data, the discrete version of this distribution is considered, with $f_d(\vartheta)$ being the probability that the delay experienced is in the dth interval of length δ, $[\delta d; \delta d + \delta)$.

Once a delay distribution has been defined, the time series of severe infections ΔI_t^{S}, $(t = 1, 2, \dots T)$ can be linked to the time series of the severe cases at the time of their diagnosis, denoted by ΔC_t^{S}, via convolution:

$$\Delta C_t^{\text{S}} = \sum_{d=0}^{t} \Delta I_{t-d}^{\text{S}} f_d(\vartheta) \tag{3.6}$$

More often, Equation 3.6 denotes the average behaviour of the observed severe cases, with this average denoted by μ_t^{S}:

$$\mu_t^{\text{S}} = \mathbb{E}(\Delta C_t^{\text{S}}) = \sum_{d=0}^{t} \Delta I_{t-d}^{\text{S}} f_d(\vartheta)$$

The observations are then distributed randomly around a function of μ_t^{S} (Brookmeyer and Gail, 1994) as explained below.

Detection and measurement error

The last elements of the observation process to define is the detection process, *i.e.* the process through which severe cases are included in the surveillance system.

A choice is to be made about the distribution of the observations. Common choices are Binomial (Shubin et al., 2016), Poisson (Birrell et al., 2012; Ong et al., 2010) or Negative Binomial (Birrell et al., 2011; Tuite and Fisman, 2018) distributions, the latter allowing more flexibility through an over-dispersion parameter.

The data-distributional assumption, in the case of deterministic transmission, severity and delay models, allows the likelihood of a time series of independent observations to be written down analytically. When transmission is stochastic and, more pertinently, severity is modelled stochastically, a dependence across time and states arises. This dependence will be addressed in Chapter 5.

Remarks on the observation process

In the previous paragraphs, many of the possible aspects of an observational process have been highlighted. Nevertheless, in reality they are often difficult to disentangle.

The data available are a corrupted and delayed signal of the underlying transmission process but the amount of corruption imputable to one aspect of the severity process (for example, the probability of symptoms given infection), or to another (for example, the probability of hospitalization given severe symptoms), or to a detection element (for example the probability of a hospitalization being reported) is indistinguishable.

This is one of the challenges mentioned in the opening of this Chapter (Becker, 1989) and the availability of prior information on the observation process, together with evidence synthesis methods, might enable the inference of both transmission and severity/observation dynamics (Birrell, De Angelis, and Presanis, 2018; De Angelis and Presanis, 2018).

3.1.8 Extensions and other modelling schemes

In order to more realistically model epidemic data, a number of extensions to the simple epidemic models presented can be made.

A very useful extension consists of allowing *time-varying* transition rates. The infection rate $\beta(t)$ can be thought to vary over time due to several different factors. Many infectious diseases show a seasonal oscillation that is believed to depend on the fluctuation of weather conditions (Lowen et al., 2007), which might inflate or deflate the magnitude of transmission (Dushoff et al., 2004). Moreover, the rate of contact within some population groups might depend highly on calendar time and hence might affect the rate of new infections. An example of this can be found in the influenza-modelling literature that has been shown to depend highly on the opening or closure of schools (House et al., 2011; Te Beest et al., 2015; Vynnycky and Edmunds, 2008). Moreover, time-varying infection rates of SIR-type models can be modelled using structured time-series, leading to the so called T-SIR model (Bjornstad, Finkenstadt, and Grenfell, 2002; Finkenstädt and Grenfell, 2000).

Another aspect that has intrigued modellers is the *Markovianity assumption*. This implies that the average time spent in each compartment distributes according to an Exponential r.v..

Some work on non-Markovian models can be found in Streftaris and Gibson (2002). Nevertheless, there exists a simpler way to allow a more flexible distribution for the time between events, while still preserving a Markovian model. A compartment (e.g. E) can be divided in multiple sub-compartments ($E_1, E_2, \ldots, E_k$); if the time spent in each sub-compartment is Exponentially distributed, then the total time spent in the E compartment is distributed according to a Gamma r.v. which is more flexible than an Exponential since it allows waiting time with modes different than 0 as well as rates varying over time-to-compartment-exit.

3.2 Estimating transmission parameters

Methods for the estimation of transmission in epidemic models are reviewed in this section, in terms of both statistical inference (Section 3.2.1) and applications to influenza (Section 3.3.2)

3.2.1 Review of estimation methods

Different inferential challenges are posed by different types of data, depending on their characteristics.

A type of data that is not treated in this thesis is *final outcome data*. They usually consist of the total number of cases generated within an epidemic. Although final outcome data do not contain information on the temporal evolution on the disease, they are very useful for the estimation of quantities such as R_0, according to the mathematical relationship between final size and R_0. For a concise reference see Britton (2010) and for proofs and derivations see Andersson and Britton (2012).

Temporal data instead, consist of time-series referring either to the number of people transiting from one state to another during an interval (incidence data) or to the number of people in a compartment at a specific time (prevalence data).

Methods for inference from temporal data change according to the assumptions of the transmission and observational models. The transmission process can be modelled in continuous or discrete time, and as deterministic or stochastic; the severity process is often approximated by its deterministic formulation and a specific discussion on this approximation is postponed to Chapter 5; finally, the detection and measurement process are often assumed to follow a parametric distribution that describes the noise in the data.

A review (O'Neill, 2010) lists the most common methods that relate transmission models to data. Here a similar, but briefer, attempt is made, which includes more recent development but without providing a comprehensive enumeration of all the possible approaches.

When a *deterministic transmission model* is assumed, the average incident/prevalent cases, incident/prevalent removals, etc are a deterministic function of the transition rates. The deterministic model can be used to predict the expected observations, while the data distribution determines the likelihood of the transmission and observational parameters. Alternatively optimization of some error function can be performed so that the parameters that provide trajectories that are closest to the data (e.g. that minimize the least squared error) are selected. This method is sometimes referred to as *trajectory matching* (Riley et al., 2003).

Nevertheless, inference might be challenging even in these cases where the amount of stochasticity is low. This is because often only data on one compartment or one specific transition are

available, and information on the initial state of the system is often absent. In this setting of partial data, the likelihood of the parameters is frequently affected by identifiability issues (Gustafson, 2010) as is shown in the following chapter.

By dropping the assumption of deterministic dynamics and dealing with *stochastic transmission models*, inference becomes even more demanding. In this case, the number of people within each compartment and moving between compartments can be described, in the SSM setting, as the state process, of which only a noisy time series is observed. In this context, inference can be performed using one of the following principles.

- *Martingale methods*, such as the ones developed in Becker (1993), are based on the formulation of ad-hoc martingales for the counting processes of the transmission systems. However the high price of complex mathematical formulation is not offset by much gain since these methods are rarely applicable as they rely on very strict hypothesis (O'Neill, 2010).

- *Likelihood free methods* have been recently developed, exploiting the availability of powerful computational resources and the fact that epidemic models are very easy to simulate. These methods are usually developed within the Bayesian literature and they might fall in the wider family of Approximate Bayesian Computation (ABC)(Kypraios, Neal, and Prangle, 2017) or they might be Markov chain Monte Carlo (MCMC) algorithms specifically tailored for epidemic data (McKinley et al., 2014).

- *Data augmentation methods* consider the problem of inference within a perspective of missing data, where the underlying transmission process is also to be estimated. These methods were developed in the context of continuous-time stochastic epidemic models where the infection times are unknown since observations are at discrete times and affected by delays (e.g. incubation time). These methods, combined with the use of MCMC methods, have been intensively used in the past 25 years (O'Neill and Roberts, 1999).

- *Filtering methods* are general methods for SSMs from the family of sequential Monte Carlo (SMC) methods. These methods have been developed both in a frequentist (Ionides, Bretó, and King, 2006) and in Bayesian frameworks (Andrieu, Doucet, and Holenstein, 2010) and they are based on sequential integration of the state process. Their use in the estimation of transmission models is becoming more and more frequent (e.g. Birrell et al. (2017) and Magpantay et al. (2015)).

3.2.2 Review of applications of epidemic models to influenza

This section reviews some studies on the transmission of influenza. The literature on this topic is copious even when limited to only those studies that use the transmission models described in Section 3.1.

There are many criteria for classifying the models, e.g. discriminating according to their time-scale or levels of stochasticity, the data used and the epidemic analysed or the inferential methodology applied. In this section, the focus is on the latter since the evolution of statistical methodology has enabled more advanced analysis of infectious disease data. Within each section,

notable works in the field are highlighted, and for each work the epidemic model used, the data used and the novelty of the approach are pinpointed.

The first applications of epidemic models to influenza used to consider the process as being fully observed and therefore explicitly write and maximise the likelihood. An example of this is Longini (1986), who uses a discrete time epidemic model (specifically, a chain-Binomial model) to fit data on influenza spreading on small and large populations. The former are data on households in the United States, the latter are time series of cases in England and Wales.

Other examples of a likelihood-based methods for influenza data are the ones that use trajectory matching, such as Wearing, Rohani, and Keeling (2005), where SIR and SEIR models are fitted to data on an epidemic of influenza in a boarding school (Anon., 1978). Despite the problem being relatively simple (with a small population and small observational noise) the paper explores interesting aspects on the choice of the compartments of the epidemic model and on the initial evolution of an epidemic. Similarly, Hall et al. (2007) fits, via maximum likelihood, a deterministic model with Normal noise, to data from a historical flu epidemic, studying the potential for prediction from early epidemic data. A more modern example of fitting transmission models via maximum likelihood can be found in Yaari et al. (2016), which exploits both clinical and environmental data to gain information on the 2009 pandemic in Israel.

The advent of Bayesian reasoning revolutionised the inference of influenza transmission with epidemic models because, on the one hand, it offers a simple way to integrate previous knowledge with data and, on the other hand, MCMC provides a way to sample from posterior distributions that often overcomes problems of complicated, multi-parameter likelihoods. An example of this analysis can be found in (Birrell et al., 2011) that analyses data on General Practitioner (GP) consultations from the 2009 pandemic fitting a deterministic discrete-time epidemic model via MCMC. Dorigatti et al. (2012), Dorigatti, Cauchemez, and Ferguson, 2013 and Te Beest et al. (2015) perform a similar analysis using a continuous-time model for data from the UK, Italy and the Netherlands respectively. Likewise, Baguelin et al. (2013) fits a multi-strain deterministic epidemic model to data on GP consultations for 15 seasons of influenza in England aimed at informing evaluation policies. In Merl et al. (2009) similar methods are used sequentially on the aforementioned boarding school data, assuming a stochastic epidemic model with a focus on intervention evaluation.

Recently, with the advent of more powerful computers, methods heavily based on simulation are increasingly used to model infectious disease data, and, among them influenza data. Filtering/SMC methods are proliferating in this field: an early example is Ong et al. (2010) that fits via SMC a chain-Binomial model to GP-consultation data during the 2009 pandemic in Singapore. This study was followed up by a similar analysis of data from Malta (Marmara, Cook, and Kleczkowski, 2014). Similarly, Dukic, Lopes, and Polson (2012) uses a sequential learning algorithm to estimate transmission parameters and states dynamics of a stochastic transmission model from Google trends data. Ensemble methods are filtering methods based on the Kalman filtering approximation (Katzfuss, Stroud, and Wikle, 2016), and they have been widely used, coupled with discrete time transmission models, on observed and predicted GP consultation data (Shaman et al., 2013) and on Google Flu Trend data (Yang, Lipsitch, and Shaman, 2015). Other works that have applied SMC methods are Shubin et al. (2016), that fitted a discrete-time SIR model to hospitalization, death and GP data from the 2009 pandemic

in Finland, and Birrell et al. (2017) that enhanced classical SMC methods to model pandemic data with shocks, using the same data of Birrell et al. (2011). Yang, Karspeck, and Shaman (2014) reviews and compares some of these SMC methods on Google Flu Trend data and GP consultations during seven influenza seasons in 115 cities in the US.

There is also a moderate literature on review papers. Biggerstaff et al. (2014) collects and compares all the estimates of R_0 available in the literature. Chretien et al. (2014) instead reports and comments on papers aimed at prediction of future influenza dynamics.

3.3 An epidemic model to estimate transmission from severe case data

This Section, studying three seasons of influenza in England, has been published in Corbella et al. (2018). This analysis provides an example application of epidemic models to real influenza data, illustrating the challenges of such an analysis.

This study fits in the literature presented above by providing a simple way to exploit readily available data to infer seasonal influenza incidence. Most of the papers cited investigate data such as GP consultations for influenza-like illness (ILI) (Baguelin et al., 2013; Birrell et al., 2011) or health-related online queries (Yang, Lipsitch, and Shaman, 2015), which are only loosely related to the actual burden and are characterized by highly volatile noise. By contrast, more specific timely data on a sample of confirmed cases (e.g. confirmed influenza hospitalizations) might be collected routinely by national health systems. An example of these data is the UK Severe Influenza Surveillance System (USISS) (Public Health England, 2014) that records counts of the weekly Intensive Care Unit (ICU) and High Dependence Unit (HDU) admissions and deaths with confirmed influenza in all hospital trusts in England.

Only Shubin et al. (2016), to the knowledge of the author, has attempted to estimate pandemic influenza transmission from routinely collected confirmed-case data. Nevertheless, his work has entailed the development of a highly complex model which is difficult to use in a seasonal monitoring setting with a prediction goal; when less effort is placed on data collection.

Conversely, here a much simpler model is proposed and applied to seasonal and pandemic influenza, relying on severe cases data alone, which are promptly available.

In Section 3.3.1 the data used are described. In the following section, transmission, severity and detection models are chosen. In Section 3.3.3 the methods used to draw inference are stated. Results are presented in Section 3.3.4 and discussed in Section 3.3.5.

3.3.1 Data

Following the 2009 pandemic, the World Health Organization (WHO) declared the beginning of a post-pandemic phase (WHO, 2010), encouraging national public health agencies to establish hospital-based surveillance systems to monitor the epidemiology of severe influenza. In response to these guidelines, and to understand the baseline epidemiology of severe influenza, the UK developed a surveillance system to monitor severe cases of influenza, the USISS. After a pilot phase in 2010/11, USISS has run for each influenza season, providing data on laboratory-confirmed ICU/HDU influenza cases and on laboratory-confirmed hospitalized cases.

According to the USISS protocol (Health Protection Agency, 2011a), all National Health Service (NHS) trusts should report the weekly number of laboratory-confirmed influenza cases admitted to ICU/HDU and the number of confirmed influenza deaths in ICU/HDU via a web tool. An ICU/HDU case is defined as a person who is admitted to ICU/HDU and has a laboratory-confirmed influenza A (including H1, H3 or novel) or B infection.

USISS runs annually from week 40 to week 20 of the following year but, in the event of a pandemic, it can be activated out of this window and will collect the same data at all levels of care, not only ICU/HDU.

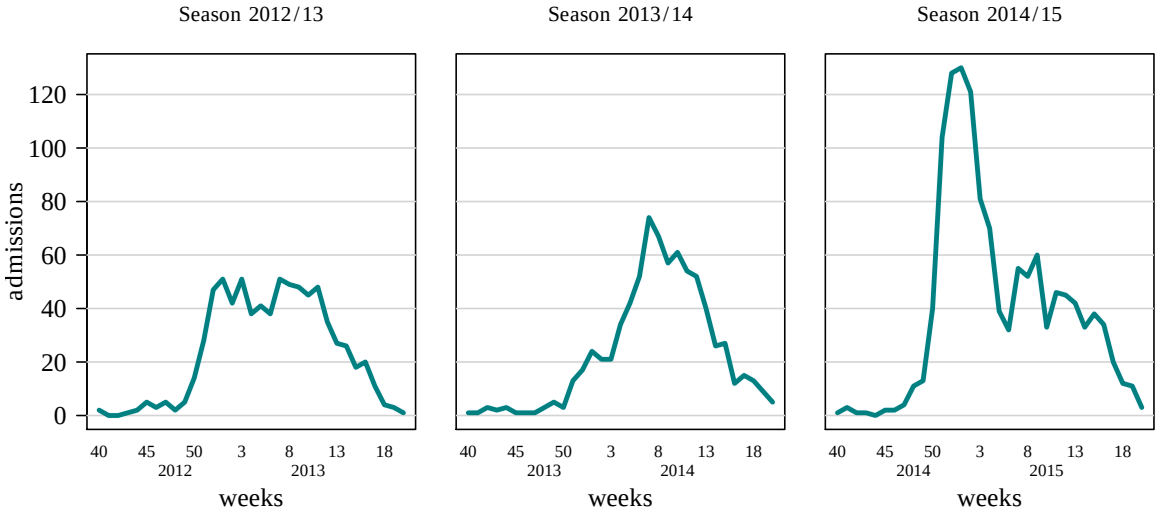


Figure 3.7: Weekly ICU/HDU admissions by season, time is measured in week number as reported on the x axis.

Data are available by age group and influenza type/subtype. However, when stratified by both, as well as week, many zero counts are observed. The total ICU/HDU admissions by week only are displayed in Figure 3.7. Each season between 2012 and 2015 is shown, with each epidemic varying substantially across seasons. In the 2012/13 season, mainly characterized by influenza B and influenza A(H3N2) outbreaks, the number of admissions peaks early, maintaining this plateau for several months (Public Health England, 2013). In 2013/14, when the predominant strain was A(H1N1), the time series displays a smoother increase, a well localized peak and a subsequent regular decrease (Public Health England, 2014). Lastly, in 2014/15, the number of ICU admissions peaks earlier and has a dramatic drop at the beginning of the new year, which is followed by a smaller wave resulting in a time series characterized by a double peak. During this season, influenza A(H3N2) was the predominant virus circulating and the total number of ICU admissions was higher; this strain is well-known to lead to more severe outcomes, particularly in the elderly (Public Health England, 2015).

In addition to the mandatory scheme, a subgroup of NHS trusts in England is recruited every year to participate in the USISS sentinel scheme (Boddington et al., 2017; Health Protection Agency, 2011b), which reports weekly numbers of laboratory-confirmed influenza cases hospitalised at all levels of care. From this scheme, individual-level data on all ICU/HDU admissions (until season 2012/13) or on hospital admissions in the young ($\leq$ 17 years old) population (from season 2013/14 onwards) are available, including clinical details such as date of symptom onset,

of hospital and ICU admission, and date of discharge from ICU.

These data provide useful information on the process between influenza infection and ICU admission (e.g. the time elapsing from symptom onset to ICU admission). Further information on this process (e.g. proportion of symptomatic cases) can be found in the existing literature about the incubation period of influenza (Tom et al., 2011) and the hospitalization fatality rate (Presanis et al., 2014).

3.3.2 Model

A deterministic, continuous-time epidemic model is chosen to model the spread of influenza in England.

The population is divided according to health status into: susceptible (S), exposed (E), infectious (I) and removed (R) compartments. The E and I compartments are further divided into two (E_1, E_2 and I_1, I_2, respectively) so that the waiting times in the E and I states are distributed according to Gamma rather than Exponential distributions (Wearing, Rohani, and Keeling, 2005). The total size of the population is fixed over every season and denoted by $N \approx 50,000,000$. The movements between compartments are determined by the transition rates: $\lambda(t)$, σ and γ explained below.

The infection rate $\lambda(t)$ is proportional to the fraction of people in the infectious compartment at t, $\frac{I_1(t)+I_2(t)}{N}$ and a time varying transmission rate $\beta(t)$:

$$\lambda(t) = \beta(t)\frac{I_1(t) + I_2(t)}{N}.$$

$\beta(t)$ is a function of time and it allows for a scaling factor $\kappa \in (0, 2]$ that expresses the change due to school closure, applied to the transmission rate during school opening β_0 Te Beest et al., 2015 as reported in Equation 3.7:

$$\beta(t) = \begin{cases} \kappa \cdot \beta_0, & t \in \text{school holidays} \\ \beta_0, & \text{otherwise.} \end{cases} \tag{3.7}$$

The transition rates σ and γ are related to the mean latent period, d_L, and the mean infectious period, d_I, by:

$$\sigma = 2/d_L, \qquad \gamma = 2/d_I$$

The system of differential equations that defines the epidemic model is reported in Equation

3.8:

$$\begin{aligned}
\frac{dS}{dt} &= -\lambda(t) \cdot S \\
\frac{dE_1}{dt} &= \lambda(t) \cdot S - \sigma \cdot E_1 \\
\frac{dE_2}{dt} &= \sigma \cdot E_1 - \sigma \cdot E_2 \\
\frac{dI_1}{dt} &= \sigma \cdot E_2 - \lambda \cdot I_1 \\
\frac{dI_2}{dt} &= \lambda \cdot I_1 - \lambda \cdot I_2 \\
\frac{dR}{dt} &= \lambda \cdot I_2
\end{aligned} \tag{3.8}$$

Exponential growth rate

The model is parametrised with the transition rate parameters (*i.e.* $\beta, \kappa, \sigma, \gamma$) and the initial state of the epidemic (*i.e.* $S(0), E_1(0), E_2(0), I_1(0), I_2(0), R(0)$).

The epidemic model can be re-parametrized assuming that the E and I compartments grow Exponentially with rate ψ (Vynnycky and White, 2010; Wearing, Rohani, and Keeling, 2005):

$$\begin{aligned}
\frac{dE_1}{dt} &= \psi \cdot E_1 \\
\frac{dE_2}{dt} &= \psi \cdot E_2 \\
\frac{dI_1}{dt} &= \psi \cdot I_1 \\
\frac{dI_2}{dt} &= \psi \cdot I_2
\end{aligned}$$

and, it is possible to equate each line of this system to the respective line of Equation 3.8 at t=0:

$$\begin{aligned}
\lambda(0) \cdot S(0) - \sigma \cdot E_1(0) &= \psi \cdot E_1(0) \\
\sigma \cdot E_1(0) - \sigma \cdot E_2(0) &= \psi \cdot E_2(0) \\
\sigma \cdot E_2(0) - \lambda \cdot I_1(0) &= \psi \cdot I_1(0) \\
\lambda \cdot I_1(0) - \lambda \cdot I_2(0) &= \psi \cdot I_2(0)
\end{aligned}$$

To solve this system, denote by I_0^{tot} the total number of infectious individuals at $t = 0$ (*i.e.* $I_0^{\text{tot}} = I_1(0) + I_2(0)$). The solution of the system is reported in Equation 3.9, where the first and last line are derived by the assumption of a closed population and the definition of initial

immunity $(1-\pi)$, respectively.

$$\begin{aligned} S(0) &= N - I_1(0) - I_2(0) - E_1(0) - E_2(0) - R(0) \\ I_1(0) &= I_0^{\text{tot}} \frac{1}{1+\frac{\gamma}{\gamma+\psi}} \\ I_2(0) &= I_0^{\text{tot}} - I_1(0) \\ E_2(0) &= I_1(0)\left(\frac{\gamma+\psi}{\sigma}\right) \\ E_1(0) &= E_2(0)\left(\frac{\sigma+\psi}{\sigma}\right) \\ R(0) &= (1-\pi)N \end{aligned} \tag{3.9}$$

This result was derived in Wearing, Rohani, and Keeling, 2005 and further used in Birrell et al., 2011 to model the spread of influenza in the UK. For this reason, it seems a sensible re-parametrisation for the data we are analysing to reduce the number of parameters of our system: the number of people in each state at $t = 0$ $(S(0), E_1(0), E_2(0), I_1(0), I_2(0), R(0))$ is now replaced by functions of $(N, \psi, I_0^{\text{tot}}, \pi)$ and (γ, σ), already parameters of (3.8).

Wearing, Rohani, and Keeling, 2005 proved that, under this parametrisation, the basic reproduction number R_0 can be expressed as a function of the rate ψ, of the average infectious period d_I, and of the average incubation period d_L:

$$R_0 = \psi \cdot d_I \cdot \left\{ \frac{\left(\frac{\psi d_L}{2}+1\right)^2}{1-\left(\frac{\psi d_I}{2}+1\right)^{-2}} \right\}$$

A further re-parametrization is given by defining a new parameter $\lambda_0 = \lambda(0)$, the hazard of infection at the beginning of the epidemic:

$$\lambda_0 = \beta(0) \cdot I_0^{\text{tot}}$$

which leads to a further definition of I_0^{tot}:

$$I_0^{\text{tot}} = \frac{d_I \lambda(0) N}{R_0} \tag{3.10}$$

Observational model

This transmission model is linked to the data on ICU admissions through an observational model that defines the time elapsing from infection to ICU admission and the probability of ICU admission conditional on infection.

Denote with $f_w^{\text{ICU|I}}$ the probability that w weeks elapse from infection to ICU admission, and with p_{ICU} the probability of ICU admission given infection. μ_w, the average number of ICU admissions during week w, can be related to the weekly new infections in the previous weeks via the following convolution:

$$\mu_w = \sum_{v=0}^{w} f_{w-v}^{\text{ICU|I}} \cdot \Delta I_v p_{ICU} \tag{3.11}$$

where $\Delta I_w = (S(w-7) - S(w)) \cdot N$ is the count of the new infections during week w.

To formulate the likelihood of the data, the observed number of ICU admissions is assumed to follow a Negative Binomial distribution centred on μ_w with over dispersion parameter η:

$$Y_w^{ICU} \sim \text{NegBin}(\mu_w, \eta),$$

i.e Y_w^{ICU} has density function:

$$f(Y_w^{ICU} = y) = \frac{\Gamma(y + r_w)}{\Gamma(y)\Gamma(y+1)} \left(\frac{1}{\eta}\right)^{r_w} \left(1 - \frac{1}{\eta}\right)^{y}$$

with $r_w = \frac{\mu_w}{\eta - 1}$.

Appendix C contains the full derivation of $f_w^{ICU|I}$ from the survival data.

3.3.3 Inference

Due to the exponential growth rate parametrization, the transmission model is defined by the parameter vector $\boldsymbol{\theta} = (\lambda_0, \psi, \pi, \sigma, \gamma, \kappa)$.

The parameters σ and γ are assumed known from previous studies (Birrell et al., 2011; Tom et al., 2011), as they can be inferred only with detailed information at the individual level. Likewise, the population size N is assumed known and fixed to the values estimated by the Office of National Statistics (ONS) (Office of National Statistics, 2012-2015).

Prior distribution

Inference for the remaining parameters is performed under two prior scenarios. In both scenarios prior distributions are set on the parameter $\boldsymbol{\theta}$, however, most interpretable parameters (e.g. R_0 and β) are reported in the table: limits have been set for these quantities and the transformations listed in the previous section have been used to derive the priors on $\boldsymbol{\theta}$.

In the first scenario, no prior information on the values of the parameters except for lower and upper bounds is known, hence the prior distributions on all the parameters are flat. Table 3.1 lists the lower and upper limits of some transformations of the parameters and the values assumed known in this scenario.

In the second scenario, a prior distribution for the initial susceptibility π is formulated using sero-prevalence data from the 2010/11 season (Hoschler et al., 2012). The use of sero-prevalence data to describe the immunity of a population could be debatable, since the results may be generalisable only to seasons with similar predominant strains circulating. Here, sero-samples were taken during an H1 predominant season: this sub-type was prevalent also in the 2012/13 season, but not in 2014/15. However, combining this prior with the data allows testing of how much prior knowledge is needed to overcome the lack of information about susceptibility from the data. An informative prior distribution on p_{ICU} is derived by combining estimates of the probability of hospitalization given infection from a previous severity study Presanis et al., 2014 with estimates of the probability of ICU/HDU admission given hospitalization from the aggregate data of the USISS sentinel scheme. Table 3.2 lists the prior distributions of the two parameters that change in the informative scenario. The remaining parameters are again assumed to be uniformly distributed.

Unknown parameters		Lower limit	Upper limit
Susceptibility	π	0	1
Initial number of infectives	$I^{tot}(0) = (I_1(0) + I_2(0))$	0	10000
Transmission rate	β	0	1.12
Over-dispersion	η	1	100
P of ICU admission given infection	p_{ICU}	0	1
Scaling factor for school closure	κ	0	2
Parameters assumed known		**Value**	
Rate of becoming infectious	σ	1	
Rate of recovery	γ	0.5797	
Population of 2012/13	$N_{2012/13}$	53,679,750	
Population of 2013/14	$N_{2013/14}$	54,091,200	
Population of 2014/15	$N_{2014/15}$	54,551,450	

Table 3.1: Prior distributions of the parameters in the flat scenario.

Parameters		Distribution
Susceptibility	π	$\sim \text{LogNorm}(\log \mu = \log(0.401), \log \sigma = 0.2)$
P of ICU admission given infection	p_{ICU}	$\sim \text{LogNorm}(\log \mu = \log(0.000239), \log \sigma = 1)$

Table 3.2: Prior distributions of the parameters that change in the informative scenario.

Inference

For both the prior settings, two types of analysis are carried out: firstly, all the data reported in Figure 3.7 were considered and analysed retrospectively. Secondly, to assess the predictive ability of the model, estimation and forecasting is performed assuming only an initial portion of the data is available. Data up to week w are used as a training dataset to estimate the parameters. Then, the evolution of the epidemic after week w is predicted, based on the estimates from the training dataset. The following prediction time points were tested: $w = 3, 8, 13$, and 18 from the beginning of the new year.

A Metropolis Hastings (MH) block updated sampling algorithm (Robert, Casella, and Casella, 2010) was formulated to sample from the posterior distribution of $\boldsymbol{\theta} = (\pi, \lambda_0, \psi, \eta, p_{ICU}, \kappa)$. Values are proposed by sampling from a truncated Log-Normal random walk. The elements of $\boldsymbol{\theta}$ were very correlated: this was diagnosed while running a preliminary unblocked MH algorithm on each element of the vector $\boldsymbol{\theta}$, conditional on the others. $\boldsymbol{\Sigma}$, the variance covariance matrix of the parameter vector, was estimated by $\widehat{\boldsymbol{\Sigma}}$, the observed variance covariance matrix of the sampled values of $\boldsymbol{\theta}$.

This matrix is used to formulate a blocked MH algorithm that allows the joint sampling of all the elements of the vector $\boldsymbol{\theta}$ (Sherlock, Fearnhead, and Roberts, 2010). In the blocked MH algorithm, a single sample is drawn from a multivariate Log-Normal distribution for all the elements of $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}^* \sim \text{LogNormal}(\boldsymbol{\theta}, \widehat{\boldsymbol{\Sigma}}\nu)$$

The proposed $\boldsymbol{\theta}^*$ is accepted with probability ρ, a function of the prior ($\pi(\boldsymbol{\theta})$), the likelihood

($p(\mathbf{y}|\boldsymbol{\theta})$) and the transition probabilities ($q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$, $q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$, *i.e.* the density function of the proposed value):

$$\rho = \frac{\pi(\boldsymbol{\theta}^*) \cdot p(\mathbf{y}|\boldsymbol{\theta}^*) \cdot q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}) \cdot p(\mathbf{y}|\boldsymbol{\theta}) \cdot q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}$$

Finally, ν denotes the scaling parameter, tuned during an adaptive phase of the algorithm in order to provide an acceptance rate between 0.2 and 0.3 and to allow good mixing of the sampled chain (Sherlock, Fearnhead, and Roberts, 2010) .

The algorithm is coded using the `R` programming language (R Core Team, 2018). The system of differential equations (3.8) is solved using the `R` package `deSolve` (Soetaert, Petzoldt, and Setzer, 2010). The algorithm used is available at the web address `http://www.mrc-bsu.cam.ac.uk/software/miscellaneous-software/`.

Likelihood

As introduced above, the posterior distribution of the parameters π, ψ and p_{ICU}, conditional on the full data, are highly correlated. The causes of this can be further explored by investigating the log-likelihood, and detecting regions (*i.e.* combination of the parameters) where the log-likelihood is equally maximised (these correspond to the red regions of Figure 3.8). Locally flat likelihoods, such as those obtained in this study, proved to be a cause of lack of identifiability (Gustafson, 2010), so that only some combinations/functions of these parameters can be estimated.

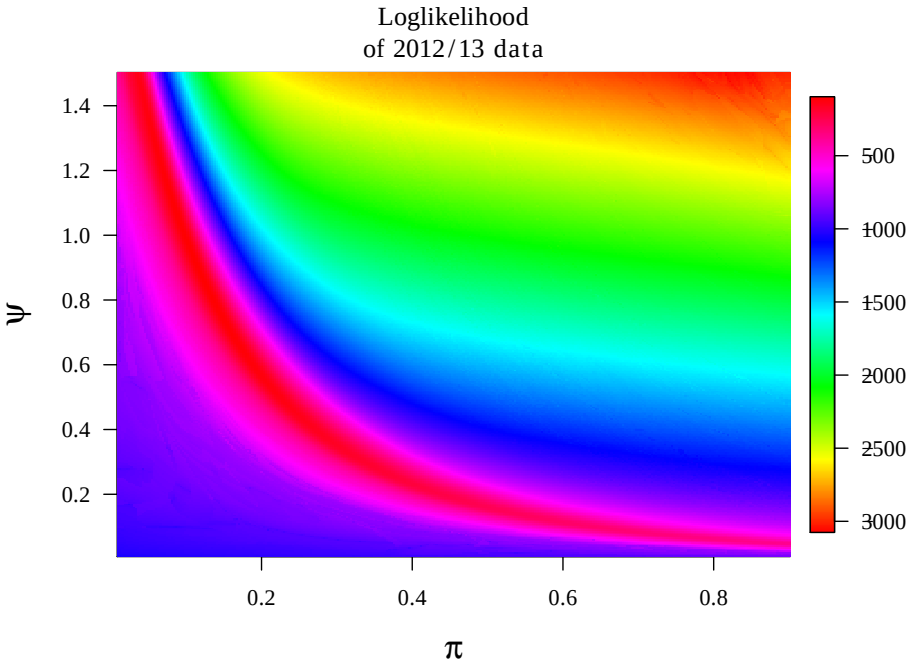


Figure 3.8: Contour plot of the log likelihood of the full dataset from season 2012/2013 as a function of the parameters ψ and π. All the other parameters are set to the median posterior values.

Parameters are even more affected by this problem when fewer data are available (e.g. at the beginning of an epidemic). This can be visualized in Figure 3.9: here the two-dimensional posterior samples from the parameters are plotted as the dataset is updated. The lighter points in yellow represent the samples of the parameters when data up to week 3 are available, and, as

more information is acquired, the posterior samples are plotted in darker colours. Other pairs of parameters (e.g. p_{ICU} and κ) were initially very correlated but, as data arrive, they become identifiable: the correlation between sample-pairs decreases and the two posterior distributions can be recognised.

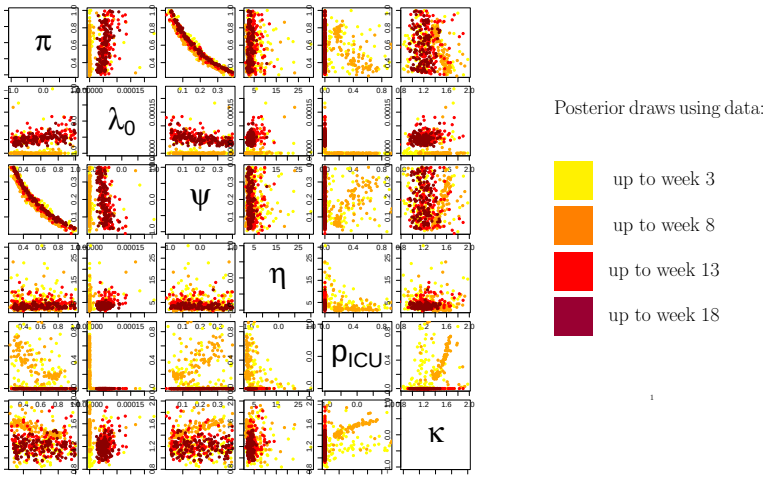


Figure 3.9: Bivariate scatter plot of a sub-sample of the posterior draws of all the parameters as the data available progress. Darker points are obtained using more data.

In Bayesian reasoning, the other method to resolve posterior distribution identifiability issues is to inject information via the definition of more-informative prior distributions. The same scatter plots of Figure 3.9, within the informative scenario, are displayed in Figure 3.10, where the constraints imposed by the data (i.e. the high correlation) are combined with independent informative priors on the parameters π and p_{ICU}.

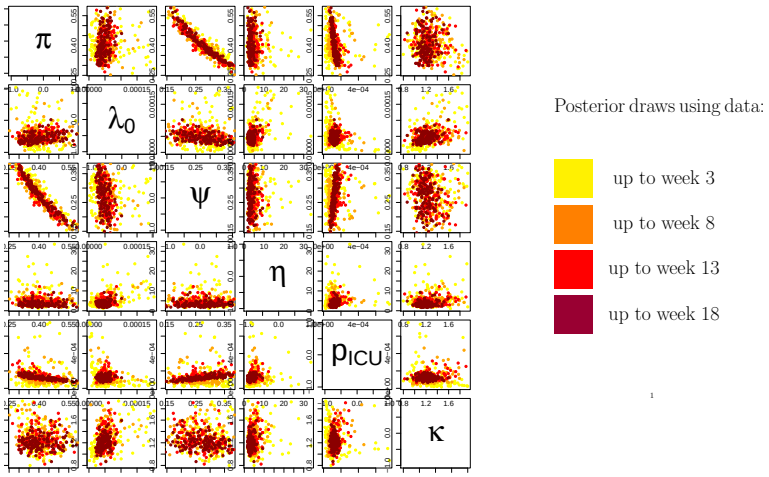


Figure 3.10: Bivariate scatter plot of a sub-sample of the posterior draws of all the parameters as the data available progress. Darker points are obtained using more data. Informative priors from Table 3.2 have been used.

3.3.4 Results

Prior to drawing inference from real data the model was tested on 5 simulated datasets. Results, both in terms of model fit and parameter estimation were satisfactory, hence the model was

further applied to real data. Results on these tested scenarios are not reported in this thesis; a comparable and more meaningful analysis of simulated hospital data is reported in Appendix C and commented upon at the end of this section. This analysis both validates the model used on real data and shows the consequences of accounting further information available during a pandemic.

Retrospective analysis

The retrospective analysis of the data is first performed in the uninformative scenario. The resulting posterior distributions of relevant parameters are displayed in Figures 3.11 and 3.12; the posterior median and 95% credible interval (CrI)s of these parameters are reported in Table 3.3. Note that the posterior distribution of the basic reproduction number R_0 is almost identical to the prior. This is due to the fact that the information contained in the data is not sufficient to determine separately the values of the parameters describing both the initial immunity and the transmission rate. For the same reason the posterior distribution of the parameter π does not change significantly from its prior, only excluding those small values that would completely prevent an epidemic to take place. This problem is due to the lack of identifiability (Gustafson, 2010) explored in Section 3.3.3.

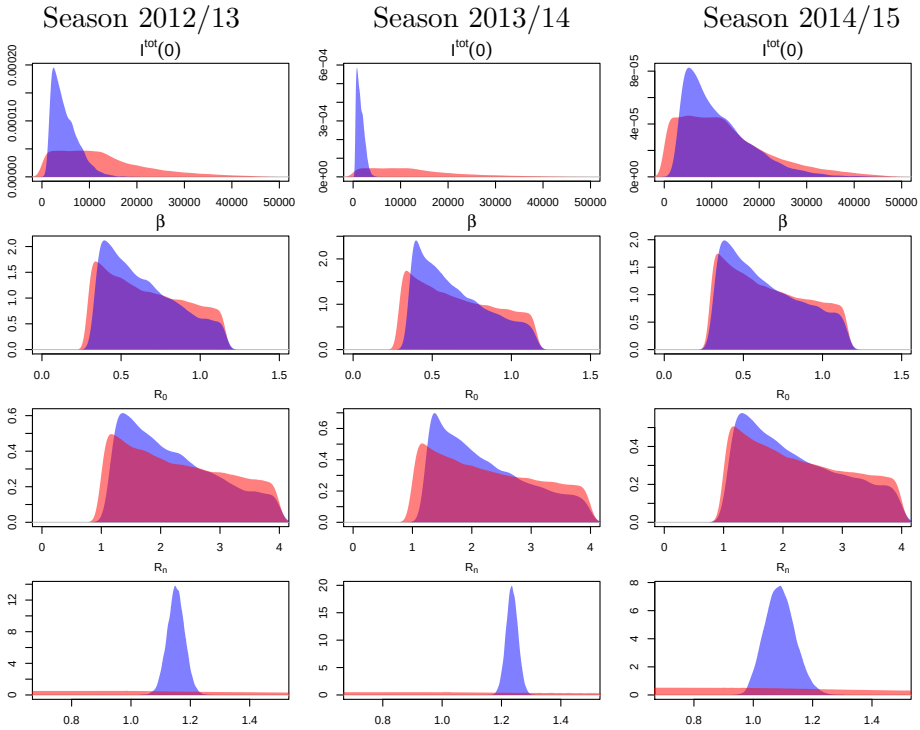


Figure 3.11: Flat scenario - Prior (red) and posterior (blue) distributions of: the total number of initial infectious I_0^{tot}, the basic transmission rate β, the basic reproduction number R_0 and the effective reproduction number R_n.

Data are much more informative about parameters η, p_{ICU} and κ. The highly variable behaviour of the ICU admissions count in season 2014/15 is reflected by the over-dispersion parameter η, whose distribution is significantly higher compared to the ones estimated from the 2012/13 and 2013/14 seasons. The range of the probability of going to ICU given infection, p_{ICU}, is always between 0.004% and 0.04%. Its median is higher in season 2014/15, in agreement

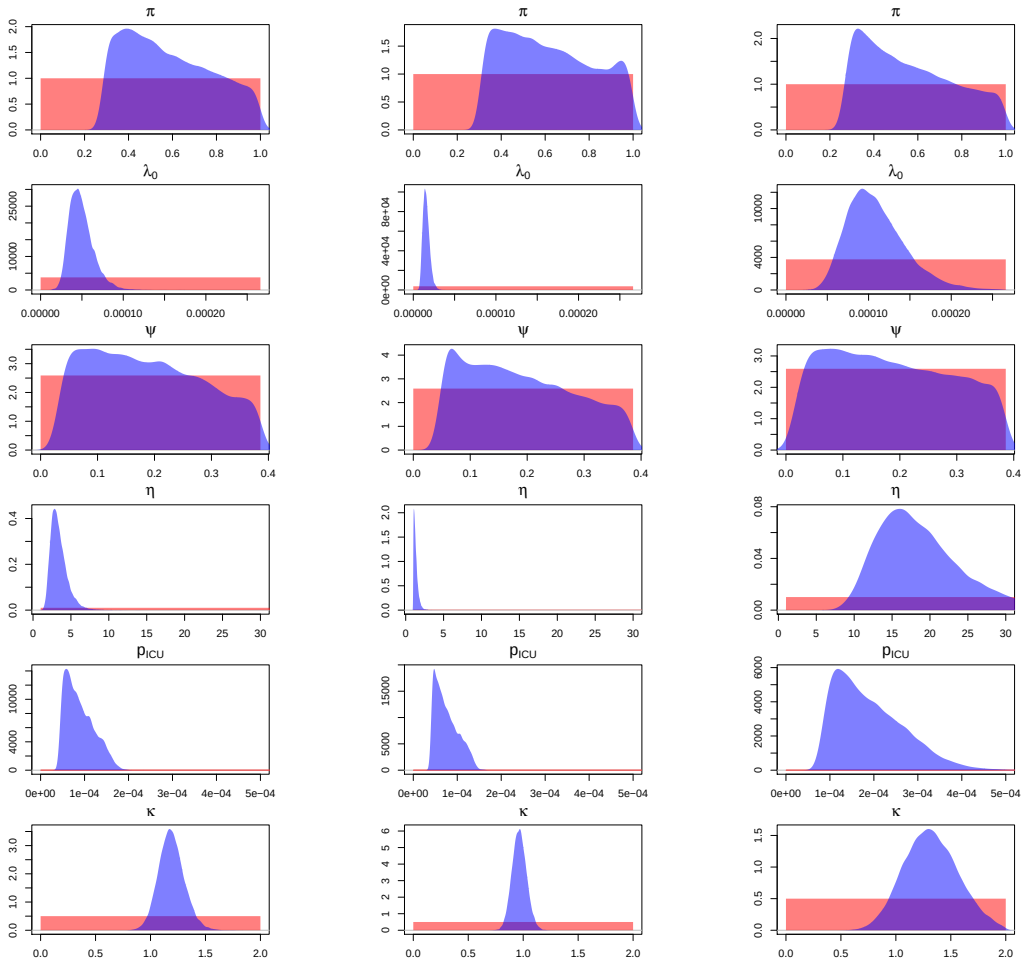


Figure 3.12: Flat scenario - Prior (red) and posterior (blue) distributions of: the initial susceptibility (π); the infection rate at $t = 0$ (λ_0; the exponential growth rate (ψ) the over-dispersion parameter (η); the probability of ICU admission given infection (p_{ICU}); and the scaling parameter (κ). The results are derived from season 2012/13 (left column), season 2013/14 (centre) and season 2014/15 (right column).

with the higher severity that was detected during this influenza season (Boddington et al., 2017). The multiplicative factor κ introduced to allow for a school-closure effect is centred on 1 for season 2013/14 and centred around higher values in the remaining seasons. A possible explanation for this counter-intuitive phenomenon relies on the age distribution of the sample population. The analysed data have a different distribution compared to the English population (Boddington et al., 2017; Office of National Statistics, 2012-2015), with patients over 65 being over represented and children in school years being under represented. The elderly individuals might be more likely to meet other potential influenza spreaders (e.g. children) during school closures, particularly over the Christmas holiday. It makes sense, therefore, to observe an inverse relationship between school closure and the transmission rate, in contrast to results that might be expected from a more representative sample of the population (Te Beest et al., 2015). However, this piecewise increment in transmission rate may incorporate other time-varying phenomena that affect the force of infection. The Christmas holiday often coincides with the beginning of a colder and more humid period and changes in vapour pressure, that

might imply an increasing spread of influenza (Dushoff et al., 2004; Lipsitch and Viboud, 2009). Lastly the posterior median of the effective reproduction number R_n is equal to 1.152, 1.235, 1.089 in seasons 2012/13, 2013/14 and 2014/15 respectively. Since this parameter contains the product of transmission-rate parameters and initial-immunity parameters it can be identified better than R_0 and hence it is more representative of the overall transmission of each season.

Parameter		Season 2012/13 Posterior Me (CrI)	Season 2013/14 Posterior Me (CrI)	Season 2014/15 Posterior Me (CrI)
Susceptibility	π	0.546 (0.297 - 0.969)	0.589 (0.32 - 0.977)	0.531 (0.28 - 0.968)
Initial number of infectious	I^{tot}	4106 (1441 - 11510)	1357 (484 - 3312)	9590 (3053 - 28493)
Transmission rate	β	0.611 (0.344 - 1.126)	0.608 (0.367 - 1.118)	0.596 (0.324 - 1.119)
Over-dispersion	η	3.204 (1.888 - 6.101)	1.25 (1.011 - 2.096)	17.925 (10.412 - 35.812)
P of ICU given infection	$p_{\text{ICU}}10^4$	0.841 (0.458 - 1.614)	0.713 (0.419 - 1.338)	1.749 (0.848- 3.745)
Factor for school closure	κ	1.185 (0.971 - 1.434)	0.965 (0.841 - 1.1)	1.313 (0.866 - 1.824)
Eff. reproduction number	R_n	1.152 (1.093 - 1.211)	1.235 (1.196 - 1.275)	1.089 (0.997 - 1.195)

Table 3.3: Posterior medians and 95% CrIs from the retrospective analysis of the ICU admissions with uninformative priors.

Under a goodness-of-fit perspective, the introduction of the parameter κ allows the flexibility needed to represent the specific features of each season, even though its posterior predictive distribution always include the non-effect value 1. This can be observed in the posterior predictive distribution of the weekly ICU admissions reported in Figure 3.13. Specifically in season 2012/13 the posterior predictive distribution manages to reproduce the plateau that takes place from the end of the Christmas vacation to the February half term. Regarding instead the double peaking season of 2014/15, the 95% Credible bounds are not narrow, but the timing of the peak of the distribution is predicted substantially better than in the case of constant infection rate (results not shown). The high variability of the data considered, combined with the constraint of a deterministic model, cause an overall weaker fitting of the model to the data of this season.

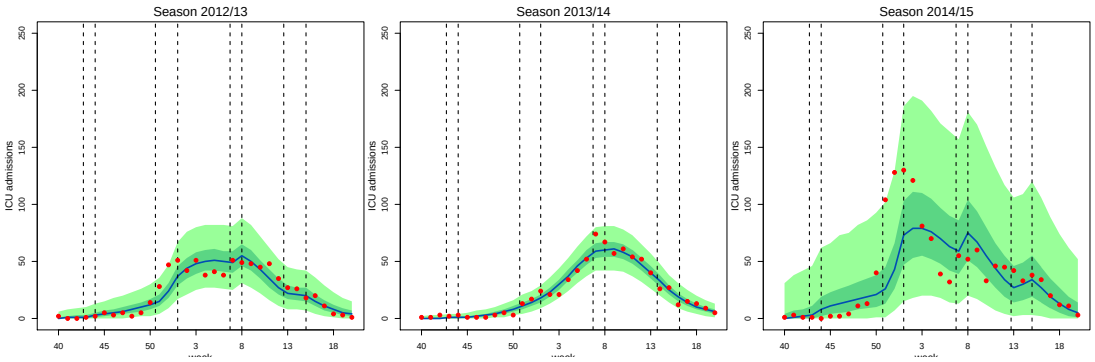


Figure 3.13: Flat scenario - Median (blue), 95 % CrI (light green) and quartile (dark green) of the posterior predictive distributions and observed values (red) for the weekly ICU/HDU admissions across seasons. The vertical dashed lines represent the breakpoints for the piecewise-constant transmissibility $\beta^*(t)$ (*i.e.* start and end of each school holiday).

The same analysis is performed in the second scenario, i.e. allowing informative priors on the susceptibility π and on p_{ICU} as defined in Table 3.2. The introduction of these prior distributions

compensates for the lack of information, allowing the identification of π and improving the precision of the posterior distribution of p_{ICU}. This affects also other functional parameters such as β and R_0. However, their posterior distributions are driven by the prior distributions alone, and they do not learn from the data. In terms of fit there was no improvement. Results are reported in Appendix C.

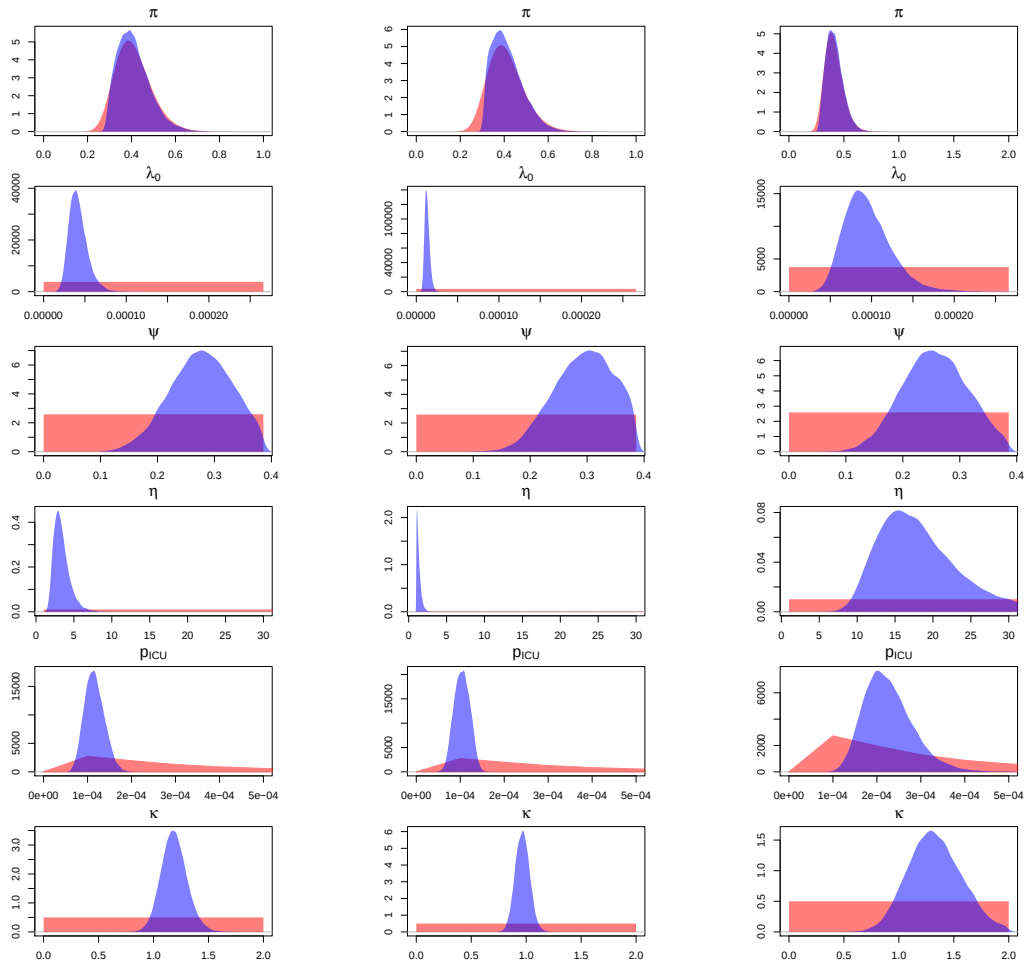


Figure 3.14: Informative scenario - Prior (red) and posterior (blue) distributions of: the initial susceptibility (π); the infection rate at $t = 0$ (λ_0; the exponential growth rate (ψ) the over-dispersion parameter (η); the probability of ICU admission given infection (p_{ICU}); and the scaling parameter (κ). The results are derived from season 2012/13 (left column), season 2013/14 (centre) and season 2014/15 (right column).

Prediction

The prospective analysis of the data in the uninformative scenario results in very wide predictions of the future dynamics, therefore the informative priors reported in Table 3.2 are used for predictions. The performance of the model at different times is plotted in Figure 3.15 for each season.

Season 2013/14, despite displaying the most regular data, is the most difficult to predict: the well-defined initial growth biases the predictions towards a major outbreak. This leads to the median and the CrIs of the posterior predictive distribution over-estimating the data until

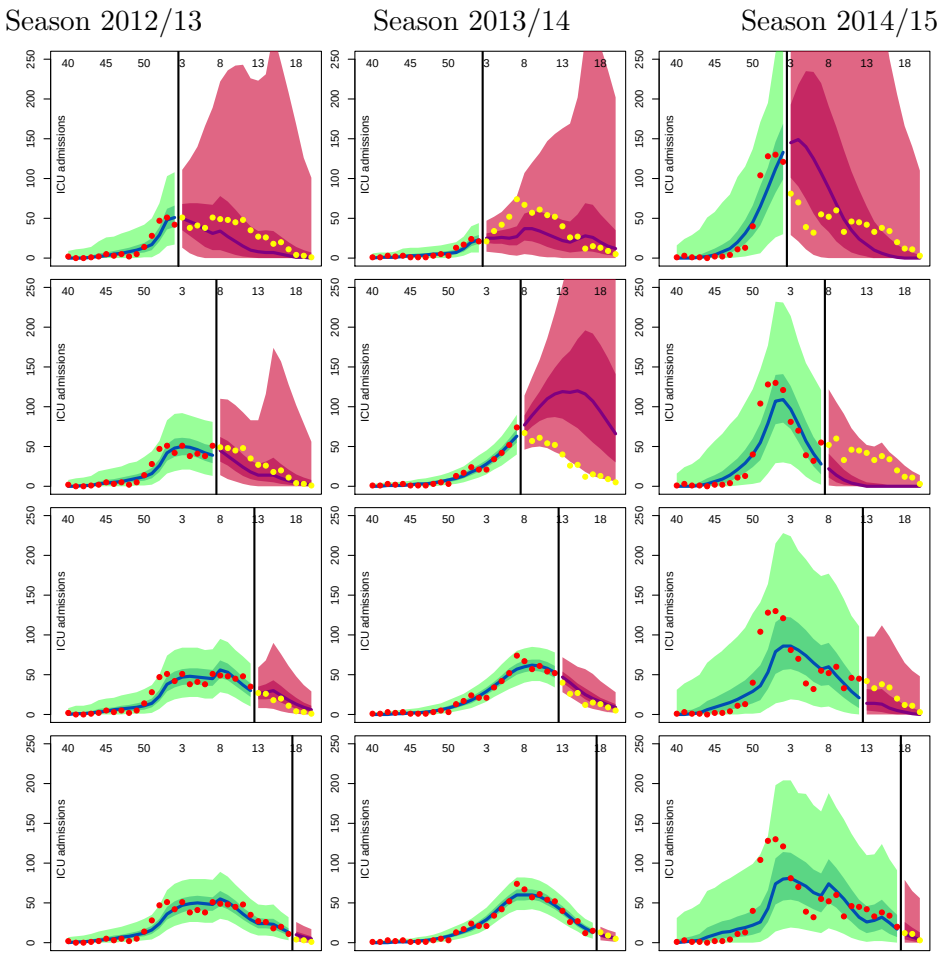


Figure 3.15: Informative scenario - The black line displays the analysis time; the blue line and green shaded area represent median, quartile (dark green) and 95% CrIs (light green) of the posterior predictive distribution for the training weeks. The pink area displays posterior quartiles (deep pink) and 95% CrIs (light pink) for the predicted future observations, and the purple line displays the median; the red dots are the training data and the yellow dots are the objects of prediction.

mid-march (week 13 from the beginning of the year). For the other two seasons, the median predicted weekly ICU admissions is always very close to the data points, but the CrIs narrow to reasonable bounds only towards the end of February (week 8 from the beginning of the year).

Prediction is challenging, as demonstrated by the variability of the predictions. For example, the 95% CrI of the predicted number of ICU admissions 3 weeks in advance, when the epidemic is still taking off (i.e. at the third week of January) is as wide as 138 for season 2012/2013 (from 2 to 140 ICU admissions), 52 for season 2013/2014 (from 6 to 58 ICU admissions) and 473 for season 2014/2015 (from 11 to 484 ICU admissions). Due to the different sizes of the epidemics, the coefficient of variation (i.e. the ratio of the posterior standard deviation to the posterior mean) can be used to compare them: it is equal to 0.751 for season 2012/13, 0.491 for season 2013/14, and 0.742 for season 2014/15, highlighting that prediction precision increases when the epidemic is smaller and less over-dispersed.

In spite of the simplicity of the model presented, the flexibility introduced by the parameter κ allows for the correction "on the fly" of the prediction, adapting to new peaks (e.g. season

2014/15) or periods of constant influenza circulation (e.g. season 2012/13).

Nonetheless, similarly to most epidemic models attempting predictions (Birrell et al., 2011; Ong et al., 2010), results are not useful (i.e. precise enough to determine a health policy response) until after the epidemic has peaked.

Simulated pandemic scenario

As mentioned in Section 3.3.1, the data-collection scheme is enhanced in the event of a pandemic. If this is the case, USISS reports also the weekly count of hospital admissions in all the trusts in England.

A simulated dataset containing the number of hospital admissions at all levels of care is obtained from the parameter set $\boldsymbol{\theta} = (\pi = 0.5, \lambda_0 = 0.0001, \psi = 0.2, \eta = 20, p_H = 0.002, \kappa = 1.5)$, where p_H replaces p_{ICU} and represents the proportion of infections leading to hospitalization, and η denotes the over-dispersion of hospital data.

Both analysis with flat priors and informative priors were run. For the former scenario, prior to posterior plots of the values for the parameters are reported in Figure 3.16 and for the latter scenario they are reported in Figure 3.17.

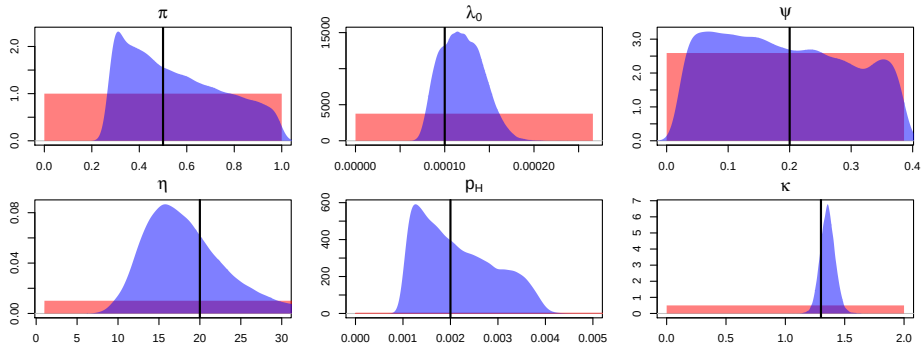


Figure 3.16: Flat Scenario - Prior (red) and posterior (blue) distributions of the parameters. The parameters are: π, ψ, λ_0, η, p_H and κ. The black vertical lines denote the values used to simulate the dataset.

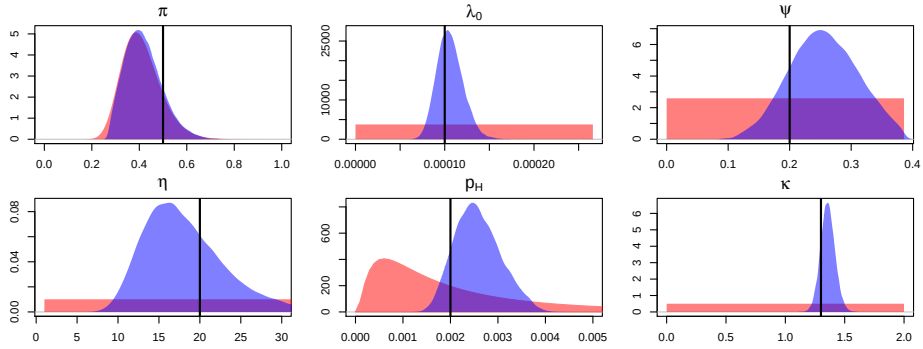


Figure 3.17: Informative Scenario - Prior (red) and posterior (blue) distributions of the parameters. The parameters are: π, ψ, λ_0, η, p_H and κ. The black vertical lines denote the values used to simulate the dataset.

Despite the increased number of observations, the model performs very similarly to the case of non-pandemic ICU-counts data. Identifiability issues are diagnosed in the uniform-priors scenario and predictions are good only when more informative prior distributions (on

the susceptibility and probability of hospitalization) are assumed. Additional results from this analysis are reported in Appendix C together with further results on real data.

3.3.5 Discussion

Corbella et al. (2018) proposes a model to estimate and predict influenza outbreaks from routinely collected data on admissions to ICU/HDU.

The performance of the proposed model is investigated both on simulated and on real data. By fitting the model to simulated numbers of weekly ICU/hospital admissions, satisfactory estimates of some of the main parameters can be obtained, even with very vague prior information, including the initial infection rate, the probability of going to ICU/hospital given infection, the effective reproduction number R_n and the scaling factor for school holidays κ. When information on the distribution of the average immunity $(1 - \pi)$ and on p_{ICU} is injected, meaningful estimates of the previously unidentifiable parameters are obtained. The evolution of the outbreak can also be forecast by analysing the first months of the epidemic using data up to the peak of influenza activity.

The model is applied to real data on the weekly number of ICU admissions from seasons 2012/13, 2013/14 and 2014/15, confirming the performance obtained on the simulated data. The estimated values of the effective reproduction number R_n are similar to those estimated during the past decade of seasonal influenza (Baguelin et al., 2013). A scaling parameter allows the transmission rate to vary between school and holiday/half-term periods, which results in a good fit of the model to the data for most of the seasons considered. A more complete investigation of the temporal variation of the transmission rate might improve the flexibility of the proposed model, and therefore the fit to more anomalous epidemics.

Recently, a similar analysis was performed on the Finnish influenza pandemic of 2009 (Shubin et al., 2016) using a more elaborate model, analysing confirmed data on both hospitalizations and GP consultations. Their inclusion of GP data enhances the performance of the inference. Nevertheless, these data are harder to collect in a larger population (England is almost 10 times more populated than Finland) and out of pandemic emergencies. By contrast, the inference performed through the model of Corbella et al. (2018) is driven by few data, though readily available, even in real time, in seasonal settings. A further advance of the model by Shubin et al., 2016 is that the transmission parameter is time-varying according to a Gaussian Process: this allows an accurate description of the past dynamics but makes prediction infeasible, since this temporal variation cannot be forecast. By contrast, the simple piecewise constant model presented here is able to forecast well the future trend and it includes enough flexibility to describe appropriately the present and the past data.

This work has also some limitations: firstly, the transmission model is non-age-specific. The assumption of homogeneous mixing across regions and age groups is very strong but this was dictated by the very small sample sizes which did not allow sub-grouping. Secondly, the quality of some estimates and predictions strongly relies on prior information on the proportion of non-immune people. As this information is needed to overcome the lack of identifiability in the parameters, sero-prevalence data following the 2010/11 epidemic were used. This is not likely to be correct for all the three seasons analysed, as the predominant strain circulating was different across seasons. Likewise, the model that describes the time elapsing between infection and ICU

admission, is assumed to be fixed and mostly known, but this assumption is not likely to be valid. The other element that defines the observational process, i.e. the probability of ICU admission given infection, is also sensitive to the choice of prior distribution.

The work presented here is a proof of concept of the potential for estimation and prediction of influenza transmission from USISS data. At the same time, the results highlight the need of collecting external data to formulate an appropriate prior distribution on the initial immunity of the population, particularly in the event of a pandemic.

The availability of this information, together with the set-up provided here, allows to retrospectively infer the epidemic parameters from routinely collected data on severe cases during seasonal outbreaks and to predict the temporal dynamics of new epidemics.

3.4 Conclusions

This chapter provides a brief introduction to epidemic models and an example of their use.

While analysing their essential features, epidemic models appear to be very flexible tools, adaptable to different levels of stochasticity, granularities of time, patterns of disease, etc. Some criteria for model choice, prior to the analysis, are exposed in the chapter. However, this list is not exhaustive: example-specific criteria for model choice (e.g. criteria based on goodness-of-fit measures or predictive-ability measures) have not been presented since the literature has not agreed on their use is an epidemic-modelling context yet.

Epidemic models could also be approached within a SSM perspective; this would allow to better identify the stochastic relationship among r.v.s. Some methods for inference in SSMs are described and applied to the analysis of epidemic data in the next chapters.

In this context, the model used in Section 3.3 could be made more realistic. The model presented is fully deterministic and the variability detected in the data is attributed only to the Negative Binomial observational noise. A more sensible assumption would be to attribute some noise to the process of severity, which describe the counts of individuals going to ICU given infection, and the delays between events. This improvement is developed in Chapter 5.

Chapter 4

Sequential Monte Carlo methods for inference in State Space Models

Chapters 2 and 3 introduced the processes of infected individuals developing severe symptoms and of the spread of an infectious disease, respectively. Both models can be seen as multi-state models (MSMs); hence the counts of individuals in the states or moving across the states can be seen as state space models (SSMs).

The inferential methods used to estimate the parameters governing these systems showed some pitfalls, above all due to the fact that the available data are only a partial signal from a much more complex hidden process. Sequential Monte Carlo (SMC) methods have been developed precisely to deal with partially observed dynamic systems, i.e. SSMs (presented in more detail in Schön et al. (2018)).

The SSM perspective introduced in Chapter 1 is here presented more in detail (Section 4.1) and some methods for the inference in such dynamic systems are introduced (Sections 4.2 and 4.3). SMC methods are used here to approximate the likelihood of data arising from an SSM, conditioning on a given value of the parameters of interest. The parameter space is explored using more classical Monte Carlo (MC) algorithms; hence the literature on SMC for the sequential exploration of the parameter space is omitted.

Two applications, one investigating a severity process (Section 4.4) and the other exploring a transmission process (Section 4.5), are also presented.

4.1 State space model specification

A SSM representation of an observable phenomenon has been widely used in many fields, from indoor positioning problems in engineering (Solin et al., 2018), to environmental studies (Anderson, 1996), to epidemic models (Magpantay et al., 2015). Hence, the literature has proliferated across subjects. Some of the most relevant aspects of SSMs are summarised below; comprehensive reviews can be found in volumes on the topic (Brockwell and Davis, 2016), key papers (Commandeur, Koopman, and Ooms, 2011; Dureau, Ballesteros, and Bogich, 2014) as well as in recent course material (Schön and Lindsten, 2017).

4.1.1 Elements of a SSM

From Definition 1 (Page 12), a SSM is a stochastic process that makes use of a latent variable representation to describe dynamical phenomena (Schön and Lindsten, 2017). It has two components: a latent process, denoted by $\{X_t\}_{t\geq 1}$ representing the underlying dynamics; and an observed process denoted by $\{Y_t\}_{t\geq 1}$. Without loss of generality, the processes are assumed to evolve over discrete time since this is the most common form encountered later in applications. Nevertheless, the following can be extended straightforwardly to other dimensions.

In this thesis, the state process is assumed Markovian over time, hence the subset of SSMs analysed can be also classified in the category of partially observed Markov processs (POMPs) (King, Nguyen, and Ionides, 2016) or hidden Markov models (HMMs) (Churchill, 2005). From now onwards, unless otherwise specified, *SSM* will refer to *Markovian SSM*.

A parameter-driven SSM can be defined through the state and the observation equation

$$\begin{aligned} X_t|(X_{t-1},\boldsymbol{\theta}) &\sim p(x_t|x_{t-1},\boldsymbol{\theta}) \\ Y_t|(X_t,\boldsymbol{\theta}) &\sim p(y_t|x_t,\boldsymbol{\theta}) \end{aligned}$$

both characterised by a vector of parameters $\boldsymbol{\theta}$ (Birrell, De Angelis, and Presanis, 2018; Brockwell and Davis, 2016). To fully define the model, the state process at $t = 0$ must be also specified; this can either be a fixed value (which usually is a parameter to estimate), or take its own distribution: $X_0|\boldsymbol{\theta} \sim p(x_0|\boldsymbol{\theta})$.

The state process $\{X_t\}_{t\geq 0}$ is often called the *dynamic parameter*, while $\boldsymbol{\theta}$ is often referred to as *static parameter(s)* and the observational process $\{Y_t\}_{t\geq 1}$ are called the *measurements*.

The full specification of a SSM is:

$$\begin{aligned} X_t|(X_{t-1},\boldsymbol{\theta}) &\sim p(x_t|x_{t-1},\boldsymbol{\theta}) \\ Y_t|(X_t,\boldsymbol{\theta}) &\sim p(y_t|x_t,\boldsymbol{\theta}) \\ X_0|\boldsymbol{\theta} &\sim p(x_0|\boldsymbol{\theta}) \end{aligned} \tag{4.1}$$

which will be adopted from here onwards. Model 4.1 defines the following full probability model:

$$p(x_{0:T},y_{0:T}|\boldsymbol{\theta}) = \prod_{t=1}^{T} p(y_t|x_t,\boldsymbol{\theta}) \prod_{t=1}^{T} p(x_t|x_{t-1},\boldsymbol{\theta}) p(x_0|\boldsymbol{\theta}) \tag{4.2}$$

that describes the joint distribution of measurements and dynamic parameters conditional on the static parameters. Thanks to Markovianity and conditional independence this distribution is decomposed: the state density and the observation density can be recognised in the right hand side of the equation.

A SSM can also be represented by a graphical model which is a probabilistic model where a graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$ represents the conditional independence structure (edges $\mathcal{E}$) between random variables (r.v.s) (nodes $\mathcal{V}$). A graphical model representing the state and observational process of a SSM is illustrated in Figure 4.1.

4.1.2 Inference from a SSM: problem set-up

The information carried by the measurements can be used to estimate the state process $\{X_t\}_{t\geq 0}$, usually conditional on specific values of the static parameter $\boldsymbol{\theta}$, or to infer the static parameter $\boldsymbol{\theta}$, usually marginally on the distribution of the state process.

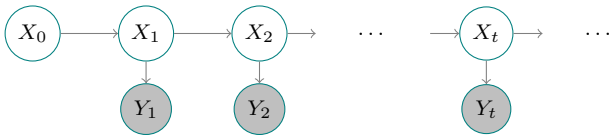


Figure 4.1: Graphical model for a SSM. Grey nodes correspond to observed variables and white nodes are latent variables. The arrows express the dependence among the variables.

State inference problems can take any of the following forms (Lindsten, 2013):

- deriving $p(x_t|y_{1:t}, \boldsymbol{\theta})$, i.e. the distribution of of the state process at t conditionally on the data up until t, is called *marginal filtering*;
- deriving $p(x_{0:t}|y_{1:t}, \boldsymbol{\theta})$, $t = 1, \ldots, T$, i.e. the distribution of the whole state process up until t conditionally on the data up until t, is called *joint filtering* and is often constructed sequentially;
- deriving $p(x_{t+s}|y_{1:t}, \boldsymbol{\theta})$, $t = 1, \ldots, T, s = 1, \ldots,$ i.e. the distribution of the state process for future intervals until time $t + s$, conditionally on the data up until t, is called *prediction*;
- deriving $p(x_{0:t}|y_{1:T}, \boldsymbol{\theta})$, $t \leq T$ i.e. the distribution of the whole state process up until T conditionally on the full data, is called *joint smoothing*;
- deriving $p(x_t|y_{1:T}, \boldsymbol{\theta})$, $t \leq T$, i.e. the distribution of the state process at t conditionally on the full data, is called *marginal smoothing*.

This thesis, and this chapter, mainly addresses filtering problems. Such filtering can be approached exploiting Bayes' theorem, conditional independence, and the Markovianity of the system. The filtering distribution at time t can be expressed in terms of: a normalizing factor $p(y_t|y_{1:t-1}, \boldsymbol{\theta})$; the, so called, measurement update $p(y_t|x_t, \boldsymbol{\theta})$, that is the data distribution at time t conditional on the hidden state at time t; and a prediction update $p(x_t|y_{1:t-1}, \boldsymbol{\theta})$, i.e. the distribution of the hidden state at t conditional on the previous data, as follows:

$$\begin{aligned}
p(x_t|y_{1:t}, \boldsymbol{\theta}) &= p(x_t|y_t, y_{1:t-1}, \boldsymbol{\theta}) \\
&= \frac{p(y_t, y_{1:t-1}|x_t, \boldsymbol{\theta})p(x_t|\boldsymbol{\theta})}{p(y_t, y_{1:t-1}|\boldsymbol{\theta})} \\
&= \frac{p(y_t|y_{1:t-1}, x_t, \boldsymbol{\theta})p(y_{1:t-1}|x_t, \boldsymbol{\theta})p(x_t|\boldsymbol{\theta})}{p(y_t|y_{1:t-1}, \boldsymbol{\theta})(y_{1:t-1}|\boldsymbol{\theta})} \\
&= \frac{p(y_t|y_{1:t-1}, x_t, \boldsymbol{\theta})\frac{p(x_t|y_{1:t-1}, \boldsymbol{\theta})p(y_{1:t-1}|\boldsymbol{\theta})}{p(x_t|\boldsymbol{\theta})}p(x_t|\boldsymbol{\theta})}{p(y_t|y_{1:t-1}, \boldsymbol{\theta})p(y_{1:t-1}|\boldsymbol{\theta})} \\
&= \frac{p(y_t|x_t, \boldsymbol{\theta})p(x_t|y_{1:t-1}, \boldsymbol{\theta})}{p(y_t|y_{1:t-1}, \boldsymbol{\theta})}
\end{aligned} \tag{4.3}$$

The prediction update in the numerator can also be decomposed exploiting Markovianity:

$$\begin{aligned}
p(x_t|y_{1:t-1}, \boldsymbol{\theta}) &= \int_{X_{t-1}} p(x_t, x_{t-1}|y_{1:t-1}, \boldsymbol{\theta}) \mathrm{d}x_{t-1} \\
&= \int_{X_{t-1}} p(x_t|x_{t-1}, \boldsymbol{\theta})p(x_{t-1}|y_{1:t-1}, \boldsymbol{\theta}) \mathrm{d}x_{t-1}
\end{aligned} \tag{4.4}$$

Plugging Equation 4.4 into Equation 4.3, the filtering distribution becomes:

$$p(x_t|y_{1:t},\boldsymbol{\theta}) = \frac{p(y_t|x_t,\boldsymbol{\theta})\int_{X_{t-1}} p(x_t|x_{t-1},\boldsymbol{\theta})p(x_{t-1}|y_{1:t-1},\boldsymbol{\theta})\mathrm{d}x_{t-1}}{p(y_t|y_{1:t-1},\boldsymbol{\theta})}$$

A closed-form solution to the integral in the numerator, as well as to the normalizing constant in the denominator, is only derivable in very rare cases such as the linear Gaussian SSM, whose solution is the Kalman Filter (Kalman, 1960). More often, the filtering distribution at time t can be approximated recursively by a two-step procedure that alternates the approximation of the measurement update and the sampling from prediction update below.

$$p(x_t|y_{1:t},\boldsymbol{\theta}) = \frac{p(y_t|x_t,\boldsymbol{\theta})p(x_t|y_{1:t-1},\boldsymbol{\theta})}{p(y_t|y_{1:t-1},\boldsymbol{\theta})} \qquad \text{(Measurement update)}$$

$$p(x_t|y_{1:t-1},\boldsymbol{\theta}) = \int_{X_{t-1}} p(x_t|x_{t-1},\boldsymbol{\theta})p(x_{t-1}|y_{1:t-1},\boldsymbol{\theta})\mathrm{d}x_{t-1} \qquad \text{(Time update)}$$

The most common among these two-step procedures is presented in Section 4.2.

The problem of **parameter inference** instead consists of deriving information on the static parameters $\boldsymbol{\theta}$ conditionally on the available observations $y_{1:t}$. Parameter inference can be approached either in a Bayesian context (where the parameters are considered as r.v.s $\boldsymbol{\Theta}$ with prior distribution $\boldsymbol{\Theta} \sim \pi(\boldsymbol{\theta})$, and the posterior distribution $\boldsymbol{\Theta}|y_{1:t}$ is to be derived) or in a purely likelihood context. In both cases the most challenging task is the computation (or approximation) of the probability of the measurements conditional on the parameter values: $p(y_{1:t}|\boldsymbol{\theta})$, i.e. the data distribution.

This distribution can be obtained by marginalizing the joint distribution of the latent states and the observed measurements (Equation 4.2), with respect to the sequence of the hidden states:

$$p(y_{1:t}|\boldsymbol{\theta}) = \int_{X_{0:T}} p(x_{0:T}, y_{1:T}|\boldsymbol{\theta})\mathrm{d}x_{0:T} \tag{4.5}$$

which can be either maximised (in a non-Bayesian context) or fully explored together with $\pi(\boldsymbol{\theta})$ for the derivation of a posterior distribution.

In either case Equation 4.5 is not available in closed form, except from a few rare SSMs.

4.2 Methods for state inference

Methods for state inference highly exploit basic simulation techniques such as MC integration and importance sampling, which are briefly recalled in Appendix D. Subsection 4.2.1 contains the general setting for most of the simulation methods used to derive filtering distributions for a given parameter $\boldsymbol{\theta}$, and hence acquire insight on the states $X_{0:t}$. In Subsection 4.2.2 one of these methods, the bootstrap particle filter (BPF), is illustrated.

Most of the results reported below can be found in common textbooks on MC and/or SMC methods (Brooks et al., 2011; Robert and Casella, 2013) as well as in key papers (Andrieu, Doucet, and Holenstein, 2010; Arulampalam et al., 2002; Gilks and Berzuini, 2001).

All distributions in this section are assumed conditional on $\boldsymbol{\theta}$, the static parameter, and for ease of presentation, this dependence is omitted, denoting $p(x_t|\boldsymbol{\theta})$ by $p(x_t)$, $p(x_t|y_{1:t},\boldsymbol{\theta})$ by $p(x_t|y_{1:t})$, etc.

4.2.1 Sequential importance (re)sampling

Sequential importance sampling (Arulampalam et al., 2002) can be used to approximate the distribution of the hidden states conditionally on the data, by decomposing the problem into simpler, lower-dimensional, approximation steps. The sequential importance sampling method presented below targets the joint filtering distribution $p(x_{0:t}|y_{1:t})$.

The target distribution at time t can be simplified using the conditional probability definition and Markovianity:

$$\begin{aligned}
p(x_{0:t}|y_{1:t}) &= p(x_{0:t}|y_t, y_{1:t-1}) \\
&= \frac{p(x_{0:t}, y_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}, && \text{by definition of conditional probability} \\
&= \frac{p(y_t|y_{1:t-1}, x_{0:t})p(x_{0:t}|y_{1:t-1})}{p(y_t|y_{1:t-1})}, && \text{conditioning on } x_{0:t} \\
&= \frac{p(y_t|y_{1:t-1}, x_{0:t})p(x_t, x_{0:t-1}|y_{1:t-1})}{p(y_t|y_{1:t-1})} \\
&= \frac{p(y_t|x_t)p(x_t|x_{0:t-1}, y_{1:t-1})p(x_{0:t-1}|y_{1:t-1})}{p(y_t|y_{1:t-1})}, && \text{by Markovianity and} \\
& && \text{definition of conditional probability} \\
&\propto p(y_t|x_t)p(x_t|x_{t-1})p(x_{0:t-1}|y_{1:t-1}), && \text{by conditional independence.}
\end{aligned}$$

To obtain samples from this target distribution, assume that, at time t, a weighted sample $\left\{x_{0:t-1}^{(n)}, w_{t-1}^{(n)}\right\}_{n=1}^{N}$, of the target distribution at $t-1$, $p(x_{0:t-1}|y_{1:t-1})$, is available. Letting $\delta_a(x)$ define a Dirac point mass in a, the sample provides the following approximation to the target distribution at $t-1$:

$$\widehat{p}(x_{0:t-1}|y_{1:t-1}) = \sum_{n=1}^{N} \delta_{x_{0:t-1}^{(n)}}(x) w_{t-1}^{(n)}$$

To propose values for the next approximation step, assume an importance distribution $q(x_{0:t}|y_{1:t})$ that is factorisable as follows:

$$\begin{aligned}
q(x_{0:t}|y_{1:t}) &= q_t(x_t|x_{t-1}, y_t)q(x_{0:t-1}|y_{1:t-1}) \\
&= q_t(x_t|x_{t-1}, y_t)q_{t-1}(x_{t-1}|x_{t-2}, y_{t-1})q(x_{0:t-2}|y_{1:t-2}) \\
&= \dots \\
&= q_0(x_0)\prod_{s=1}^{t} q_s(x_s|x_{s-1}, y_s)
\end{aligned}$$

where $q_t(x_t|x_{t-1}, y_t)$ and $q_0(x_0)$ could be generic importance distributions for x_t. A sample from the importance distribution $q(x_{0:t}|y_{1:t})$ could then be obtained recursively from the sample from the importance sample at the previous time step: $\left\{x_{0:t}^{(n)}\right\}_{n=1}^{N} = \left\{x_t^{(n)}, x_{0:t-1}^{(n)}\right\}_{n=1}^{N}$ with $x_t^{(n)}$ sampled from $q_t(x_t|x_{0:t-1}^{(n)}, y_t)$, $\left\{x_{t-1}^{(n)}\right\}_{n=1}^{N}$ sampled from $q_{t-1}(x_{t-1}|x_{0:t-2}^{(n)}, y_{t-1})$ and so on. Note that usually the importance density is chosen also according to the data at time t, to better match with the target density: $p(x_{0:t}|y_{1:t})$.

The resulting importance weight for the sample $\left\{x_{0:t}^{(n)}\right\}_{n=1}^{N}$ is obtained by dividing the target

density by the importance density, i.e.:

$$\begin{aligned}\omega(x_{0:t}, y_{1:t}) &= \frac{p(y_t|x_t)p(x_t|x_{t-1})p(x_{0:t-1}|y_{1:t-1})}{q_0(x_0)\prod_{s=1}^{t} q_s(x_s|x_{s-1}, y_s)} \\ &= \frac{p(y_t|x_t)p(x_t|x_{t-1})p(x_{0:t-1}|y_{1:t-1})}{q_t(x_t|x_{t-1}, y_t)q(x_{0:t-1}|y_{1:t-1})} \\ &= \frac{p(y_t|x_t)p(x_t|x_{t-1})}{q_t(x_t|x_{t-1}, y_t)}\omega(x_{0:t-1}, y_{1:t-1}) \\ &= \frac{p(y_t, x_t|x_{t-1})}{q_t(x_t|x_{t-1}, y_t)}\omega(x_{0:t-1}, y_{1:t-1})\end{aligned} \tag{4.6}$$

Applying recursively this derivation (for $t, t-1, \ldots, 1, 0$) leads to a weighted sample from the target distribution $p(x_{0:t}|y_{1:t})$ as illustrated in the pseudo code of Algorithm 1.

Result: $\left\{x_{0:t}^{(n)}, w_t^{(n)}\right\}_{n=1}^{N}$: a weighted sample from $p(x_{0:t}|y_{1:t})$
Input: N,$q_0(x)$, $q_t(x)$, $p(x_0)$, $p(x_t|x_{t-1})$, $p(y_t|x_t)$ for $t = 0, 1, \ldots t$
for $n = 1, \ldots N$ **do**
  sample
  $x_0{}^{(n)} \sim q_0(x)$
  compute the importance weights
  $\widetilde{w}_0^{(n)} = \frac{p(x_0^{(n)})}{q_0(x_0^{(n)})}$
end
normalize the importance weights
$w_0^{(n)} = \frac{\widetilde{w}_0^{(n)}}{\sum_{i=1}^{N} \widetilde{w}_0^{(n)}} \qquad \forall n = 1, \ldots N$
for $s = 1, \ldots t$ **do**
  for $n = 1, \ldots N$ **do**
    sample
    $x_s{}^{(n)} \sim q_s(x|x_{s-1}^{(n)}, y_s)$
    compute the importance weights
    $\widetilde{w}_s^{(n)} = \frac{p(y_s|x_s^{(n)})p(x_s^{(n)}|x_{s-1}^{(n)})}{q_s(x_s^{(n)}|x_{s-1}^{(n)}, y_s)} w_{s-1}{}^{(n)}$
  end
  normalize the importance weights
  $w_s^{(n)} = \frac{\widetilde{w}_s^{(n)}}{\sum_{i=1}^{N} \widetilde{w}_s^{(n)}} \qquad \forall n = 1, \ldots N$
end

Algorithm 1: Sequential Importance Sampling for $p(x_{0:t}|y_{1:t})$.

With increasing dimensionality of the target distribution, however, the weights degenerate: a small number of particles are assigned relatively large weights and most of the particles have weight zero. To overcome weight degeneracy, resampling steps can be inserted in order to re-rejuvenate the sequential sample $\left\{x_{0:s}^{(n)}, w_s^{(n)}\right\}_{n=1}^{N}$ for $s = 1, \ldots t$.

The sequential importance re-sampling algorithm (Arulampalam et al., 2002), introduces a resampling step at each time-update of the algorithm, effectively obtaining equally-weighted samples from the target distribution as reported in Algorithm 2. Hence, the importance weights

of this algorithm are computed anew at every time step, independently from the weights at the previous time step.

In the resampling step, the notation $x_s^{(n)} \sim \mathcal{C}\left\{x_s^{(i)}, w_s^{(i)}\right\}_{i=1}^{N}$ denotes the sampling of a particle $x_s^{(n)}$ from a discrete distribution that takes values $x_s^{(1)}, x_s^{(2)}, \ldots, x_s^{(i)}, \ldots, x_s^{(N)}$ with respective weights $w_s^{(1)}, w_s^{(2)}, \ldots, w_s^{(N)}$. I.e. X_s is distributed as a sum of Dirac point-mass variables:

$$p(x_s^{(n)}) = \sum_{i=1}^{N} \delta_{x_s^{(i)}}(x) w_s^{(i)}.$$

Result: $\left\{x_{0:t}^{(n)}\right\}_{n=1}^{N}$: a uniform sample from $p(x_{0:t}|y_{1:t})$
Input: N,$q_0(x)$, $q_t(x)$, $p(x_0)$, $p(x_t|x_{t-1})$, $p(y_t|x_t)$ for $t = 0, 1, \ldots t$
for $n = 1, \ldots N$ **do**
  sample
  $x_0{}^{(n)} \sim q_0(x)$
  compute the importance weights
  $\widetilde{w}_0^{(n)} = \frac{p(x_0^{(n)})}{q_0(x_0^{(n)})}$
end
normalize the importance weights
$w_0^{(n)} = \frac{\widetilde{w}_0^{(n)}}{\sum_{i=1}^{N} \widetilde{w}_0^{(n)}} \qquad \forall n = 1, \ldots N$
resample the particles with importance weights
$x_0^{(n)} \sim \mathcal{C}\left\{x_0^{(n)}, w_0^{(n)}\right\}_{n=1}^{N} \qquad \forall n = 1, \ldots N$
set equal weights
$w_0^{(n)} = \frac{1}{N} \qquad \forall n = 1, \ldots N$
for $s = 1, \ldots t$ **do**
  for $n = 1, \ldots N$ **do**
    sample
    $x_s{}^{(n)} \sim q_s(x|x_{s-1}^{(n)}, y_s)$
    compute the importance weights
    $\widetilde{w}_s^{(n)} = \frac{p(y_s|x_s^{(n)}) p(x_s^{(n)}|x_{s-1}^{(n)})}{q_s(x_s^{(n)}|x_{s-1}^{(n)}, y_s)} \widetilde{w}_{s-1}^{(n)}$
  end
  normalize the importance weights
  $w_s^{(n)} = \frac{\widetilde{w}_s^{(n)}}{\sum_{i=1}^{N} \widetilde{w}_s^{(n)}} \qquad \forall n = 1, \ldots N$
  resample the particles with importance weights
  $x_s^{(n)} \sim \mathcal{C}\left\{x_s^{(i)}, w_s^{(i)}\right\}_{i=1}^{N} \qquad \forall n = 1, \ldots N$
  set equal weights
  $w_s^{(n)} = \frac{1}{N} \qquad \forall n = 1, \ldots N$
end

Algorithm 2: Sequential Importance Re-sampling for $p(x_{0:t}|y_{1:t})$.

4.2.2 Bootstrap Particle Filter

The BPF is a sequential importance re-sampling algorithm where specific choices of the importance distribution (and hence of the weights) are made. This algorithm was first introduced by Stewart and McCarthy Jr (1992) and by Gordon, Salmond, and Smith (1993), taking for the first time its current name. The BPF targets the joint filtering distribution $p(x_{0:t}|y_{1:t})$.

The key idea of the algorithm is to generate a set of N particles and apply three steps sequentially (over times $t = 1, 2, \dots$):

- *resample* to obtain a equally-weighted sample from the target distribution at $t-1$:
$$\left\{x_{0:t-1}^{(n)}, \frac{1}{N}\right\}_{n=1}^{N}$$
- *propagate* this sample from the importance distribution, chosen to be the state equation of the SSM:
$$q_t(x_t|x_{t-1}, y_t) = p(x_t|x_{t-1})$$
obtaining the sample:
$$\left\{x_{0:t-1}^{(n)}, x_t^{(n)}\right\}_{n=1}^{N}$$
- *weight* the proposed sample according to the target and importance density:
$$\omega_t = \frac{p(y_s, x_s^{(n)}|x_{s-1}^{(n)})}{q_t(x_t|x_{t-1}, y_t)} = \frac{p(y_s|x_s^{(n)})p(x_s^{(n)}|x_{s-1}^{(n)})}{p(x_s^{(n)}|x_{s-1}^{(n)})} = p(y_s|x_s^{(n)}).$$

Hence, the BPF is simply a sequential importance re-sampler having the state equation as importance distribution. The full pseudo-code of the BPF is reported in Algorithm 3.

Result: $\left\{x_{0:t}^{(n)}, w_t^{(n)}\right\}_{n=1}^{N}$: a weighted sample from $p(x_{0:t}|y_{1:t})$
Input: N, $p(y_t|x_t)$, $p(x_t|x_{t-1})$, for $t = 0, 1, \dots t$
for $n = 1, \dots N$ **do**
  sample
  $x_0{}^{(n)} \sim p(x_0)$
  propagate: sample
  $x_1{}^{(n)} \sim p(x_1|x_0{}^{(n)})$
  compute the importance weights
  $\widetilde{w}_0^{(n)} = \frac{p(y_1|x_1^{(n)})p(x_1^{(n)}|x_0{}^{(n)})p(x_0^{(n)})}{p(x_1^{(n)}|x_0{}^{(n)})p(x_0^{(n)})} = p(y_1|x_1^{(n)})$
end
normalize the importance weights
$w_0^{(n)} = \frac{\widetilde{w}_0^{(n)}}{\sum_{n=1}^{N}\widetilde{w}_0^{(n)}} \qquad \forall n = 1, \dots, N$
for $s = 1, \dots t$ **do**
  for $n = 1, \dots N$ **do**
    resample to obtain a uniformly weighted sample
    $x_{s-1}{}^{(n)} \sim \mathcal{C}\left\{x_{0:s-1}^{(i)}, w_s^{(i)}\right\}_{i=1}^{N}$
    propagate according to the state equation
    $x_s{}^{(n)} \sim p(x|x_{s-1}^{(n)}, y_s)$
    compute the importance weights
    $\widetilde{w}_s^{(n)} = \frac{p(y_s|x_s^{(n)})p(x_s^{(n)}|x_{s-1}^{(n)})}{p(x_s^{(n)}|x_{s-1}^{(n)})} = p(y_s|x_s^{(n)})$
  end
  normalize the importance weights
  $w_s^{(n)} = \frac{\widetilde{w}_s^{(n)}}{\sum_{n=1}^{N}\widetilde{w}_s^{(n)}} \qquad \forall n = 1, \dots, N$
end

Algorithm 3: Bootstrap Particle Filter for $p(x_{0:t}|y_{1:t})$.

There are many flavours that can be added to enhance the basic BPF, such as the use of auxiliary variables (Pitt and Shephard, 1999) that improves matching between the importance and target distributions. However, over this thesis, only the algorithms presented in this chapter have been adopted.

4.3 Methods for parameter inference

Subsection 4.3.1 illustrates how SMC algorithms, such as the sequential importance sampler, resampler and the BPF, provide a method to approximate the likelihood, i.e. the data distribution conditional on a parameter value $\boldsymbol{\theta}$. The question on how this approximated likelihood can be used in an estimation context is addressed in Subsection 4.3.2. Since the focus of this section is the inference of the static parameter, $\boldsymbol{\theta}$ is reintroduced in the notation below.

4.3.1 Approximation of the likelihood via SMC

Given the Markovianity of the model and the conditional independence of the observations, the data distribution can be factorised into single-time-step densities

$$p(y_{1:T}|\boldsymbol{\theta}) = \prod_{t=1}^{T} p(y_t|y_{1:t-1}, \boldsymbol{\theta}) \tag{4.7}$$

where each element of the likelihood can be seen as the marginalization over the hidden states of the joint distribution, and further simplified thanks to conditional independence:

$$\begin{aligned} p(y_t|y_{1:t-1}, \boldsymbol{\theta}) &= \int_{X_{0:t}} p(y_t, x_{0:t}|y_{1:t-1}, \boldsymbol{\theta}) \, \mathrm{d}x_{0:t} \\ &= \int_{X_{0:t}} p(y_t, x_t|x_{0:t-1}, y_{1:t-1}, \boldsymbol{\theta}) p(x_{0:t-1}|y_{1:t-1}, \boldsymbol{\theta}) \, \mathrm{d}x_{0:t} \\ &= \int_{X_{0:t}} p(y_t, x_t|x_{t-1}, \boldsymbol{\theta}) p(x_{0:t-1}|y_{1:t-1}, \boldsymbol{\theta}) \, \mathrm{d}x_{0:t} \end{aligned}$$

In the final line, the numerator of a sequential importance re-sampler weight can be recognised. Hence, the integrand can be rewritten as:

$$p(y_t|y_{1:t-1}, \boldsymbol{\theta}) = \int_{X_{0:t}} \omega(x_t, x_{t-1}, y_t) \underbrace{q_t(x_t|x_{t-1}, y_t) p(x_{0:t-1}|y_{1:t-1}, \boldsymbol{\theta})}_{*} \mathrm{d}x_{0:t}$$

In a sequential importance re-sampler (e.g. in the BPF) the trajectories $\left\{x_{0:t}^{(n)}\right\}_{n=1}^{N} = \left\{\breve{x}_t^{(n)}, x_{0:t-1}^{(n)}\right\}_{n=1}^{N}$ are sampled from distribution $*$. This means that the integral can be approximated by vanilla MC using the observed weights:

$$\begin{aligned} &\approx \frac{1}{N} \sum_{n=1}^{N} \omega(x_t^{(n)}, x_{t-1}^{(n)}, y_t) \\ &\approx \frac{1}{N} \sum_{n=1}^{N} \widetilde{w}_t^{(n)} \end{aligned} \tag{4.8}$$

Equation 4.8 can be plugged into 4.7, leading to an approximation of the likelihood of the data conditionally on the parameter $\boldsymbol{\theta}$, i.e.:

$$\begin{aligned} p(y_{1:T}|\boldsymbol{\theta}) &= \prod_{t=1}^{T} p(y_t|y_{1:t-1}, \boldsymbol{\theta}) \\ &\approx \prod_{t=1}^{T} \frac{1}{N} \sum_{n=1}^{N} \widetilde{w}_t^{(n)} \end{aligned} \tag{4.9}$$

Finally, it is easy to see that this approximation, for the BPF takes the following form:

$$p(y_{1:T}|\boldsymbol{\theta}) \approx \prod_{t=1}^{T} \frac{1}{N} \sum_{n=1}^{N} \widetilde{w}_t^{(n)} = \prod_{t=1}^{T} \frac{1}{N} \sum_{n=1}^{N} p(y_s|x_s^{(n)})$$

4.3.2 Inference with the approximated likelihood

There is not a unique way in which this approximated likelihood can be used to drive inference on the parameter $\boldsymbol{\theta}$.

Methods belonging to the non-Bayesian literature aim at obtaining the maximum-likelihood estimate and provide point and interval estimates. These methods include Multiple Iterated Filtering (Ionides, Bretó, and King, 2006), as well grid-search methods.

Within the Bayesian framework, many iterative algorithms have been developed to sample from the posterior distribution of interest $(\boldsymbol{\Theta}|y_{1:t})$ when only an approximation of the likelihood is available. Andrieu, Doucet, and Holenstein (2010) reviews and summarizes the algorithms used more frequently for parameter inference in SSMs. This paper combines SMC methods with Markov chain Monte Carlo (MCMC) methods in a non-trivial and non-standard way, which often includes exploration of the distribution of the latent states. Among the algorithms proposed, *pseudo-marginal approaches* can be found: these were introduced previously in Andrieu and Roberts (2009) and provide a simple way to integrate SMC-based approximation of the likelihood into MCMC algorithms for Bayesian inference.

Pseudo-marginal algorithms are aimed at exploring only the posterior distribution of the parameter, marginally from the distribution of the states, and they are based on the classical Metropolis Hastings (MH) algorithm (Hastings, 1970; Metropolis et al., 1953). Nevertheless, differently from the original MH algorithm, here the unnormalised posterior distribution is approximated by the product of the prior and a SMC approximation of the likelihood in the acceptance ratio.

Two algorithms are employed throughout this thesis: grouped independence Metropolis Hastings (GIMH) (Beaumont, 2003) and Monte Carlo within Metropolis (MCWM) (Andrieu and Roberts, 2009). In the former, at iteration i, when a new value $\boldsymbol{\theta}'$ is proposed, an SMC algorithm is run to estimate the likelihood $\widehat{p}(\boldsymbol{\theta}')$, which is plugged into the numerator of the acceptance ratio, together with prior and proposal density. The denominator is composed of the previously-retained estimated likelihood for the initial parameter $\widehat{p}(\boldsymbol{\theta}^i)$ and the respective prior and proposal density. Upon acceptance, the proposed parameter $\boldsymbol{\theta}'$ and its estimated likelihood $\widehat{p}(\boldsymbol{\theta}')$ are retained; upon rejection the old parameter $\boldsymbol{\theta}^i$ and its likelihood $\widehat{p}(\boldsymbol{\theta}^i)$ are retained. On the other hand, the MCWM algorithm re-approximates the likelihood of the parameter $\boldsymbol{\theta}^i$ when computing the acceptance ratio instead of storing it and re-using it every time. GIMH and MCWM are reported and compared in algorithm 4.

GIMH was proved an exact algorithm in Andrieu and Roberts (2009), targeting the exact posterior distribution. Moreover, the authors showed that, despite MCWM being biased for small approximation size N, this bias decreases and becomes irrelevant as N increases (McKinley et al., 2014). If the likelihood approximation is precise enough, the two algorithms would perform equally well.

4.4 An application of SMC to data with high missingness

Chapter 2 explored many candidate models and inferential methods to estimate influenza severity as measured by the hospitalised case fatality risk (hCFR) or the intensive-care case fatality risk (iCFR). Nevertheless, none of the estimators proposed are applicable to a situation where only data on the hospital admissions (or Intensive Care Unit (ICU) admissions) and deaths

Result: $\left\{\boldsymbol{\theta}^{(n)}\right\}_{n=1}^{N}$: a sample from $p(\boldsymbol{\theta}|y_{1:t})$
Input: B iterations; N approximation size ; $\widehat{p}(y_{1:T}|\boldsymbol{\theta})$, e.g. estimate 4.9; $k(\cdot|\boldsymbol{\theta})$ transition kernel; $\boldsymbol{\theta}$ initialization

for $i = 1, \dots B$ **do**

sample
$\boldsymbol{\theta}' \sim k(\cdot|\boldsymbol{\theta}^i)$

generate from a SMC algorithm with N particles

GIMH	MCWM
$\widehat{p}(y_{1:T}\|\boldsymbol{\theta}')$	$\widehat{p}(y_{1:T}\|\boldsymbol{\theta}')$
-	$\widehat{p}(y_{1:T}\|\boldsymbol{\theta}^i)$

sample
$u \sim \text{Uniform}(0, 1)$

compute
$\alpha = \min\left[1, \frac{\widehat{p}(y_{1:T}|\boldsymbol{\theta}')\widehat{p}(\boldsymbol{\theta}')k(\boldsymbol{\theta}^i|\boldsymbol{\theta}')}{\widehat{p}(y_{1:T}|\boldsymbol{\theta}^i)\widehat{p}(\boldsymbol{\theta}^i)k(\boldsymbol{\theta}'|\boldsymbol{\theta}^i)}\right]$

if $u < \alpha$ **then**

set

GIMH	MCWM
$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}'$	$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}'$
$\widehat{p}(y_{1:T}\|\boldsymbol{\theta}^{i+1}) = \widehat{p}(y_{1:T}\|\boldsymbol{\theta}')$	-

else

set

GIMH	MCWM
$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^i$	$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^i$
$\widehat{p}(y_{1:T}\|\boldsymbol{\theta}^{i+1}) = \widehat{p}(y_{1:T}\|\boldsymbol{\theta}^i)$	-

end

end

Algorithm 4: GIMH and MCWM algorithms

counts are available, with no information on recoveries/discharges. This subsection exploits the SSM setting and SMC methods to address this problem.

4.4.1 Formulation of the problem

Severity estimators based only on the cumulative counts of ICU admissions and deaths, such as the one adopted by the World Health Organization (WHO) are biased, since their denominator includes people whose survival outcome is unknown or censored, due to not having died nor recovered yet.

Unfortunately, in many settings, information on the number of individuals who have not had the severe event, in this application death, might not be available. This is the case, for example, of seasonal influenza in the UK: severity is monitored via the UK Severe Influenza Surveillance System (USISS), that records the weekly counts of confirmed-influenza ICU admissions and deaths, but not the discharges. Despite these data being very representative (they are a census and should record all the events), no information is available on the recovery process

and therefore only biased estimators can be obtained.

Similarly to the formulation of the severity estimator based on counting processes (of Section 2.5), it is useful to see severity as a MSM and consequently as a SSM.

At an individual level every subject, once admitted to hospital/ICU will eventually either recover or die after some waiting time. If the related counting process is observed in discrete time, during each interval those in hospital might either recover, die or remain in the hospital for more time.

Taking weekly time intervals, indexed by $t = 1, \dots, T$, where T is the end of the epidemic, denote by h_t the number of confirmed influenza admissions to ICU during interval t (which is deterministic and known at time t); denote with X_t, Y_t and Z_t the r.v.s representing the number of patients with confirmed influenza remaining in ICU, dying and being discharged/recovering during interval t, respectively. Denote with ICU_t the number of confirmed influenza patients present in ICU at the end of interval t. Note that ICU_t is just a (deterministic) transformation of the r.v. X_t since $ICU_t = X_t + h_t$.

The state dynamics assumed are the same as Section 2.5. At every interval t, the number of people who recover, die and stay in ICU (Z_t, Y_t and X_t respectively) are a Multinomial sample of the number of people in ICU at the end of the previous interval (icu_{t-1}). Denote the probability of recovery/discharge and death respectively during an interval by the static parameter $\boldsymbol{\theta} = \{p_{\mathrm{R}}, p_{\mathrm{D}}\}$.

$$(Z_t, Y_t, X_t | x_{t-1}) \sim \text{Multi}(\text{n}{=}icu_{t-1} = x_{t-1} + h_{t-1}, \text{p}{=}(p_{\mathrm{R}}, p_{\mathrm{D}}, (1 - p_{\mathrm{D}} - p_{\mathrm{R}}))) \tag{4.10}$$

The r.v. Y_t, i.e. the weekly count of deaths, is observed exactly, so the SSM can be defined via the joint distribution of the unobserved and observed processes, where the observation process depends on the previous value of the unobserved process as in Equation 4.11, differently from the general notation in (4.1):

$$Z_t, Y_t, X_t | (X_{t-1} = x_{t-1}, \boldsymbol{\Theta} = \boldsymbol{\theta}) \sim p(z_t, y_t, x_t | x_{t-1}, \boldsymbol{\theta}) \tag{4.11}$$

where $p(\cdot | x_{t-1}, \boldsymbol{\theta})$ is the probability mass function of a Multinomial r.v. with probability vector $\boldsymbol{\theta} = (p_{\mathrm{D}}, p_{\mathrm{R}})$ and size $x_{t-1} + h_{t-1}$. Lastly, let the system at $t = 0$ have $Y_0 = 0$, $Z_0 = 0$, while the initial number of people in hospital with confirmed influenza X_0 is Poisson distributed with known rate 1.5:

$$X_0 \sim \text{Pois}(1.5)$$

The SSM is represented in Figure 4.2.

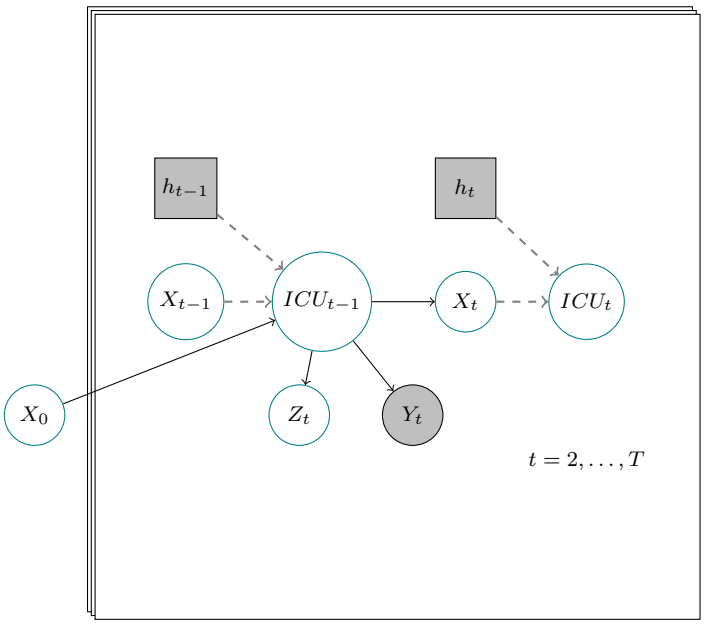


Figure 4.2: Graphical model for the chain Multinomial dynamics. Circles represent r.v.s and squares represent deterministic quantities. Gray-filled nodes represent observed quantities. Solid arrows represent random relations and dashed arrows represent deterministic relations.

4.4.2 Proposal of an SMC method for the approximation of the likelihood

To obtain estimates of the system parameters marginalizing over the hidden (recovery) states, a sequential importance sampler is used to approximate the likelihood of the data given a parameter value $\boldsymbol{\theta}$, as reported in Algorithm 5.

The importance distribution $q_{1:T}$ for all the hidden states $x_{1:T}$ is chosen to be factorisable in a Markovian fashion, so that it is composed of conditional distributions on the previous hidden state x_{t-1}. At each time point, the importance distribution q_t depends on: the previous state of the system $\bar{x}_{t-1}^{(n)}$ (sampled at the previous step); the known number of admissions h_{t-1}; and the data y_t, and it is chosen to have Binomial distribution:

$$q_t(x_t^{(n)}|y_t, x_{t-1}^{(n)}) = \text{Binom}\left(\text{n}=(x_{t-1}^{(n)} + h_{t-1} - y_t), \text{p}= \left(1 - \frac{p_\text{R}}{1 - p_\text{D}}\right)\right)$$

This choice is made so that the importance distribution at time t covers the same space as the likelihood distribution.

The target density is the SSM defined in Equation 4.10.

Algorithm 5 proposes the hidden state of the number of people that remain in hospital every week; these particles are re-sampled according to the state and observation densities and the weights are used to approximate the likelihood of parameter $\boldsymbol{\theta} = (p_\text{D}, p_\text{R})$.

The proposed algorithm was tested on simulated data. ICU admissions from a seasonal SEEIIR epidemic were simulated; the consequent death and recovery events were also simulated assuming parameters $\boldsymbol{\theta} = (p_\text{D} = 0.2, p_\text{R} = 0.4)$. The complete dataset (including the hidden number of people recovering/staying in the hospital) is plotted in Figure 4.3.

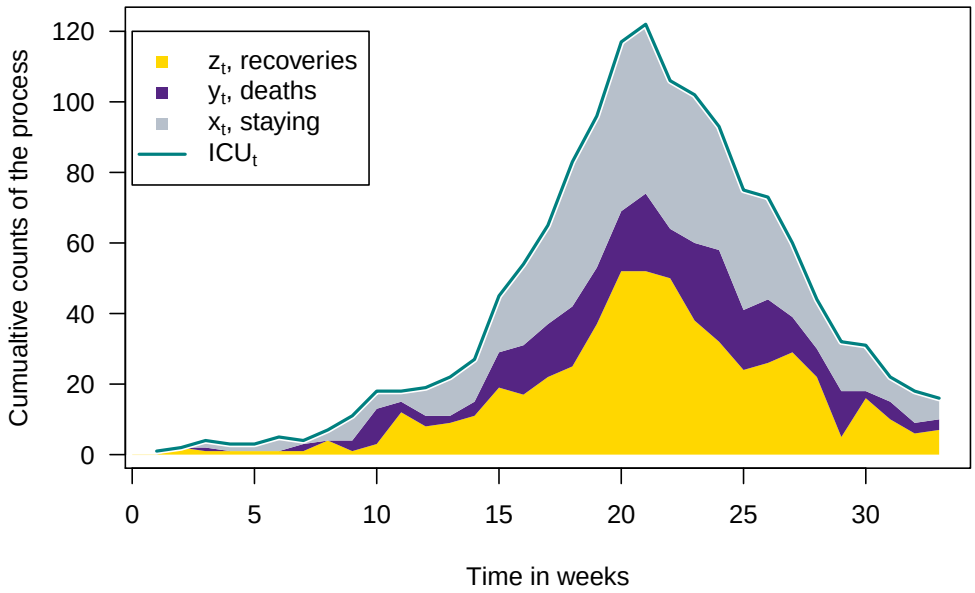


Figure 4.3: Simulated data, cumulative counts

input : data vector $y_{1:T}$; $\boldsymbol{\theta} = (p_{\text{D}}, p_{\text{R}})$; number of particles N

output: an approximation of the likelihood $p(y_{1:T}|\boldsymbol{\theta})$

for $n = 1, \ldots, N$ **do**

$x_0^{(n)} \sim \text{Pois}(1.5)$

initialize the weights:

$w_0^{(n)} = \frac{1}{N}$

propagate:

$x_1^{(n)}|x_0^{(n)}, y_1 \sim \text{Binom}\left(\text{n}=(x_0^{(n)} + h_0 - y_1), \text{p}=\left(1 - \frac{p_{\text{R}}}{1-p_{\text{D}}}\right)\right)$

compute the weights:

$\widetilde{w}_1^{(n)} = \frac{p(z_1^{(n)}, y_1, x_1^{(n)}|x_0^{(n)})}{q_t(x_1^{(n)}, y_1|x_0^{(n)})}$

end

normalize the weights

$w_1^{(n)} = \frac{\widetilde{w}_1^{(n)}}{\sum_{n=1}^N \widetilde{w}_1^{(n)}}$ for $n = 1, \ldots, N$

set the weighted sample from the target density to be:

$\left\{x_{0:1}^{(n)}, w_1^{(n)}\right\}_{n=1}^N$

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

for $n = 1, \ldots, N$ **do**

resample

$x_{1:t-1}^{(n)} \sim \mathcal{C}\left\{x_{0:t-1}^{(j)}, w_{t-1}^{(j)}\right\}_{j=1}^N$

re-set the weights

$w_{t-1}^{(n)} = \frac{1}{N}$

propagate

$x_t^{(n)}|x_{t-1}^{(n)}, y_t \sim \text{Binom}\left(\text{n}=(x_{t-1}^{(n)} + h_{t-1} - y_t), \text{p}=\left(1 - \frac{p_{\text{R}}}{1-p_{\text{D}}}\right)\right)$

compute the weights:

$\widetilde{w}_t^{(n)} = \frac{p(z_t^{(n)}, y_t, x_t^{(n)}|x_{t-1}^{(n)})}{q_t(x_t^{(n)}, y_t|x_{t-1}^{(n)})} w_{t-1}^{(n)}$

end

normalize the weights

$w_t^{(n)} = \frac{\widetilde{w}_t^{(n)}}{\sum_{n=1}^N \widetilde{w}_t^{(n)}}$ for $n = 1, \ldots, N$

set the weighted sample from the target density to be:

$\left\{x_{0:t-1}^{(n)}, x_t^{(n)}, w_t^{(n)}\right\}_{n=1}^N$

and denote it by:

$\left\{x_{0:t}^{(n)}, w_t^{(n)}\right\}_{n=1}^N = \left\{x_{0:t-1}^{(n)}, x_t^{(n)}, w_t^{(n)}\right\}_{n=1}^N$

end

approximate the likelihood by $p(y_{1:T}|\boldsymbol{\theta}) \approx \prod_{t=1}^T \frac{1}{N} \sum_{n=1}^N \widetilde{w}_t^{(n)}$

Algorithm 5: Sequential importance re-sampler to approximate the likelihood of severe data.

A run of the algorithm at the data-generating parameters is plotted in Figure 4.4. The importance and filtering particles are plotted, together with the true, latent trajectory of the process that they approximate.

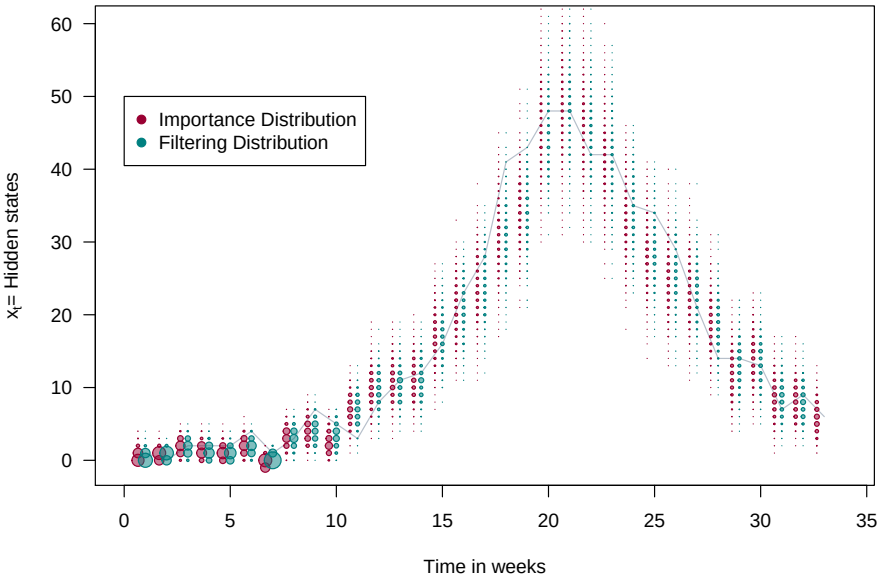
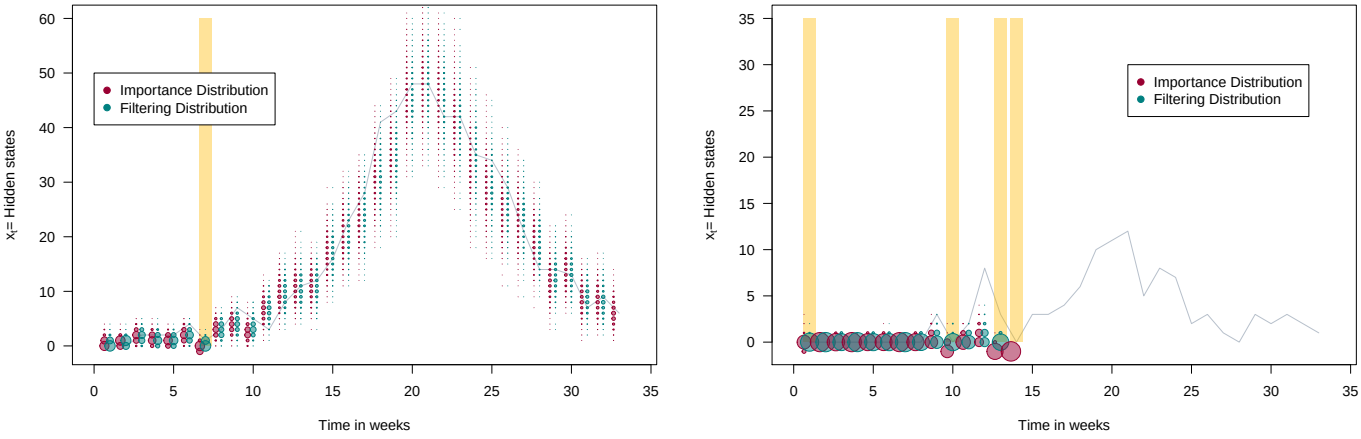


Figure 4.4: Importance distribution (red) and filtering distribution(green) of the hidden states x_t for $t = 1, \ldots, T$ derived with Algorithm 5 with N = 1000 particles. The grey line is the true trajectory $x_{0:t}$.

This algorithm however, has a peculiar pitfall: many simulated values might have weight 0. This happens mostly because the state and observational process is discrete and bounded (being Multinomial). It might happen that some of the simulated trajectories have probabilities 0, since, within their simulated history, the sampled number of people remaining in ICU is lower than the observed number of deaths at a following time step (i.e. $\exists\ t$ s.t. $icu_{t-1} < y_t$). This event happens more frequently as the simulated ICU cases become smaller: Figure 4.5 highlights in yellow time points where a single particle fails and shows a case where all the particles used in Algorithm 5 have probability 0.

These *particle failure* events happen more frequently when the counts take low values and therefore they pose tighter constraints; and when the parameters tested are different from the data-generating parameters.



(a) (b)

Figure 4.5: Importance distribution (red) and filtering distribution(green) of the hidden states x_t for $t = 1, \dots, T$ derived with Algorithm 5 with N = 1000 particles. The grey line is the true trajectory $x_{0:t}$. The vertical yellow bars denote the times when one or more trajectory has probability 0: in the panel (b) all the sampled trajectories have probability 0 at time $t = 14$.

Rather than being a pitfall of the sequential importance sampler, this issue is caused by the SSM defined, which is based on discrete and bounded r.v.s and is highly observed. These characteristics makes the application of SMC methods non-trivial.

4.4.3 An extension to avoid particle failure

To address the issue of failing particles, an alternative version of Algorithm 5 is formulated.

The key idea consists of preventing a complete failure of all the particles by including at least one case that always leads to a strictly positive weight. This is the case when everyone who does not die stays in the hospital: in this case there is no possibility for the trajectory to propose too many individuals recovering.

This is done by setting deterministically one particle to take value $x_t = x_{t-1} + h_{t-1}$, this will be a *life belt* in case all the others fail. To do so, a new sampling method is introduced.

Deterministic mixture importance sampling

Deterministic mixture importance sampling was first introduced in Veach and Guibas, 1995 and re-proposed in Owen and Zhou, 1998, together with proofs of unbiasedness. In a more recent paper (Cornuet et al., 2012), the same importance sampler is reported in its essential elements.

The framework is the same as a classical importance sampler (hence there is no reference to the sequential aspect and the indexes t are dropped), targeting a distribution $f(x)$ from which direct sampling is impossible and that can be evaluated point-wise up to a proportionality constant: $f(x) = \frac{\widetilde{f}(x)}{z}$. Instead of choosing an importance distribution $q(x)$, results from different importance sampling distributions $q_1(x), q_2(x), \dots, q_G(x)$ are pooled together so that the sample

is drawn as follows:

$$\begin{aligned} x_1^{(n)} &\sim q_1(x) \qquad \text{for } n = 1, \dots, N_1 \\ x_2^{(n)} &\sim q_2(x) \qquad \text{for } n = 1, \dots, N_2 \\ &\dots \\ x_G^{(n)} &\sim q_G(x) \qquad \text{for } n = 1, \dots, N_G \end{aligned}$$

Let N be the sum of the importance samples: $N = \sum_{g=1}^{G} N_g$. Owen and Zhou, 1998 propose to pool the samples together recomputing the weights as if the total sample was drawn from a mixture of all the densities $q_1(x), q_2(x), \dots, q_G(x)$:

$$\widetilde{q}(x^{(n)}) = \frac{1}{N} \sum_{g=1}^{G} N_g q_g(x_g^{(n)})$$

resulting in the deterministic mixture weight:

$$w_g^{(n)} = \frac{\widetilde{f}(x_g^{(n)})}{\frac{1}{N} \sum_{l=1}^{G} N_l q_l(x_g^{(n)})}$$

Despite the importance samples not being drawn from a mixture distribution, these weights allow exact MC estimation.

Consider the estimator $\hat{I}$ for the expected value μ of a function $h(x)$ where x has probability density function $f(x)$ ($\mu = \mathbb{E}[h(x)] = \int h(x) f(x) \mathrm{d}x$) reported below

$$\hat{I} = \frac{1}{N} \sum_{g=1}^{G} \sum_{n=1}^{N_g} h\left(x_g^{(n)}\right) \frac{f(x_g^{(n)})}{\frac{1}{N} \sum_{l=1}^{G} N_l q_l\left(x_g^{(n)}\right)}$$

The expected value is:

$$\begin{aligned} \mathbb{E}\left[\hat{I}\right] &= \mathbb{E}\left[\frac{1}{N} \sum_{g=1}^{G} \sum_{n=1}^{N_g} h\left(x\right) \frac{f(x)}{\frac{1}{N} \sum_{l=1}^{G} N_l q_l\left(x\right)}\right] \\ \mathbb{E}\left[\hat{I}\right] &= \frac{1}{N} \sum_{g=1}^{G} \sum_{n=1}^{N_g} \mathbb{E}\left[h\left(x\right) \frac{f(x)}{\frac{1}{N} \sum_{l=1}^{G} N_l q_l\left(x\right)}\right], \qquad \text{for linearity of } \mathbb{E} \end{aligned}$$

The N_g samples are drawn independent and identically distributed (iid) from the g-th importance distribution, hence:

$$\begin{aligned} \mathbb{E}\left[\hat{I}\right] &= \frac{1}{N} \sum_{g=1}^{G} N_g \int h\left(x\right) \frac{f(x)}{\frac{1}{N} \sum_{l=1}^{G} N_l q_l\left(x\right)} q_g\left(x\right) \mathrm{d}x \\ &= \sum_{g=1}^{G} N_g \int h\left(x\right) \frac{f(x)}{\sum_{l=1}^{G} N_l q_l\left(x\right)} q_g\left(x\right) \mathrm{d}x \\ &= \int h\left(x\right) f(x) \sum_{g=1}^{G} N_g \frac{q_g\left(x\right)}{\sum_{l=1}^{G} N_l q_l\left(x\right)} \mathrm{d}x \\ &= \int h\left(x\right) f(x) \mathrm{d}x = \mu \end{aligned}$$

and exactness of the estimator is proved. Cornuet et al., 2012 underlines that the samples obtained are valid importance-sampling approximation for the target f if every sub-sample of

size N_g is a valid importance sample, i.e. if the support of q_g contains the support of f. In this case, the deterministic mixture importance sampling can be seen as a method that simply *pools* importance-sample estimators obtained from many different importance distributions.

However there is at least one case for which it is not strictly necessary for the support of each of the importance distributions to cover the support of the target distribution. This is the case when the q_gs are defined on a partition of the domain of f, Ω, i.e. the support of q_g is Ω_g, for $g = 1, \ldots, G$, with $\Omega = \bigcup_{g=1}^{G} \Omega_g$ and $\Omega_g \bigcap \Omega_h = \varnothing \forall g \neq h$. In this case, the sum at the denominator of the weight can be simplified since $q_g(x_h^{(n)}) = 0 \ \forall g \neq h$, hence:

$$\hat{I} = \frac{1}{N} \sum_{g=1}^{G} \sum_{n=1}^{N_g} h\left(x_g^{(n)}\right) \frac{f(x_g^{(n)})}{\frac{1}{N} \sum_{l=1}^{G} N_l q_l\left(x_g^{(n)}\right)} = \frac{1}{N} \sum_{g=1}^{G} \sum_{n=1}^{N_g} h\left(x_g^{(n)}\right) \frac{f(x_g^{(n)})}{\frac{N_g}{N} q_l\left(x_g^{(n)}\right)}$$

The estimator unbiasedness can be derived similarly to before

$$\begin{aligned}
\mathbb{E}\left[\hat{I}\right] &= \mathbb{E}\left[\frac{1}{N} \sum_{g=1}^{G} \sum_{n=1}^{N_g} h(x) \frac{f(x)}{\frac{N_g}{N} q_g(x)}\right] \\
&= \sum_{g=1}^{G} \frac{1}{N_g} \sum_{n=1}^{N_g} \mathbb{E}\left[h(x) \frac{f(x)}{q_g(x)}\right] \\
&= \sum_{g=1}^{G} \int_{\Omega} \frac{f(x) h(x)}{q_g(x)} q_g(x) \, \mathrm{d}x
\end{aligned}$$

Before simplification, the integral can be separated in all the disjoint sets of the partition for Ω:

$$\begin{aligned}
\mathbb{E}\left[\hat{I}\right] &= \sum_{g=1}^{G} \int_{\Omega} \frac{f(x) h(x)}{q_g(x)} q_g(x) \, \mathrm{d}x \\
&= \sum_{g=1}^{G} \left[\int_{\Omega_1} \frac{f(x) h(x)}{q_g(x)} q_g(x) \, \mathrm{d}x + \cdots + \int_{\Omega_l} \frac{f(x) h(x)}{q_g(x)} q_g(x) \, \mathrm{d}x + \cdots + \int_{\Omega_G} \frac{f(x) h(x)}{q_g(x)} q_g(x) \, \mathrm{d}x\right]
\end{aligned}$$

where $q_g(x) = 0$ for all the sets Ω_h with $h \neq g$, hence:

$$\begin{aligned}
\mathbb{E}\left[\hat{I}\right] &= \sum_{g=1}^{G} \int_{\Omega_g} \frac{f(x) h(x)}{q_g(x)} q_g(x) \, \mathrm{d}x \\
&= \sum_{g=1}^{G} \int_{\Omega_g} f(x) h(x) \mathrm{d}x \\
&= \int_{\Omega} f(x) h(x) \mathrm{d}x = \mu.
\end{aligned}$$

Algorithm formulation

Deterministic mixture importance sampling is used at each time step $t = 0, 1, \ldots, T$ assuming $G = 2$ importance distributions having $N_1 = N - 1$ and $N_2 = 1$. The importance distributions considered are:

- $x_t^{(n)} \sim q_1(x_t^{(n)})$ for $n = 1 \ldots, N-1$ where $q_1(x)$ is the same Binomial distribution used in Algorithm 5:

$$x_t^{(n)} | x_{t-1}^{(n)}, y_t \sim \text{Binom}\left(\text{n}=(x_{t-1}^{(n)} + h_{t-1} - y_t), \text{p}=\left(1 - \frac{p_{\text{R}}}{1 - p_{\text{D}}}\right)\right) \tag{4.12}$$

- $x_t^{(n)} \sim q_2(x_t^{(n)})$ for $n = N$ where $q_2(x^N)$ is a Dirac point mass distribution centred in the *life belt* case where all the people who enter the hospital and do not die stay in the hospital:

$$x_t^{(N)} | \overline{x}_{t-1}^{(N)}, y_t \sim \delta_{\overline{x}_{t-1}^{(N)} + h_{t-1} - y_t}(x)$$

with $N_1 = N - 1$ and $N_2 = 1$.
In the system considered, the domain of the hidden states keeps expanding, taking as upper limit the value of the lifebelt particle. Nevertheless, the proposed algorithm composes two distributions that do not fully cover this support. For this reason the current formulation of the algorithm does not lead to an unbiased approximation. Extension to achieve unbiasedness are discussed at the end of the current section.

These importance distributions are used in the full algorithm reported in Algorithm 6.

Algorithm 6 performs well in cases when Algorithm 5 was failing, allowing the approximation of the likelihood of the data given a value for $\boldsymbol{\theta}$, as shown in Figure 4.6.

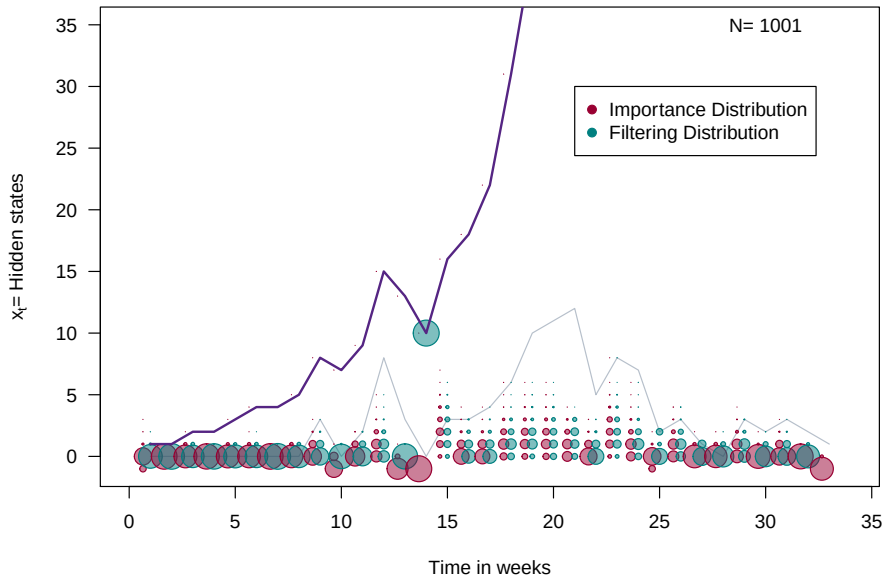


Figure 4.6: Importance distribution (red) and filtering distribution(green) of the hidden states x_t for $t = 1, \ldots, T$ derived with Algorithm 4 with N = 1001 particles one of which is a *life belt* particle whose trajectory is drawn in purple. The grey line is the true trajectory $x_{0:t}$.

input : data vector $y_{1:T}$; $\boldsymbol{\theta} = (p_{\text{D}}, p_{\text{R}})$; number of particles N

output: an approximation of the likelihood $p(y_{1:T}|\boldsymbol{\theta})$

for $n = 1, \ldots, N-1$ **do**

$x_0^{(n)} \sim \text{Pois}(1.5)$

$w_0^{(n)} = \frac{1}{N}$ set the weights

$x_1^{(n)}|x_0^{(n)}, y_1 \sim \text{Binom}\left(\text{n}=(x_0^{(n)} + h_0 - y_1), \text{p}=\left(1 - \frac{p_{\text{R}}}{1-p_{\text{D}}}\right)\right)$ propagate

$\widetilde{w}_1^{(n)} = \frac{p(x_1^{(n)}, y_1, z_1^{(n)}|x_0^{(n)})}{q(x_1^{(n)}, y_1|x_0^{(n)})}$ compute the weights

end

for $n = N$ **do**

$x_0^{(N)} \sim \text{Pois}(1.5)$

$w_0^{(N)} = \frac{1}{N}$ set the weights

$x_1^{(N)}|x_0^{(N)}, y_1 = x_0^{(N)} + h_0 - y_1$ propagate

$\widetilde{w}_1^{(N)} = \frac{p(x_1^{(N)}, y_1, z_1^{(N)}|x_0^{(N)})}{1/N}$ compute the weights

end

normalize the weights

$w_1^{(n)} = \frac{\widetilde{w}_1^{(n)}}{\sum_{n=1}^N \widetilde{w}_1^{(n)}}$ for $n = 1, \ldots, N$

set the weighted sample from the target density to be:

$\left\{x_{0:1}^{(n)}, w_1^{(n)}\right\}_{n=1}^N$

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

for $n = 1, \ldots, N-1$ **do**

$x_{0:t-1}^{(n)} \sim \mathcal{C}\left\{x_{0:t-1}^{(j)}, w_{t-1}^{(j)}\right\}_{j=1}^N$ resample

$x_t^{(n)}|x_{t-1}^{(n)}, y_t \sim \text{Binom}\left(\text{n}=(x_{t-1}^{(n)} + h_{t-1} - y_t), \text{p}=\left(1 - \frac{p_{\text{R}}}{1-p_{\text{D}}}\right)\right)$ propagate

$\widetilde{w}_t^{(n)} = \frac{p(x_t^{(n)}, y_t, z_t^{(n)}|x_{t-1}^{(n)})}{q(x_t^{(n)}, y_t|x_{t-1}^{(n)})(N-1)}$ compute the weights

end

for $n = N$ **do**

$x_{0:t-1}^{(N)} = \left\{x_{0:t-1}^{(N)}, w_{t-1}^{(N)}\right\}$ set the *life belt* trajectory

$x_t^{(N)} = x_{t-1}^{(N)} + h_{t-1} - y_t$ propagate deterministically

$\widetilde{w}_t^{(N)} = \frac{p(x_t^{(N)}, y_t, z_t^{(N)}|x_{t-1}^{(N)})}{1/N} \cdot \widetilde{w}_{t-1}^{(N)}$ compute the weight

end

normalize the weights

$w_t^{(n)} = \frac{\widetilde{w}_t^{(n)}}{\sum_{n=1}^N \widetilde{w}_t^{(n)}}$ for $n = 1, \ldots, N$

set the weighted sample from the target density to be:

$\left\{x_{0:t-1}^{(n)}, x_t^{(n)}, w_t^{(n)}\right\}_{n=1}^N$

and denote it by:

$\left\{x_{0:t}^{(n)}, w_t^{(n)}\right\}_{n=1}^N = \left\{x_{0:t-1}^{(n)}, x_t^{(n)}, w_t^{(n)}\right\}_{n=1}^N$

end

approximate the likelihood by $p(y_{1:T}|\boldsymbol{\theta}) \approx \prod_{t=1}^T \frac{1}{N} \sum_{n=1}^N \widetilde{w}_t^{(n)}$

Algorithm 6: Sequential importance re-sampler to approximate the likelihood of severe data with *life belt* particle.

4.4.4 Performance assessment

To assess whether the likelihood obtained via SMC could help in estimating the iCFR, a small simulation study is conducted.

500 datasets from the chain Multinomial model defined in Equation 4.10 are drawn in different scenarios determined by values of the parameters p_{D} and p_{R}. This model can be re-parametrised as:

$$\begin{cases} \beta_1 = p_{\text{D}} + p_{\text{R}} \\ \beta_2 = \frac{p_{\text{D}}}{p_{\text{D}}+p_{\text{R}}} \end{cases}$$

where β_1 is the probability of either event happening in a week and β_2 is the iCFR.

A grid of values for $(\beta_1; \beta_2)$ is set and the likelihood is approximated via SMC for all the values on the grid. The observed most-likely value of β_2 is then taken as an estimator of iCFR. This value is compared with the WHO estimator:

$$i\widehat{CFR}^{(\text{WHO})} = \frac{\sum_t y_t}{\sum_t h_t}$$

and with the unbiased true estimator:

$$i\widehat{CFR}^{(\text{UNBIASED})} = \frac{\sum_t y_t}{\sum_t y_t + z_t}$$

which cannot be derived with information only on deaths and admissions. This analysis is repeated every 5 weeks from the beginning of the epidemic so that data are censored.

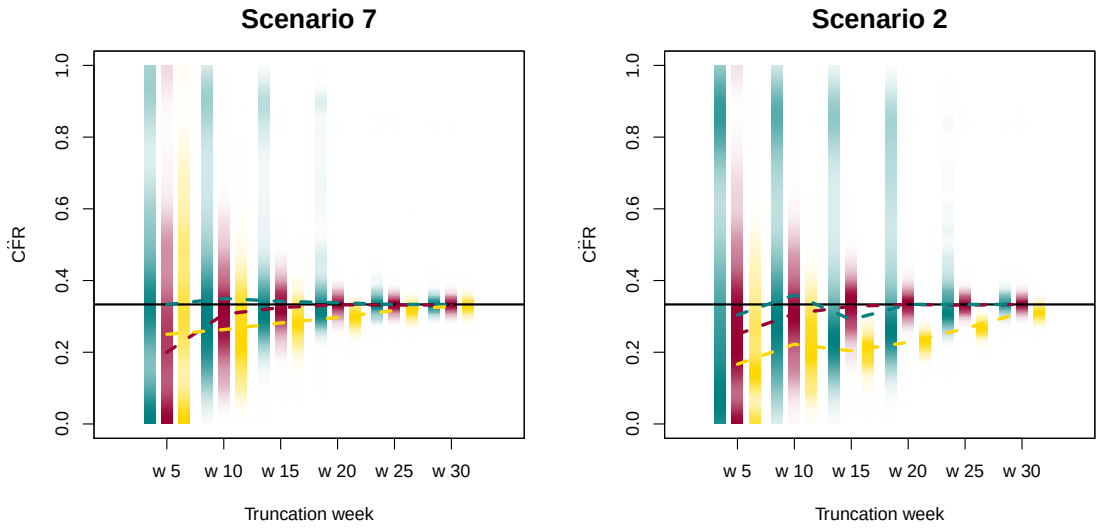
The density-strip plots of the distributions of the three estimators are plotted at the different stages of the analysis (Figure 4.7) for three simulated scenarios. The distribution of the unknown gold-standard estimator $i\widehat{CFR}^{(\text{UNBIASED})}$ is plotted in red, while the distribution of $i\widehat{CFR}^{(\text{WHO})}$ is plotted in yellow and the distribution of the maximum-likelihood estimate according to the SMC approximated likelihood $i\widehat{CFR}^{(\text{SMC})}$ is plotted in green. The WHO estimates underestimate the iCFR especially in the early phase of the epidemic. The estimates based on the SMC likelihood, instead, better reflect in their high variance the fact that there is a lack of information due to censoring, moreover the estimates are often centred in the true value (Panel (a)). This effect is enforced as the censoring effects become bigger due to smaller β_1, with the SMC-based estimate being much more variable than the WHO estimator, accounting for the hidden information (Panel (b)). In a realistic setting where the iCFR is low and half of the cases recover within a week the SMC-based estimate and the WHO estimate are close in median but the WHO estimate is still affected by a misleading low variance, even lower than the one of the unbiased estimate (Panel (c)).

Findings discussion

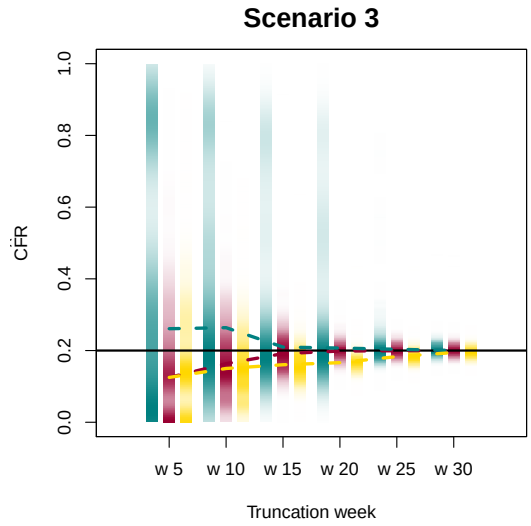
The results presented are promising and show that the estimation of severity could benefit from an SSM perspective and from SMC approaches.

Algorithm 6 addresses the issue of full particle failure proposing one way to prevent this situation. This algorithm could be potentially extended by using more than one *life belt* particle to cover the whole tails of the trajectories space. This extension can be formulated so that the importance approximation is exact, for example by defining a distribution of life belt particles on the complement of the support of the non-life belt importance distribution of Equation 4.12.

Moreover this algorithm could be used in a Bayesian context and combined with external information on the time to any event β_1.



(a) $p_{\text{D}} = 0.2, p_{\text{R}} = 0.4, \beta_1 = 0.6, \beta_2 = 0.33$ (b) $p_{\text{D}} = 0.1, p_{\text{R}} = 0.2, \beta_1 = 0.3, \beta_2 = 0.33$



(c) $p_{\text{D}} = 0.1, p_{\text{R}} = 0.4, \beta_1 = 0.5, \beta_2 = 0.2$

Figure 4.7: Density-strip and median (dashed) of the unbiased unknown estimator (red), of the WHO estimator (gold) and of the SMC-based estimator (green) at censoring points $t = 5, 10, \ldots, 30$ in three different scenarios. The true value is denoted by an horizontal black line.

4.5 Application of SMC to epidemics with multiple data sources

The SSM modelling approach and the consequent inference via SMC can be applied to the evolution and observation of stochastic epidemic models in a natural way. As reported in Chapter 3, these models are usually Markovian, as the distribution of the state at the next time only depends on the current state of the system. In this section a stochastic epidemic model is formulated together with a stochastic severity and detection process. SMC methods are then used to propose an approximation of the likelihood of the full model.

4.5.1 State process

Assume a SEIR stochastic model evolving over discrete intervals of length δ, and consider a large susceptible population, so that the number of people becoming infectious can be approximated by a Poisson r.v.:

$$
\begin{aligned}
A_t &\sim \text{Pois}\left(S_t \beta \frac{I_t}{N} \delta\right) \\
B_t &\sim \text{Bin}(E_t, \sigma\delta) \\
C_t &\sim \text{Bin}(I_t, \gamma\delta)
\end{aligned}
$$

and

$$
\begin{aligned}
S_{t+1} &= S_t - A_t \\
E_{t+1} &= E_t + A_t - B_t \\
I_{t+1} &= I_t + B_t - C_t \\
R_{t+1} &= R_t + C_t
\end{aligned}
$$

where, the parameters and r.v.s are represented in Figure 4.8.

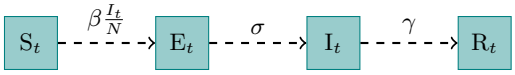


Figure 4.8: Graphical representation of the SEIR epidemic model assumed here. The boxes represent each transmission compartment: susceptible S_t, infected but not yet infectious E_t, infectious I_t and removed R_t. The arrows represent the possible transitions and are overlaid by the respective rates: the infection rate, $\beta \frac{I_t}{N}$; the rate of becoming infectious σ; and the recovery rate γ.

It is convenient to re-define the system according to another notation. Let κ index the possible compartments (e.g. $\kappa = E$, exposure, $\kappa = I$ infectious phase, etc). Denote by ${}^{\eta}X_t^{\kappa}$ the r.v. representing number of people moving from compartment η to compartment κ . The r.v.s involved in the transmission model can be then re-written during the tth interval as:

$$
\begin{aligned}
{}^{\mathrm{S}}X_t^{\mathrm{E}} &\sim \text{Pois}\left(S_t \beta \frac{I_t}{N} \delta\right) \\
{}^{\mathrm{E}}X_t^{\mathrm{I}} &\sim \text{Bin}(E_t, \sigma\delta) \\
{}^{\mathrm{I}}X_t^{\mathrm{R}} &\sim \text{Bin}(I_t, \gamma\delta)
\end{aligned}
$$

This means that the number of individuals becoming infected $\{{}^{\mathrm{S}}X^{\mathrm{E}}\}_t$ is a doubly stochastic Poisson process (Kingman, 1992).

The number of people becoming exposed might undertake any of the paths represented in the MSM process of Figure 4.9, reporting the possible severity states. The events that might happen to an infected individual are:

- he recovers/dies soon after infection;
- he goes to the General Practitioner (GP) for influenza-like illness (ILI) and then recovers or dies;

- he goes to the hospital and then recovers or dies;
- he goes to the GP for ILI, then to hospital and then recovers or dies;
- he goes to the GP for ILI, then to hospital, then to ICU and then recovers or dies;
- he goes to hospital, then to ICU and then recovers or dies.

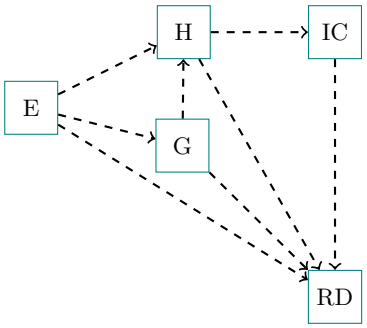


Figure 4.9: MSM for the severity states of an influenza case.

The recovery/death compartment is an absorbing compartment.

To formulate the severity process it is useful to consider the process within an immigration-death process perspective (Chiang, 1980). The individuals who become infected ${}^{\mathrm{S}}X_t^{\mathrm{E}}$ enter the compartment E, at time t, which is itself divided into three sub-compartments: those who have been infected and will visit the GP, those who will visit the hospital and those who will recover or die without being medically attended. The number of people in these conditions at time t are denoted by ${}^{\mathrm{E}}P^{\mathrm{G}}(t)$, ${}^{\mathrm{E}}P^{\mathrm{H}}(t)$ and ${}^{\mathrm{E}}P^{\mathrm{RD}}(t)$ respectively, where P stands for *prevalent*, which by definition refers to a time point and not to an interval. The number of people that join the respective prevalence compartment at time t are denoted by ${}_t^{\mathrm{E}}X^{\mathrm{G}}$, ${}_t^{\mathrm{E}}X^{\mathrm{H}}$ and ${}_t^{\mathrm{E}}X^{\mathrm{RD}}$ and are distributed according to Multinomial r.v. with size ${}^{\mathrm{S}}X^{\mathrm{E}} = {}^{\mathrm{S}}x^{\mathrm{E}}$:

$$\left({}_t^{\mathrm{E}}X^{\mathrm{G}}, {}_t^{\mathrm{E}}X^{\mathrm{H}}, {}_t^{\mathrm{E}}X^{\mathrm{RD}} \middle| {}^{\mathrm{S}}x_t^{\mathrm{E}}\right) \sim \mathrm{Multi}\left({}^{\mathrm{S}}x_t^{\mathrm{E}}; ({}^{\mathrm{E}}\theta^{\mathrm{G}}, {}^{\mathrm{E}}\theta^{\mathrm{H}}, 1 - {}^{\mathrm{E}}\theta^{\mathrm{H}} - {}^{\mathrm{E}}\theta^{\mathrm{H}})\right)$$

where ${}^{\mathrm{E}}\theta^{\mathrm{G}}$ and ${}^{\mathrm{E}}\theta^{\mathrm{H}}$ are the probability of GP consultation and hospitalization given infection, respectively.

An Exponential distribution is assumed to describe the time elapsing between events: e.g. the time from infection to GP consultation, conditionally on belonging to the group that will eventually consult the GP, is distributed according to an Exponential r.v. with rate ${}^{\mathrm{E}}\lambda^{\mathrm{G}}$. The same assumption is made for the other two waiting times with respective rates ${}^{\mathrm{E}}\lambda^{\mathrm{H}}$ and ${}^{\mathrm{E}}\lambda^{\mathrm{IC}}$. According to the Poisson process properties defined in Section 2.1.2, the exit process from the prevalence compartments are also Poisson distributed, with rate dependent on the individual rate of events ${}^{\mathrm{E}}\lambda^{\mathrm{G}}$, ${}^{\mathrm{E}}\lambda^{\mathrm{H}}$ and ${}^{\mathrm{E}}\lambda^{\mathrm{IC}}$; on the number of individuals at risk of the event, i.e. the prevalent individuals; and on the length of the intervals considered δ:

$$\begin{aligned}
{}^{\mathrm{E}}X_t^{\mathrm{G}} &\sim \mathrm{Pois}\left({}^{\mathrm{E}}\lambda^{\mathrm{G}} \cdot {}^{\mathrm{E}}P^{\mathrm{G}}(t) \cdot \delta\right) \\
{}^{\mathrm{E}}X_t^{\mathrm{H}} &\sim \mathrm{Pois}\left({}^{\mathrm{E}}\lambda^{\mathrm{H}} \cdot {}^{\mathrm{E}}P^{\mathrm{H}}(t) \cdot \delta\right) \\
{}^{\mathrm{E}}X_t^{\mathrm{RD}} &\sim \mathrm{Pois}\left({}^{\mathrm{E}}\lambda^{\mathrm{RD}} \cdot {}^{\mathrm{E}}P^{\mathrm{RD}}(t) \cdot \delta\right)
\end{aligned}$$

The prevalence at the end of the interval can then be updated as a function of the r.v.s of the individuals entering and exiting the prevalence compartment:

$$
\begin{aligned}
{}^{\mathrm{E}}P^{\mathrm{G}}(t+\delta) &= {}^{\mathrm{E}}P^{\mathrm{G}}(t) + {}^{\mathrm{E}}_{t}X^{\mathrm{G}} - {}^{\mathrm{E}}X^{\mathrm{G}}_{t} \\
{}^{\mathrm{E}}P^{\mathrm{H}}(t+\delta) &= {}^{\mathrm{E}}P^{\mathrm{H}}(t) + {}^{\mathrm{E}}_{t}X^{\mathrm{H}} - {}^{\mathrm{E}}X^{\mathrm{H}}_{t} \\
{}^{\mathrm{E}}P^{\mathrm{RD}}(t+\delta) &= {}^{\mathrm{E}}P^{\mathrm{RD}}(t) + {}^{\mathrm{E}}_{t}X^{\mathrm{RD}} - {}^{\mathrm{E}}X^{\mathrm{D}}_{t}
\end{aligned}
$$

The model for the first severity transitions is reported in Figure 4.10. The in- and out-flow for

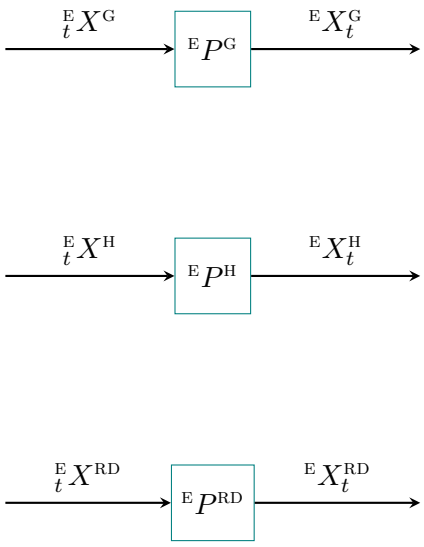


Figure 4.10: Compartmental model that describe the incidence of the events following infection as an immigration death process.

all the other severity compartments can be defined similarly. The full state process is defined in Equation 4.13 for the case described in Figure 4.9, where the first three lines represent the stochastic transmission process and the remaining lines the stochastic severity process. The distribution of the initial state of the system (the number of prevalent people in every severity state) is omitted for clarity. The model is easily extensible to as many levels of severity as there

might be.

$$
\begin{aligned}
{}^{\mathrm{S}}X_t^{\mathrm{E}} &\sim \text{Pois}\left(S_t\beta\frac{I_t}{N}\right)\\
{}^{\mathrm{E}}X_t^{\mathrm{I}} &\sim \text{Bin}(E_t,\sigma)\\
{}^{\mathrm{I}}X_t^{\mathrm{R}} &\sim \text{Bin}(I_t,\gamma)\\
\left({}_t^{\mathrm{E}}X^{\mathrm{G}},{}_t^{\mathrm{E}}X^{\mathrm{H}},{}_t^{\mathrm{E}}X^{\mathrm{RD}}\middle|{}^{\mathrm{S}}x_t^{\mathrm{E}}\right) &\sim \text{Multi}\left({}^{\mathrm{S}}x_t^{\mathrm{E}};({}^{\mathrm{E}}\theta^{\mathrm{G}},{}^{\mathrm{E}}\theta^{\mathrm{H}},1-{}^{\mathrm{E}}\theta^{\mathrm{G}}-{}^{\mathrm{E}}\theta^{\mathrm{H}})\right)\\
{}^{\mathrm{E}}X_t^{\mathrm{G}} &\sim \text{Pois}\left({}^{\mathrm{E}}\lambda^{\mathrm{G}}\cdot{}^{\mathrm{E}}P^{\mathrm{G}}(t)\cdot\delta\right)\\
{}^{\mathrm{E}}X_t^{\mathrm{H}} &\sim \text{Pois}\left({}^{\mathrm{E}}\lambda^{\mathrm{H}}\cdot{}^{\mathrm{E}}P^{\mathrm{H}}(t)\cdot\delta\right)\\
{}^{\mathrm{E}}X_t^{\mathrm{RD}} &\sim \text{Pois}\left({}^{\mathrm{E}}\lambda^{\mathrm{RD}}\cdot{}^{\mathrm{E}}P^{\mathrm{RD}}(t)\cdot\delta\right)\\
{}^{\mathrm{E}}P^{\mathrm{G}}(t+\delta) &= {}^{\mathrm{E}}P^{\mathrm{G}}(t)+{}_t^{\mathrm{E}}X^{\mathrm{G}}-{}^{\mathrm{E}}X_t^{\mathrm{G}}\\
{}^{\mathrm{E}}P^{\mathrm{H}}(t+\delta) &= {}^{\mathrm{E}}P^{\mathrm{H}}(t)+{}_t^{\mathrm{E}}X^{\mathrm{H}}-{}^{\mathrm{E}}X_t^{\mathrm{H}}\\
{}^{\mathrm{E}}P^{\mathrm{RD}}(t+\delta) &= {}^{\mathrm{E}}P^{\mathrm{RD}}(t)+{}_t^{\mathrm{E}}X^{\mathrm{RD}}-{}^{\mathrm{E}}X_t^{\mathrm{G}}\\
({}_t^{\mathrm{G}}X^{\mathrm{RD}},{}_t^{\mathrm{G}}X^{\mathrm{H}}|{}^{\mathrm{E}}x_t^{\mathrm{G}}) &\sim \text{Multi}\left({}^{\mathrm{E}}x_t^{\mathrm{G}};({}^{\mathrm{G}}\theta^{\mathrm{RS}},1-{}^{\mathrm{G}}\theta^{\mathrm{RD}})\right)\\
{}^{\mathrm{G}}X_t^{\mathrm{H}} &\sim \text{Pois}\left({}^{\mathrm{G}}\lambda^{\mathrm{H}}\cdot{}^{\mathrm{G}}P^{\mathrm{H}}(t)\cdot\delta\right)\\
{}^{\mathrm{G}}X_t^{\mathrm{RD}} &\sim \text{Pois}\left({}^{\mathrm{G}}\lambda^{\mathrm{RD}}\cdot{}^{\mathrm{G}}P^{\mathrm{RD}}(t)\cdot\delta\right)\\
{}^{\mathrm{G}}P^{\mathrm{H}}(t+\delta) &= {}^{\mathrm{G}}P^{\mathrm{H}}(t)+{}_t^{\mathrm{G}}X^{\mathrm{H}}-{}^{\mathrm{G}}X_t^{\mathrm{H}}\\
{}^{\mathrm{G}}P^{\mathrm{RD}}(t+\delta) &= {}^{\mathrm{G}}P^{\mathrm{RD}}(t)+{}_t^{\mathrm{G}}X^{\mathrm{RD}}-{}^{\mathrm{G}}X_t^{\mathrm{RD}}\\
\left({}_t^{\mathrm{H}}X^{\mathrm{IC}},{}_t^{\mathrm{H}}X^{\mathrm{RD}}\middle|{}^{\mathrm{E}}x_t^{\mathrm{H}},{}^{\mathrm{G}}x_t^{\mathrm{H}}\right) &\sim \text{Multi}\left({}^{\mathrm{E}}x_t^{\mathrm{H}}+{}^{\mathrm{G}}x_t^{\mathrm{H}};({}^{\mathrm{E}}\theta^{\mathrm{G}},{}^{\mathrm{H}}\theta^{\mathrm{RD}},1-{}^{\mathrm{H}}\theta^{\mathrm{RD}})\right)\\
{}^{\mathrm{H}}X_t^{\mathrm{IC}} &\sim \text{Pois}\left({}^{\mathrm{H}}\lambda^{\mathrm{IC}}\cdot{}^{\mathrm{H}}P^{\mathrm{IC}}(t)\cdot\delta\right)\\
{}^{\mathrm{H}}X_t^{\mathrm{RD}} &\sim \text{Pois}\left({}^{\mathrm{H}}\lambda^{\mathrm{RD}}\cdot{}^{\mathrm{H}}P^{\mathrm{RD}}(t)\cdot\delta\right)\\
{}^{\mathrm{H}}P^{\mathrm{IC}}(t+\delta) &= {}^{\mathrm{H}}P^{\mathrm{IC}}(t)+{}_t^{\mathrm{H}}X^{\mathrm{IC}}-{}^{\mathrm{H}}X_t^{\mathrm{IC}}\\
{}^{\mathrm{H}}P^{\mathrm{RD}}(t+\delta) &= {}^{\mathrm{H}}P^{\mathrm{RD}}(t)+{}_t^{\mathrm{H}}X^{\mathrm{RD}}-{}^{\mathrm{H}}X_t^{\mathrm{RD}}\\
{}^{\mathrm{IC}}X_t^{\mathrm{RD}} &\sim \text{Pois}\left({}^{\mathrm{IC}}\lambda^{\mathrm{RD}}\cdot{}^{\mathrm{IC}}P^{\mathrm{RD}}(t)\cdot\delta\right)\\
{}^{\mathrm{IC}}P^{\mathrm{RD}}(t+\delta) &= {}^{\mathrm{IC}}P^{\mathrm{RD}}(t)+{}_t^{\mathrm{IC}}X^{\mathrm{RD}}-{}^{\mathrm{IC}}X_t^{\mathrm{RD}}
\end{aligned}
\tag{4.13}
$$

This process is Markovian and, given a chosen interval-length δ, can be easily simulated using the τ-leap approximation (Gillespie, 2001) as described in Appendix D.

4.5.2 Observational process

Consider K datasets $D_1, D_2, \dots D_{\mathrm{K}}$. These datasets are usually composed of a time series at a given granularity $\Delta_{\mathrm{K}}, \forall \mathrm{K} = 1, \dots \mathrm{K}$. These time series report typically the number of incident cases moving between two severity states to the following over the interval of length Δ_{K}. The intervals at which different datasets are recorded can have different lengths; the time-index s_{K} refers to the sth interval for the dataset K according to its granularity. K here indicates both the dataset and the severity level entered, assuming that each data stream monitors only the incidence of one severe event. Again, an example of this data-collection scheme is the collection of influenza data streams. While the number of hospitalizations (D_{H}) and ICU admissions (D_{IC})

is collected on a weekly basis ($\Delta_{\text{H}} = \Delta_{\text{IC}} = 7$ days), counts of GP consultations are monitored daily (Δ_{GP}=1 day).

The average number of people observed by the dataset K during the interval s_{K} covering $[s_{\text{K}}; s_{\text{K}} + \Delta_{\text{K}})$ can be derived from the incidence of the severity state it refers to

$$\mu^{\text{K}}_{s_{\text{K}}} = \sum_{t \in s_{\text{K}}} \left(\sum_{\text{J} \neq \text{K}} {}^{\text{J}}X^{\text{K}}_t \right) d^{\text{K}}_{s_{\text{K}}}$$

where $d^{\text{K}}_{s_{\text{K}}}$ denotes the detection parameter of the dataset K, i.e. the probability of being observed given incidence. The observed cases, conditionally on the hidden states, are assumed to follow a Negative-Binomial distribution centred around the mean μ^{K}_s with over-dispersion parameter η^{K}, i.e.:

$$\left(Y^{\text{K}}_{s_{\text{K}}} \middle| {}^{\text{J}}X^{\text{K}}_t \right) \sim \text{NegBinom}(\mu^{\text{K}}_{s_{\text{K}}}, \eta^{\text{K}}). \tag{4.14}$$

4.5.3 Bootstrap Particle Filter to approximate the likelihood

Equations 4.13 and 4.14 define the states and the data distribution. The methods illustrated in Section 4.2 can be used to approximate the likelihood of the available data given a chosen parameter value.

A BPF is used to approximate the likelihood (the code for the algorithm is reported in appendix D). An innovation with respect to the majority of the applications present in the literature is that, given the different granularities of the data, this algorithm iterates over the intervals and over the data available at intervals of different granularity.

An illustration of how the model performs is reported in Figure 4.11. Here the proposed particles of the hidden state ${}^{\text{E}}X^{\text{C}}_t$ are represented in grey, while the particles re-sampled according to three different datasets are in colour. The blue dots refer to the time-steps where the particle set has been re-sampled according to the data on the flu-confirmed GP consultations during a specific day, while the green and red particles are re-sampled according to weekly hospital and ICU data respectively.

One of the most important parameters to define is δ, the length of the interval of the state process. The smaller the interval the better the τ-leap approximation will be, especially at the beginning and the end of the epidemic when numbers are small. Figure 4.12 reports the same results of Figure 4.11 adopting an interval of half a day, in the top panel, and of a quarter of day, in the bottom panel.

4.5.4 Model discussion

SMC methods marry well with epidemic models, due to the dynamical character of the phenomenon and to the timely availability of the data.

The model proposed here is very easy to simulate and, by exploiting Markovianity and conditional independence of the observations, the dependencies that usually affect epidemic data are decomposed and acknowledged correctly.

By contrast, the algorithm used for the likelihood approximation of such a model is computationally expensive, which strongly discourages its use within a Bayesian MCMC procedure such as the MCWM and the GIMH procedures mentioned above. To improve the approximation

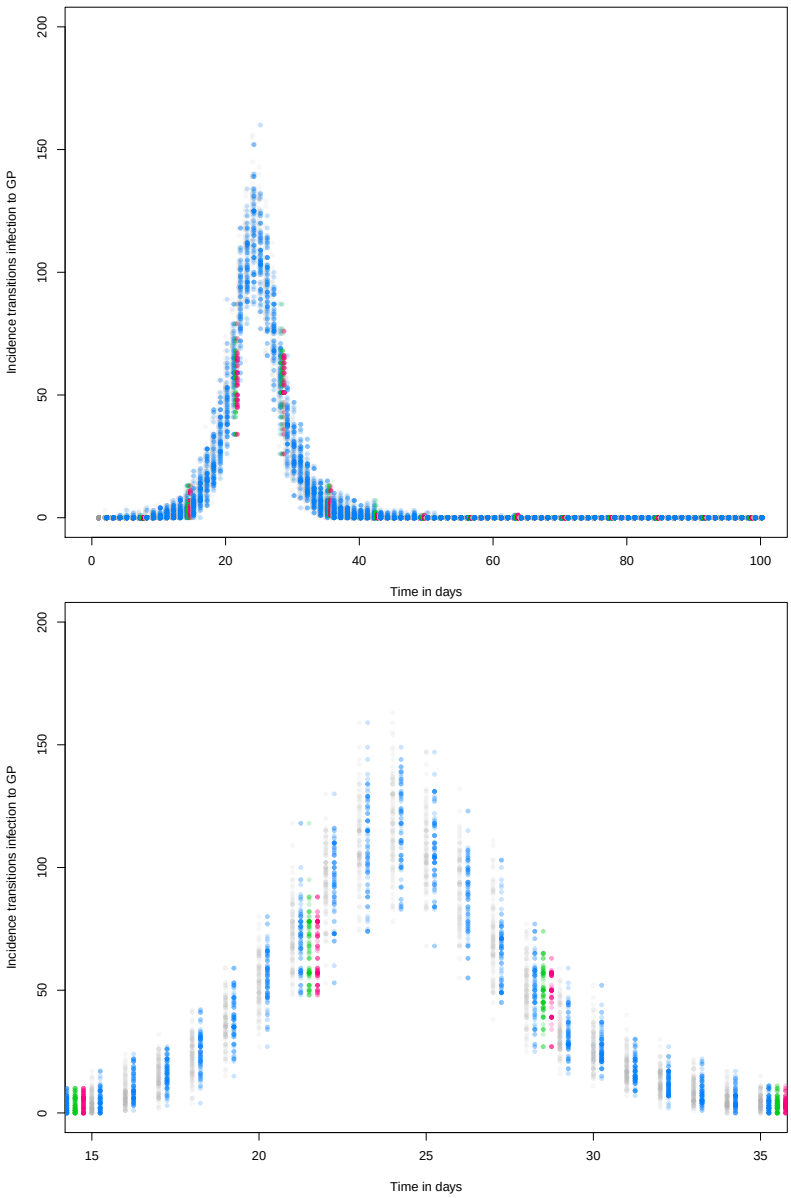


Figure 4.11: Proposed and re-sampled particles of the number of individuals visiting GPs according to the BPF (legend in text). The bottom panel is just a zoomed image of the top panel around the epidemic peak.

of the epidemic dynamics, a small time-step for the state process must be chosen. However, by doing so, the algorithm to approximate the likelihood becomes more and more time-consuming, prohibiting inference.

4.6 Conclusions

SSMs are shown to provide a valuable set up for both transmission and severity models, describing dynamic systems and the latent layers of information. Complementary SMC methods include a vast set of tools that are very flexible and that can be tailored to the assumed system and the available data.

In this chapter, an SMC algorithm is proposed to approximate the likelihood of the parameters of a chain-Multinomial severity system. This set up allows the analysis of data collected

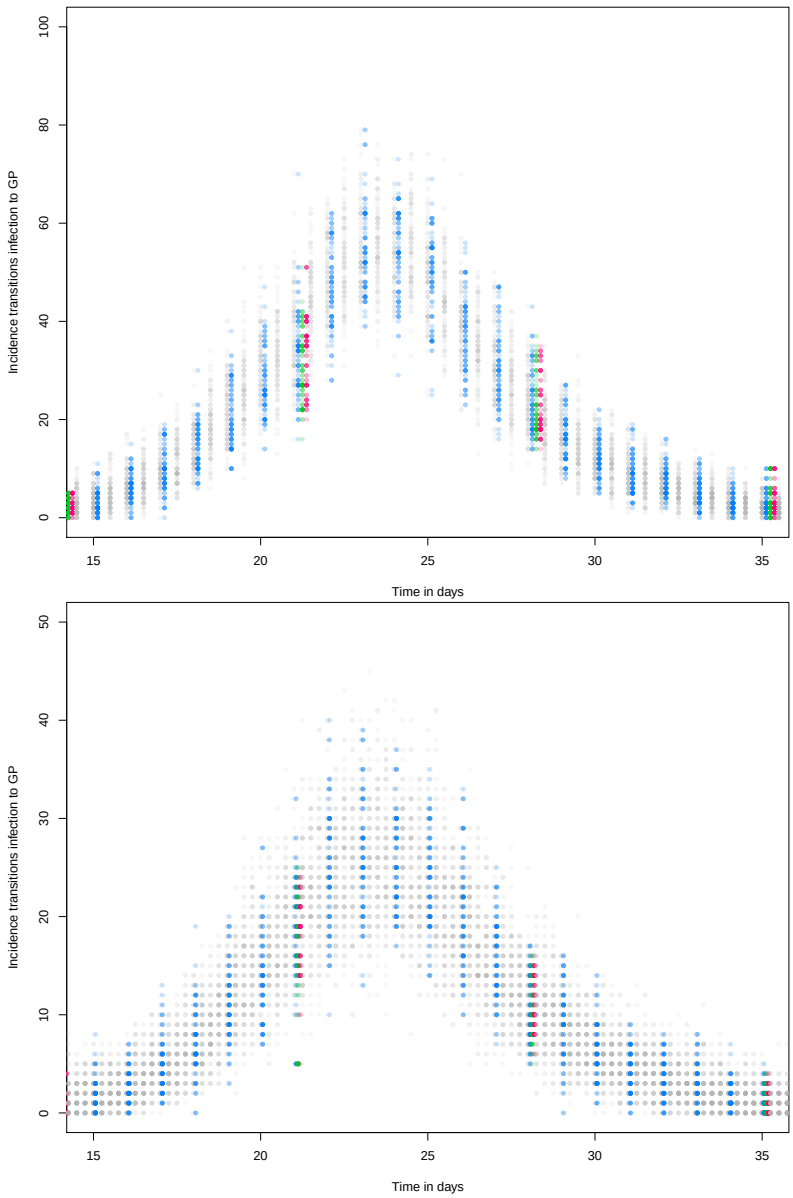


Figure 4.12: Proposed and re-sampled particles of the number of individuals visiting GPs according to the BPF (legend in text) zoomed around epidemic peak. In the top panel the interval step has length half a day and in the bottom panel a quarter of a day.

routinely in England such as the weekly count of confirmed influenza ICU admissions and deaths. According to the estimators reviewed and formulated in Chapter 2, these data were inadequate in providing information on the iCFR, since they do not include any information on the recovery process. However, here a SSM is used to account for the hidden stochastic recovery process and the SMC algorithm formulated here allows for approximation of the likelihood of a given parameter. Standard SMC has the problem of particle collapse and the traditional algorithm is amended to guard against this occurrence.

Furthermore, a SSM to model count data on cases at different levels of severity that arise from an epidemic is here formulated. SMC methods are again used to approximate the likelihood by simulating and filtering hidden states at successive time points and sequential levels of severity. Despite providing a simple framework to model and draw inference from multiple epidemic data, the proposed algorithm is highly computationally intensive. Moreover, in a context such as the

analysis of seasonal influenza, a fully stochastic transmission model is often unnecessary. A model that matches a deterministic epidemic dynamic with a stochastic severity process would be less computationally expensive and realistic enough to model available influenza data. Such a model is developed in the next chapter.

Chapter 5

Multiple data for stochastic severity

The epidemic model adopted for the case study of Section 3.3 uses a very simplistic formulation for the severity and detection model. In contrast, Chapter 2 reports several methods to model the severity dynamics and consequently to infer severity parameters.

The aim of this chapter is to explore the incorporation of more complex and realistic severity models in the study of a transmission model.

While the model becomes more complex, the inclusion of more data sources is both necessary and intricate. It is necessary because, with the proliferation of parameters and the inclusion of layers of stochasticity, a larger amount of information is needed to overcome parameter identifiability problems; it is intricate because these multiple sources might be dependent, not only within themselves (e.g. due to temporal dependence), but also between each other (e.g. due to individuals appearing in more than one dataset).

The chapter is structured as follows. Section 5.1 presents possible ways to include layers of stochasticity and the challenges of its application to multiple data. A model for the analysis of multiple dependent data is formulated in Section 5.2. Section 5.3 contains arguments on parameter inference in the context of models with stochastic severity and employs the methods introduced in Chapter 4 . Section 5.4 contains a comprehensive simulation study to verify the relevance of the method proposed. The chapter terminates with a summary of the findings.

5.1 Challenges

Corbella et al. (2018), similarly to other works (e.g. Birrell et al. (2011), Te Beest et al. (2015), Birrell et al. (2017)) combines a deterministic transmission model and a deterministic severity model with a Negative Binomial observational noise.

Precisely, the number of people being infected during the interval t, ΔI_t, is a deterministic function of the initial state of the epidemic and of the parameters:

$$\Delta I_t = S(t\delta) - S(t\delta + \delta) \qquad \forall t = 0, 1, \dots, T.$$

Likewise, the number of infected individuals who will eventually become severe and be admitted to an Intensive Care Unit (ICU) after being infected at time t is a deterministic function of ΔI_t, hence:

$$\Delta I_t^{\text{ICU}} = \Delta I_t \cdot p_{ICU} = (S(t\delta) - S(t\delta + \delta)) \cdot p_{ICU},$$

and lastly, the number of people who are admitted to ICU during the t-th interval, the ICU cases, μ_t^{ICU}, can be obtained via convolution:

$$\mu_t^{\text{ICU}} = \sum_{d=0}^{t} \Delta I_{t-d}^{\text{ICU}} f_d(\vartheta),$$

where $f_d(\vartheta)$ is the probability that the delay between infection and ICU admissions is in the dth interval of length δ, $[\delta d; \delta d + \delta)$.

Around the deterministic and unknown number of ICU admissions, the observations are assumed to be distributed according to a Negative Binomial random variable (r.v.) with over-dispersion parameter η:

$$Y_t^{\text{ICU}} \sim \text{NegBin}(\mu_t^{\text{ICU}}, \eta).$$

While the assumption of a deterministic transmission has been shown reasonable for seasonal influenza in a large population (see Section 3.1.5 of this thesis), the same cannot be said for the assumption of a deterministic severity model with observational noise. This model in fact assumes that the true unknown number of ICU admissions, μ_t^{ICU}, is just a deterministic function of the parameters of the model. This implies that, for example, if the model assumes constant transmission rates, ICU counts follow tightly the time-pattern of the outbreak, being just a scaled and delayed version of the epidemic curve of the number of new infections ΔI_t.

Nevertheless, it seems unreasonable that all the variability seen in the data $y_{1:T}^{\text{ICU}}$ would be attributable to detection/observation noise. More likely, the number of infected individuals that will eventually be admitted to ICU, ΔI_t^{ICU} is a r.v., stochastically dependent on the infected individuals ΔI_t and the severity parameters p_{ICU}. Similarly, the number of people that have been infected at interval t and have experienced d intervals of delay cannot be only described by the proportion $\Delta I_t^{\text{ICU}} f_d(\vartheta)$ but the variability of this delay process should also be considered.

Such considerations apply not only to the model of Corbella et al. (2018), but also to any transmission models for which most of the variability is caused by the severity process.

Section 5.1.1 explores the formulation of a stochastic model for disease severity, and Section 5.1.2 explores the inclusion of multiple data sources within a model with stochastic severity.

5.1.1 More complex model: layers of stochasticity

For the sake of generalizability the notation used to characterize the severity model changes from the one used at the beginning of this section and in Section 3.1.7, and the process is considered in a state space model (SSM) perspective. A more general nomenclature is adopted for the severity parameters as well. The severity SSM is defined in discrete time, through intervals of length δ so that the t-th interval is defined as $[\delta t, \delta t + \delta)$ and the intervals are indexed by $t = 0, 1, 2, \ldots, T$.

Stochastic total number of severe events

Denote by $\kappa = 0, 1, 2, \ldots, \text{K}$ the index representing the possible events of increasing severity, with $\kappa = 0$ representing the infection and all the others being some events that can be described by a pyramidal subset structure (i.e. $\text{K} \subseteq \text{K} - 1 \subseteq \cdots \subseteq 1 \subseteq 0$). Denote by X^{κ} the r.v. representing the total number of people experiencing severe event κ by the end of the epidemic.

Denote by ${}^0\theta^{\text{K}}$ the probability of experiencing event K conditionally on infection and by ${}^{\text{J}}\theta^{\text{K}}$ the probability of experiencing event K conditionally on having already experienced event J.

In a stochastic epidemic model X^0, the final size of the epidemic, is a r.v., as are the numbers of new infections in every interval X_t^0, $t = 0, 1, 2, \ldots, T$. In a deterministic epidemic model such as the one considered here, instead, these numbers are just deterministic functions of the transmission parameters, for this reason Greek letters are used, as for all the other model parameters, ξ^0 and ξ_t^0.

In the deterministic perspective, assumed for example in Corbella et al. (2018), the number of people experiencing a severe event is also a deterministic function of the parameter ξ^0 and ${}^0\theta^{\text{K}}$ or ${}^{\text{J}}\theta^{\text{K}}$. For example, in a simple and generic pyramid structure such as the one drawn in Figure 5.1 the event K affects a subset of the population that experienced J, and therefore ${}^0\theta^{\text{K}} = {}^0\theta^{\text{J}} \cdot {}^{\text{J}}\theta^{\text{K}}$. The number of people experiencing the event K can be equally derived with

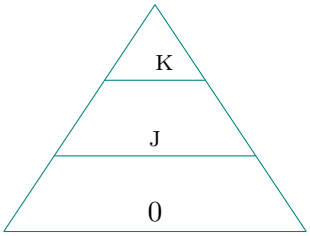


Figure 5.1: Two-levels pyramid structure.

$$\xi^{\text{K}} = \xi^0 \cdot {}^0\theta^{\text{K}}$$

or with

$$\xi^{\text{K}} = \xi^0 \cdot {}^0\theta^{\text{J}} \cdot {}^{\text{J}}\theta^{\text{K}}$$

Taking a stochastic perspective instead, the quantity of interest is the r.v. of the number people experiencing event K, X^{K}. This could be, for example, modelled as a Binomial sample from the number of infected individuals:

$$X^{\text{K}} \sim \text{Bin}(\xi^0, {}^0\theta^{\text{K}}). \tag{5.1}$$

Exploiting the pyramidal structure (Figure 5.1) the same quantity can be defined as a chain of Binomial r.v.s, one conditional on another, as follows:

$$\begin{aligned} X^{\text{J}} &\sim \text{Bin}(\xi^0, {}^0\theta^{\text{J}}) \\ (X^{\text{K}} | X^{\text{J}} = x^{\text{J}}) &\sim \text{Bin}(x^{\text{J}}, {}^{\text{J}}\theta^{\text{K}}). \end{aligned} \tag{5.2}$$

This equation introduces a dependence across different severity states since individuals at a higher severity level are a sub sample of individuals at a lower severity level (see, for example, the models in Presanis et al. (2014)).

The Binomial distribution is the most reasonable and intuitive model, yet other distributional choices might be made. If the number of infected individuals ξ^0 is large, and the probability of the first severe outcome ${}^0\theta^{\text{J}}$ is small, Equation 5.2 can be approximated by the Poisson distribution, leading to:

$$\begin{aligned} X^{\text{J}} &\sim \text{Pois}(\xi^0 \cdot {}^0\theta^{\text{J}}) \\ (X^{\text{K}} | X^{\text{J}} = x^{\text{J}}) &\sim \text{Bin}(x^{\text{J}}, {}^{\text{J}}\theta^{\text{K}}). \end{aligned}$$

The distribution of the number of events at the next level of severity X^{K}, is still a Binomial r.v. conditional on X^{J}, but, from the properties of the Poisson distribution, Equation 5.1 can be approximated by a Poisson r.v., i.e.:

$$X^{\text{K}} \sim \text{Pois}(\xi^0 \cdot {}^0\theta^{\text{K}}).$$

If the variability in the observations is high and the hypothesis of the equality of mean and variance of the Poisson r.v. is not realistic, a further parameter η^{J} can be added to account for the over-dispersion. In this case X^{J} can be approximated by a Negative Binomial r.v.:

$$X^{\text{J}} \sim \text{NegBin}(\xi^0 \cdot {}^0\theta^{\text{J}}, \eta^{\text{J}}).$$

These are some examples of stochastic processes for the total number of severe events.

Stochastic delays

When these events are analysed over time, similar processes can be considered, though the notation is to be expanded as follows. Denote by X_t^{J} the number of people experiencing severe event J during interval t; by ${}_t^{\text{J}}X^{\text{K}}$ the number of people that will eventually experience severe event K, having already experienced event J at t; and by ${}_t^{\text{J}}X_s^{\text{K}}$ the number of people that experience event K at time s, having already experienced event J at t. Thus, the subscript and superscript on the right side denote the time and type of final events and the ones on the left side denote the time and type of previous event. When only the right superscript/subscript is reported, the final events at the subscript time are counted irrespectively (i.e. summing over) the time of previous event.

A stochastic model for the time series ${}_{1:T}^{\text{J}}X^{\text{K}}$ of the number of people that have had event J at time t ($t = 1, \dots, T$) and will experience event K can be specified similarly to Equation 5.2, leading to:

$$({}_t^{\text{J}}X^{\text{K}} | x_t^{\text{J}}) \sim \text{Bin}(x_t^{\text{J}}, {}^{\text{J}}\theta^{\text{K}}) \qquad \forall t = 1, \dots, T, \tag{5.3}$$

Denote by ${}^{\text{J}}f_d^{\text{K}}({}^{\text{J}}\vartheta^{\text{K}})$ the probability that the delay experienced between event J and K is in the dth interval of length δ, $[\delta d; \delta d + \delta)$ for $d = 0, 1, \dots, D$, with D being the largest interval index for which the delay is relevant (i.e. ${}^{\text{J}}f_d^{\text{K}}({}^{\text{J}}\vartheta^{\text{K}}) \approx 0$ for $d > D$). ${}^{\text{J}}f_d^{\text{K}}({}^{\text{J}}\vartheta^{\text{K}})$ is often a parametric distribution with appropriate parameter/parameter vector ${}^{\text{J}}\vartheta^{\text{K}}$. To make the notation lighter, ${}^{\text{J}}f_d^{\text{K}}({}^{\text{J}}\vartheta^{\text{K}})$ is replaced by ${}^{\text{J}}f_d^{\text{K}}$ below, representing both the parametric choice and the parameter used to describe the delay from J to K.

Under this discrete definition of the distribution of the time to event, and conditionally on ${}_t^{\text{J}}X^{\text{K}}$, the introduction of stochastic delays consists of defining the number of people experiencing event K at time s, having already experienced event J at t, ${}_t^{\text{J}}X_s^{\text{K}}$, as a component of the Multinomial r.v.:

$$\begin{pmatrix} {}_t^{\text{J}}X_t^{\text{K}} \\ {}_t^{\text{J}}X_{t+1}^{\text{K}} \\ \vdots \\ {}_t^{\text{J}}X_{t+d}^{\text{K}} \\ \vdots \\ {}_t^{\text{J}}X_{t+D}^{\text{K}} \end{pmatrix} \Bigg| \, {}_t^{\text{J}}X^{\text{K}} = {}_t^{\text{J}}x^{\text{K}} \sim \text{Multi} \left(\text{size} = {}_t^{\text{J}}x^{\text{K}}, \text{prob} = \begin{bmatrix} {}^{\text{J}}f_0^{\text{K}} \\ {}^{\text{J}}f_1^{\text{K}} \\ \vdots \\ {}^{\text{J}}f_d^{\text{K}} \\ \vdots \\ {}^{\text{J}}f_D^{\text{K}} \end{bmatrix} \right),$$

which can be written in compact form as:

$${}_t^{\text{J}}X_{t:t+D}^{\text{K}} | {}_t^{\text{J}}x^{\text{K}} \sim \text{Multi}\left({}_t^{\text{J}}x^{\text{K}}, {}^{\text{J}}f_{0:D}^{\text{K}}\right).$$

The number of people that experience event K at each time $t = 1, \dots T$ can then be obtained by summing these stochastic terms, i.e.:

$$X_t^{\mathrm{K}} = \sum_{d=0}^{D} {}_{t-d}{}^{\mathrm{J}}X_t^{\mathrm{K}} \tag{5.4}$$

The counts of events at a higher level of severity can be modelled likewise.

This form of delay structure introduces a dependence over time besides the one already mentioned across severity states. This is evident in Equation 5.4, where the number of people experiencing K at time t depends on r.v.s defined on the previous D intervals.

5.1.2 Multiple data: the problem of dependence

This section reveals the reason why a model such as the one defined above causes problems when applied to multiple data. The severity dynamics described are again the those represented in Pyramid 5.1, i.e. two possible severity states after infection, J and K, the latter being a subset of the former. Moreover, to exploit the properties of the Poisson r.v., the case of a small ${}^{0}\theta^{\mathrm{J}}$ and a large ξ_t^0 is considered, hence:

$${}_t^0X^{\mathrm{J}} \sim \mathrm{Pois}(\xi_t^0 \cdot {}^{0}\theta^{\mathrm{J}}) \qquad \forall t = 1, \dots, T \tag{5.5}$$

Following the definition of stochastic delay and the pyramidal structure, other conditional quantities can be defined: the number of people that have event J at time t, X_t^{J}, is a convolution of the people who have been infected at $t - d$ and experience delay d between infection and event J, ${}_{t-d}{}^{0}X_t^{\mathrm{J}}$. Of these, a Binomial sample ${}_t^{\mathrm{J}}X^{\mathrm{K}}$ will experience event K. A further Multinomial r.v. ${}_t^{\mathrm{J}}X_{t+d}^{\mathrm{K}}$ denotes the people that have experienced event K after d intervals from J, the convolution of which leads to X_t^{K}, the number on individuals experiencing event K at time t. This system can be defined through the following r.v.:

$$\begin{aligned}
{}_t^0X_{t:t+D}^{\mathrm{J}} | {}_t^0x^{\mathrm{J}} &\sim \mathrm{Multi}\left({}_t^0x^{\mathrm{J}}, {}^{0}f_{0:D}^{\mathrm{J}}\right) & X_t^{\mathrm{J}} &= \sum_{d=0}^{D} {}_{t-d}{}^{0}X_t^{\mathrm{J}} \\
{}_t^{\mathrm{J}}X^{\mathrm{K}} | x_t^{\mathrm{J}} &\sim \mathrm{Bin}\left(x_t^{\mathrm{J}}, {}^{\mathrm{J}}\theta^{\mathrm{K}}\right) & \forall t &= 1, \dots, T \\
{}_t^{\mathrm{J}}X_{t:t+D}^{\mathrm{K}} | {}_t^{\mathrm{J}}x^{\mathrm{K}} &\sim \mathrm{Multi}\left({}_t^{\mathrm{J}}x^{\mathrm{K}}, {}^{\mathrm{J}}f_{0:D}^{\mathrm{K}}\right) & X_t^{\mathrm{K}} &= \sum_{d=0}^{D} {}_{t-d}{}^{\mathrm{J}}X_t^{\mathrm{K}}
\end{aligned}$$

A directed acyclic graph (DAG) representing the severity process in case of no delay and in case of stochastic delays is reported in Figure 5.2 and 5.3, respectively. In the former the model of Equation 5.2 is displayed: the node of the number of people at the most severe state, X_t^{K}, depends on the number of people at the previous severity states X_t^{J} which depends on the number of infections ${}^{0}\xi_t$. The latter figure (5.3) represents also all the temporal dependence, with the Multinomial delays and the convolutions, represented by deterministic (dashed) relations.

If multiple observations of the severity process are available, for example at different levels of severity, a detection process needs to be defined. A simple modelling assumption might be to consider Binomially-distributed observations from both severity levels with time-varying parameters as follows:

$$\begin{aligned}
Y_t^{\mathrm{J}} | X_t^{\mathrm{J}} = x_t^{\mathrm{J}} &\sim \mathrm{Bin}(x_t^{\mathrm{J}}, \zeta_t^{\mathrm{J}}) \\
Y_t^{\mathrm{K}} | X_t^{\mathrm{K}} = x_t^{\mathrm{K}} &\sim \mathrm{Bin}(x_t^{\mathrm{K}}, \zeta_t^{\mathrm{K}})
\end{aligned}$$

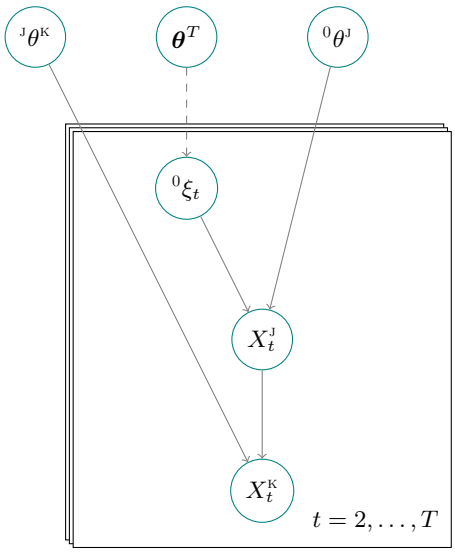


Figure 5.2: DAG of a stochastic severity model with two levels of severity and no delay.

for $t = 1, \ldots, T$, with ζ_t^{J} and ζ_t^{K} denoting the detection probability of the individuals that have had event J and K, respectively.

Using the properties of the Poisson r.v. it is easy to derive the marginal distributions of the observations, found to be Poisson (see Poisson properties reported in Section 2.1.2)

$$\begin{aligned} Y_t^{\text{J}} &\sim \text{Pois}\left(\zeta^{\text{J}} \cdot {}^0\theta^{\text{J}} \cdot \sum_{d=0}^{D} \xi_{t-d}^0 \cdot {}^0 f_d^{\text{J}}\right) \\ Y_t^{\text{K}} &\sim \text{Pois}\left(\zeta^{\text{K}} \cdot {}^{\text{J}}\theta^{\text{K}} \cdot {}^0\theta^{\text{J}} \cdot \sum_{d=0}^{D} \sum_{g=0}^{d} \xi_{t-d-g}^0 \cdot {}^0 f_d^{\text{J}} \cdot {}^{\text{J}} f_g^{\text{K}}\right) \end{aligned} \tag{5.6}$$

for $t = 1, \ldots, T$. In case each data point of the two severity time series was to be used alone for inference, the likelihood would be easily specified by the respective line of Equation 5.6.

If instead the data at the two severity levels are used jointly, the model defined above and represented in Figure 5.3 shows that each severity level shares information within itself over time for a lag D and between levels, since event K is a sub-sample of event J.

In SSM terms, the observations depend on the same hidden states and therefore, rather than each informing some hidden states independently, they share information on these hidden states.

This can be understood by thinking about the fact that some individuals will be recorded both in one dataset and in the other (these would be the people that experience both J and K and in both cases are observed). Likewise, within one time series of severe cases each observation will contain information on the infections in the current interval and in the previous ones, being correlated with the neighbouring observations.

In simpler cases (e.g. with no delays and when only two data points $(y^{\text{J}}, y^{\text{K}})$ are considered) the joint likelihood accounting for the shared hidden state can be shown to be different from the product of two independent likelihoods (see Appendix E). In Section 5.4, for more complex but realistic cases with delays and time series of observations, a simulation study is carried out to show if (and in which scenarios) accounting for this dependence would lead to different results from adopting an incorrect formulation that assumes independent observations.

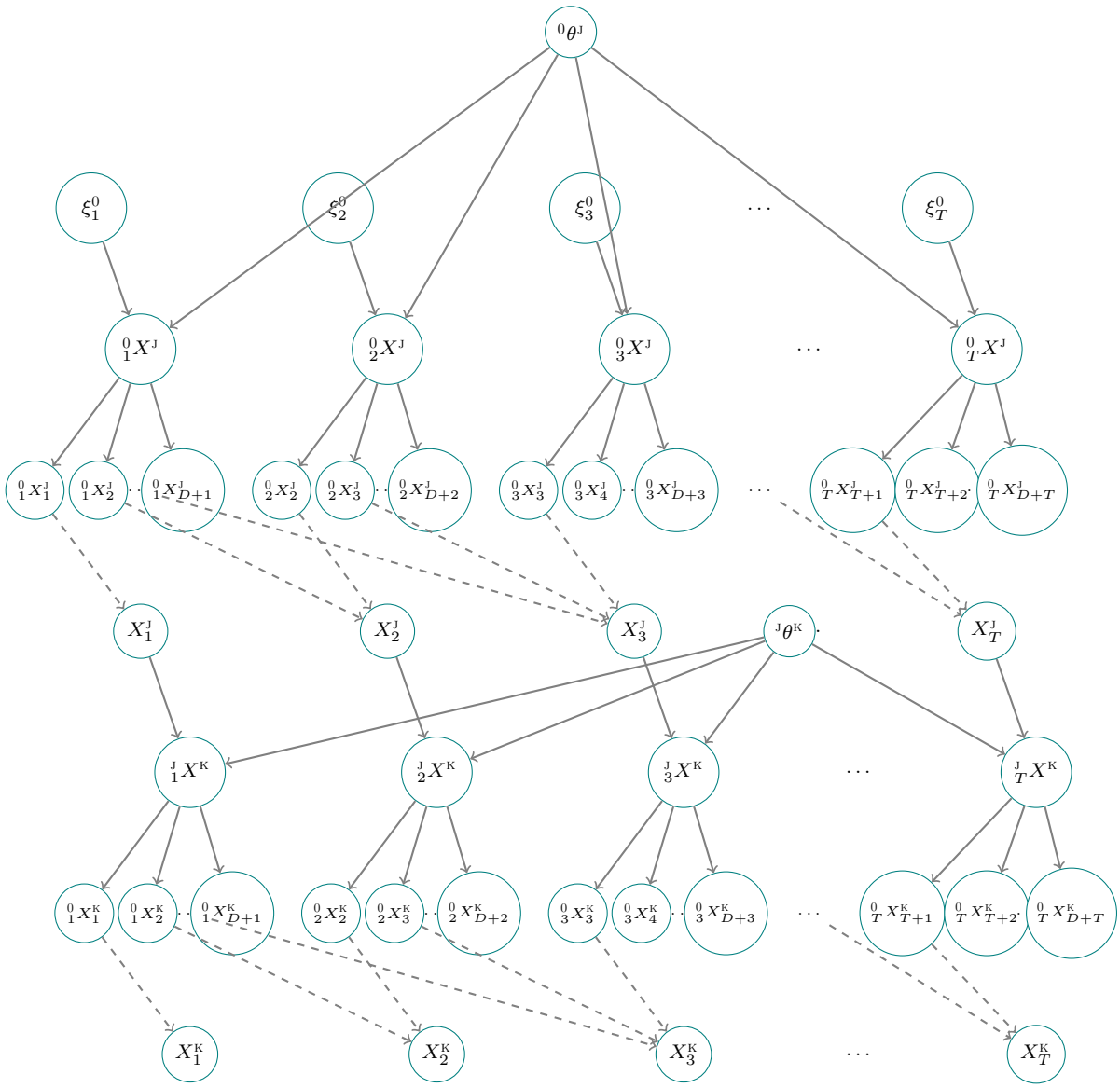


Figure 5.3: DAG of a stochastic severity model with two levels of severity and delays.

5.2 Model construction

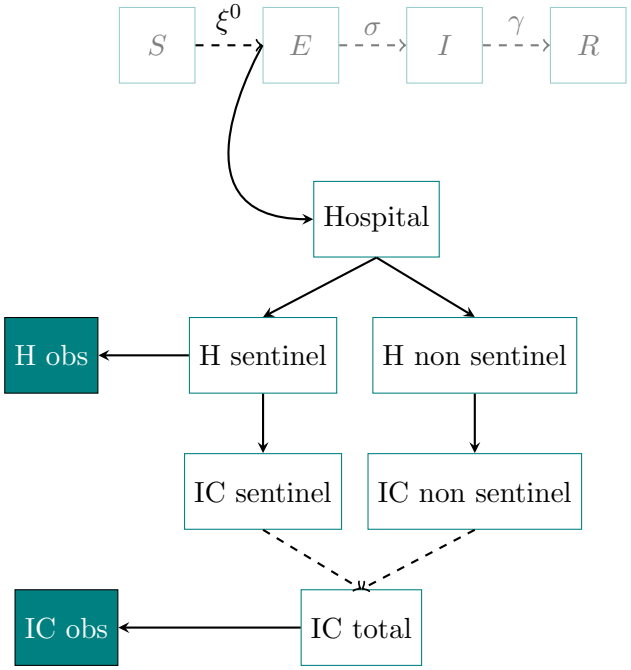
In this paragraph a motivating dataset is presented and a model of the kind described in Section 5.1 is formulated.

5.2.1 Motivation: the USISS collection scheme

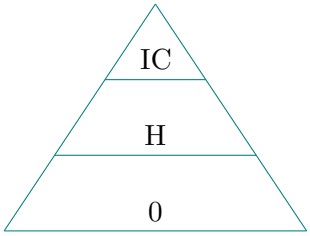
The UK Severe Influenza Surveillance System (USISS) collects information on severe cases of influenza in England. Part of its content has been discussed and used in another part of this thesis (Section 3.3), however here it is briefly recalled in the light of the considerations on dependent data expressed above.

According to the USISS protocol (Health Protection Agency, 2011a), all National Health Service (NHS) trusts in England report the weekly number of laboratory-confirmed influenza cases admitted to ICU and the number of confirmed influenza deaths in ICU. In addition to this mandatory scheme, a sentinel subgroup of NHS trusts in England, is recruited every year to participate in the USISS sentinel scheme (Boddington et al., 2017; Health Protection Agency,

2011b), which reports weekly numbers of laboratory-confirmed influenza cases hospitalised at all levels of care.



(a) Flowchart of the possible severe events recoded in USISS.



(b) Infection (0), hospitalization (H) and ICU admission (IC) pyramid structure.

Figure 5.4: Charts for USISS severity scheme: individuals might be hospitalised and then they might be admitted to ICU (b). Hospitalizations in sentinel hospitals and ICU admissions in all hospitals are recorded (a).

Figure 5.4 illustrates with a flow chart and with a pyramid the data collected through the USISS system. Some individuals might be detected in both datasets, leading to a dependence.

5.2.2 Model formulation

Denote by $\boldsymbol{\theta}$ the set of parameters, composed of $\boldsymbol{\theta}^T$, the parameters of the transmission model and $\boldsymbol{\theta}^S$, the parameters of the severity and detection model.

$\boldsymbol{\theta}^T = \{\pi, \iota, \sigma, \gamma, \beta\}$ consists of the transmission rate β; the exit rates from compartments E and I, σ and γ respectively; and the initial proportions immune, π, and of infected/infectious, ι. π and ι, together with σ, γ and the known constant N, the total size of the population, contribute to the formulation of the initial state of the epidemic.

Since a deterministic model is assumed, the information contained in $\boldsymbol{\theta}^T$, together with the known constants, provides the full time series of the number of new infections.

Consider intervals of length δ so that the t-th interval covers the time $[t\delta, t+\delta)$ and the intervals are now indexed by $t = 0, 1, 2, \ldots, T$, where $t = 0$ coincides with the beginning of the data collection period and $t = T$ is the end of the data collection period. Denote the number of susceptible individuals at the beginning of interval t by S_t and likewise for the other compartments E, I, R. Denote by $\xi^0_{1:T}$ the vector of the number of new infections in interval $t = 0, 1, 2, \ldots T$.

$\boldsymbol{\theta}^S = \left\{{}^0\theta^{\mathrm{H}}, {}^{\mathrm{H}}\theta^{\mathrm{IC}}, {}^0f^{\mathrm{H}}, {}^{\mathrm{H}}f^{\mathrm{IC}}, \zeta_t^{\mathrm{H}}, \zeta_t^{\mathrm{IC}}\right\}$ includes probabilities of severe events conditional on in-

fection at a previous severity level: ${}^{0}\theta^{\mathrm{H}}$, the probability of hospitalization given infection; and ${}^{\mathrm{H}}\theta^{\mathrm{IC}}$, the probability of ICU admission given hospitalization. ${}^{0}f^{\mathrm{H}}$ and ${}^{\mathrm{H}}f^{\mathrm{IC}}$ denotes discretised Exponentially-distributed waiting times for the time from infection to hospitalisation and from hospital admission to ICU admission, respectively. Lastly, the detection parameters are denoted by ζ_t^{H} and ζ_t^{IC}, and are the probability of detecting an hospitalised and ICU case, respectively.

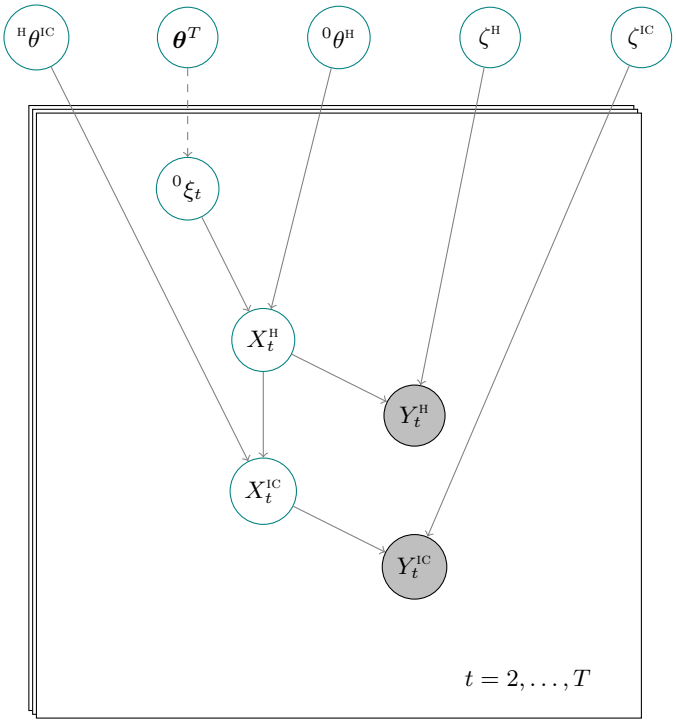


Figure 5.5: DAG for two correlated data with no delay. Each r.v. is represented as a node, where grey nodes correspond to variables that are observed and white nodes are latent variables. The arrows express the dependence among the variables: dashed arrows represent deterministic functions and solid arrows stochastic functions.

The model can be described using SSM notation, defining a state process and an observational process. The former is composed of the distributions of: ${}^{0}_{t}X^{\mathrm{H}}$, the number of hospitalizations that were infected at each interval t; X_t^{H}, the number of hospitalizations at t, obtained via a convolution of ${}^{0}_{t}X^{\mathrm{H}}$; the number of hospitalizations that eventually will be admitted to IC and have been hospitalised at t, ${}^{\mathrm{H}}_{t}X^{\mathrm{IC}}$; the number of IC admissions at t, X_t^{IC} obtained by the convolution of ${}^{\mathrm{H}}_{t}X^{\mathrm{IC}}$, for $t = 0, 1, 2, \ldots, T$. The state process is reported in Equation 5.7.

$$
\begin{aligned}
\left({}^{0}_{t}X^{\mathrm{H}}\right) &\sim \mathrm{Pois}({}^{0}\xi \cdot {}^{0}\theta^{\mathrm{H}}) \\
\left({}^{0}_{t}X^{\mathrm{H}}_{t+1:t+D} | {}^{0}_{t}X^{\mathrm{H}} = {}^{0}_{t}x^{\mathrm{H}}\right) &\sim \mathrm{Multi}({}^{0}_{t}x^{\mathrm{H}}, {}^{0}f^{\mathrm{H}}_{1:D}), \qquad X_t^{\mathrm{H}} = \sum_{s=1}^{S} {}_{t-s}{}^{0}X_s^{\mathrm{H}} \\
({}^{\mathrm{H}}_{t}X^{\mathrm{IC}} | X_t^{\mathrm{H}} = x_t^{\mathrm{H}}) &\sim \mathrm{Bin}(x_t^{\mathrm{H}}, {}^{\mathrm{H}}\theta^{\mathrm{IC}}) \\
\left({}^{\mathrm{H}}_{t}X^{\mathrm{IC}}_{t+1:t+D} | {}^{\mathrm{H}}_{t}X^{\mathrm{IC}} = {}^{\mathrm{H}}_{t}x^{\mathrm{IC}}\right) &\sim \mathrm{Multi}({}^{\mathrm{H}}_{t}x^{\mathrm{IC}}, {}^{\mathrm{H}}f^{\mathrm{IC}}_{1:D}), \qquad X_t^{\mathrm{IC}} = \sum_{s=1}^{S} {}_{t-s}{}^{\mathrm{H}}X_s^{\mathrm{IC}}
\end{aligned}
\tag{5.7}
$$

for $t = 0, 1, \ldots, T$.

Regarding the observational process, the distributions of two datasets, $y_{1:T}^{\mathrm{H}}$, the count of hospitalizations, and $y_{1:T}^{\mathrm{IC}}$, the count of IC admissions, conditional on the hidden states $X_{1:T}^{\mathrm{H}}$

and $X^{\text{IC}}_{1:T}$, are assumed to be Binomial with probabilities set to the detection probabilities ζ^{H}_t and ζ^{IC}_t respectively, as reported in Equation 5.8

$$\begin{aligned}(Y^{\text{H}}_t | X^{\text{H}}_t = x^{\text{H}}_t) &\sim \text{Bin}(x^{\text{H}}_t, \zeta^{\text{H}}_t) \\ (Y^{\text{IC}}_t | X^{\text{IC}}_t = x^{\text{IC}}_t) &\sim \text{Bin}(x^{\text{IC}}_t, \zeta^{\text{IC}}_t)\end{aligned} \tag{5.8}$$

with $t = 0, 1, 2, \ldots, T$.

Thanks to the Poisson properties, several hidden states and the data distribute marginally according to a Poisson distribution, as reported in Equation 5.9

$$\begin{aligned}X^{\text{H}}_t &\sim \text{Pois}\left({}^0\theta^{\text{H}} \cdot \sum_{d=0}^{D} \xi^0_{t-d} \cdot {}^0 f^{\text{H}}_d\right) \\ X^{\text{IC}}_t &\sim \text{Pois}\left({}^{\text{H}}\theta^{\text{IC}} \cdot {}^0\theta^{\text{H}} \cdot \sum_{d=0}^{D}\sum_{g=0}^{d} \xi^0_{t-d} \cdot {}^0 f^{\text{H}}_d \cdot {}^{\text{H}} f^{\text{IC}}_g\right) \\ Y^{\text{H}}_t &\sim \text{Pois}\left(\zeta^{\text{H}}_t \cdot {}^0\theta^{\text{H}} \cdot \sum_{d=0}^{D} \xi^0_{t-d} \cdot {}^0 f^{\text{H}}_d\right) \\ Y^{\text{IC}}_t &\sim \text{Pois}\left(\zeta^{\text{IC}}_t \cdot {}^{\text{H}}\theta^{\text{IC}} \cdot {}^0\theta^{\text{H}} \cdot \sum_{d=0}^{D}\sum_{g=0}^{d} \xi^0_{t-d-g} \cdot {}^0 f^{\text{H}}_d \cdot {}^{\text{H}} f^{\text{IC}}_g\right)\end{aligned} \tag{5.9}$$

for $t = 0, 1, 2, \ldots T$.

Despite the two data having simple marginal distributions, they are not independent since they share information on some individuals (the ones that are detected both in hospital and IC) as can be seen from the DAG in Figure 5.5, which represents the same data in the specific case of no delay between consecutive events.

For this reason the joint distribution of $(Y^{\text{H}}_{1:T}, Y^{\text{IC}}_{1:T})$ is non-trivial. The next section describes two approximations of the joint likelihood.

5.3 Model inference

A simulation algorithm is proposed to approximate the joint likelihood of hospitalization and IC data. The joint probability distribution can be decomposed in two ways:

$$\begin{aligned}p(y^{\text{H}}_{1:T}, y^{\text{IC}}_{1:T} | \boldsymbol{\theta}) &= p(y^{\text{H}}_{1:T} | y^{\text{IC}}_{1:T}, \boldsymbol{\theta}) p(y^{\text{IC}}_{1:T} | \boldsymbol{\theta}) \\ &= p(y^{\text{IC}}_{1:T} | y^{\text{H}}_{1:T}, \boldsymbol{\theta}) p(y^{\text{H}}_{1:T} | \boldsymbol{\theta})\end{aligned}$$

where, in both cases, one of the two factors is available in closed form.

Algorithm 7 exploits the first decomposition, where $p(y^{\text{IC}}_{1:T} | \boldsymbol{\theta})$ is available in closed form and a solution is needed for $p(y^{\text{H}}_{1:T} | y^{\text{IC}}_{1:T}, \boldsymbol{\theta})$ which is obtained by approximating the T-dimensional integral:

$$\begin{aligned}p(y^{\text{H}}_{1:T} | y^{\text{IC}}_{1:T}, \boldsymbol{\theta}) &= \int_{X^{\text{H}}_1} \cdots \int_{X^{\text{H}}_T} p(y^{\text{H}}_{1:T}, X^{\text{H}}_{1:T} | y^{\text{IC}}_{1:T}, \boldsymbol{\theta}) \text{d}X^{\text{H}}_1 \ldots \text{d}X^{\text{H}}_T \\ &= \int_{X^{\text{H}}_1} \cdots \int_{X^{\text{H}}_T} p(y^{\text{H}}_{1:T} | X^{\text{H}}_{1:T}, y^{\text{IC}}_{1:T}, \boldsymbol{\theta}) p(X^{\text{H}}_{1:T} | y^{\text{IC}}_{1:T}, \boldsymbol{\theta}) \text{d}X^{\text{H}}_1 \ldots \text{d}X^{\text{H}}_T \\ &= \int_{X^{\text{H}}_1} \cdots \int_{X^{\text{H}}_T} p(y^{\text{H}}_{1:T} | X^{\text{H}}_{1:T}, \boldsymbol{\theta}) p(X^{\text{H}}_{1:T} | y^{\text{IC}}_{1:T}, \boldsymbol{\theta}) \text{d}X^{\text{H}}_1 \ldots \text{d}X^{\text{H}}_T\end{aligned}$$

Result: $\widehat{p}(y^{\mathrm{H}}_{1:T}, y^{\mathrm{IC}}_{1:T}|\boldsymbol{\theta})$
Input: fixed parameter $\boldsymbol{\theta}$, number of particels N, data $y^{\mathrm{H}}_{1:T}$, $y^{\mathrm{IC}}_{1:T}$
compute
$p(y^{\mathrm{IC}}_{1:T}|\boldsymbol{\theta}) = f(y^{\mathrm{IC}}_{1:T}|\zeta^{\mathrm{IC}}_t \cdot {}^{\mathrm{H}}\theta^{\mathrm{IC}} \cdot {}^{0}\theta^{\mathrm{H}} \cdot \sum_{d=0}^{D}\sum_{g=0}^{d} \xi^{0}_{t-d-g} \cdot {}^{0}f^{\mathrm{H}}_{d} \cdot {}^{\mathrm{H}}f^{\mathrm{IC}}_{g})$
with $f(\cdot)$ being a Poisson density
for $n = 1, \ldots, N$ **do**
  for $t = 0, 1, \ldots T$ **do**
    sample
    $x^{\mathrm{IC}(n)}_t \sim \mathrm{Pois}\left((1-\zeta^{\mathrm{IC}}_t)\left[{}^{\mathrm{H}}\theta^{\mathrm{IC}} \cdot {}^{0}\theta^{\mathrm{H}} \sum_{d=0}^{D}\sum_{g=0}^{d} \xi^{0}_{t-d-g} \cdot {}^{0}f^{\mathrm{H}}_{d} \cdot {}^{\mathrm{H}}f^{\mathrm{IC}}_{g}\right]\right) + y^{\mathrm{IC}}_T$
    sample
    ${}^{\mathrm{H}}_{t-1}x^{\mathrm{IC}(n)}_1, \ldots, {}^{\mathrm{H}}_{t-S}x^{\mathrm{IC}(n)}_S \sim \mathrm{Multi}(x^{\mathrm{IC}(n)}_t, {}^{\mathrm{H}}f^{\mathrm{IC}}_{1:S})$
    compute
    ${}^{\mathrm{H}}_{t}x^{\mathrm{IC}(n)} = \sum_{s=1}^{S} {}^{\mathrm{H}}_{t}x^{\mathrm{IC}(n)}_s$
    sample
    $x^{\mathrm{H}(n)}_t|{}^{\mathrm{H}}_{t}x^{\mathrm{IC}(n)}, \boldsymbol{\theta} \sim \mathrm{Pois}((1-{}^{\mathrm{H}}\theta^{\mathrm{IC}}) \left[{}^{0}\theta^{\mathrm{H}} \sum_{d=0}^{D} \xi^{0}_{t-d} \cdot {}^{0}f^{\mathrm{H}}_{d}\right]) + {}^{\mathrm{H}}_{t}x^{\mathrm{IC}(n)}$
  end
  compute
  $p(y^{\mathrm{H}}_{1:T}|{x^{\mathrm{H}}_{1:T}}^{(n)}, \boldsymbol{\theta}) = g(y^{\mathrm{H}}_{1:T}|{x^{\mathrm{H}}_{1:T}}^{(n)}, \zeta^{\mathrm{H}}_t)$
  with $g(\cdot)$ being a Binomial density
end
$\widehat{p}(y^{\mathrm{H}}_{1:T}, y^{\mathrm{IC}}_{1:T}|\boldsymbol{\theta}) = p(y^{\mathrm{IC}}_{1:T}|\boldsymbol{\theta}) \cdot \frac{1}{N}\sum_{n=1}^{N} p(y^{\mathrm{H}}_{1:T}|{x^{\mathrm{H}}_{1:T}}^{(n)}, \boldsymbol{\theta})$

Algorithm 7: First approximation of the likelihood

by conditional independence of the state space model.

Algorithm 8 approximates the other factor. $p(y^{\mathrm{H}}_{1:T}|\boldsymbol{\theta})$ is available in closed from and a solution is needed for $p(y^{\mathrm{IC}}_{1:T}|y^{\mathrm{H}}_{1:T}, \boldsymbol{\theta})$.

$$\begin{aligned} p(y^{\mathrm{IC}}_{1:T}|y^{\mathrm{H}}_{1:T}, \boldsymbol{\theta}) &= \int_{X^{\mathrm{IC}}_1} \cdots \int_{X^{\mathrm{IC}}_T} p(y^{\mathrm{IC}}_{1:T}, X^{\mathrm{IC}}_{1:T}|y^{\mathrm{H}}_{1:T}, \boldsymbol{\theta}) \mathrm{d}X^{\mathrm{IC}}_1 \ldots \mathrm{d}X^{\mathrm{IC}}_T \\ &= \int_{X^{\mathrm{IC}}_1} \cdots \int_{X^{\mathrm{IC}}_T} p(y^{\mathrm{IC}}_{1:T}|X^{\mathrm{IC}}_{1:T}, y^{\mathrm{H}}_{1:T}, \boldsymbol{\theta}) p(X^{\mathrm{IC}}_{1:T}|y^{\mathrm{H}}_{1:T}, \boldsymbol{\theta}) \mathrm{d}X^{\mathrm{IC}}_1 \ldots \mathrm{d}X^{\mathrm{IC}}_T \\ &= \int_{X^{\mathrm{IC}}_1} \cdots \int_{X^{\mathrm{IC}}_T} p(y^{\mathrm{IC}}_{1:T}|X^{\mathrm{IC}}_{1:T}, \boldsymbol{\theta}) p(X^{\mathrm{IC}}_{1:T}|y^{\mathrm{H}}_{1:T}, \boldsymbol{\theta}) \mathrm{d}X^{\mathrm{IC}}_1 \ldots \mathrm{d}X^{\mathrm{IC}}_T \end{aligned}$$

It is straightforward to simulate from the hidden states, thanks to the structure assumed: the delays follow a Multinomial structure while the remaining process is a chain of Binomial and Poisson r.v.s. Specifically, let:

$$\begin{aligned} (X|\lambda) &\sim \mathrm{Pois}(\lambda) \\ (Y|\theta, x) &\sim \mathrm{Bin}(x, \theta) \\ (Y|\lambda, \theta) &\sim \mathrm{Pois}(\lambda\theta), \qquad \text{by property of the Poisson} \end{aligned}$$

The distribution of $X|y$ can be derived as follows:

$$\begin{aligned} p(x|y) &= \frac{p(y|x)p(x)}{p(y)} && \text{by Bayes' theorem,} \\ &= \frac{\frac{x!}{y!(x-y)!}\theta^y(1-\theta)^{x-y}\frac{\lambda^x}{x!}e^{-\lambda}}{\frac{(\lambda\theta)^y}{y!}e^{-\lambda\theta}} \\ &= \frac{\lambda^{x-y}(1-\theta)^{x-y}e^{-\lambda(1-\theta)}}{\frac{1}{(x-y)!}} \end{aligned}$$

which is the density function of a Poisson with rate $\lambda(1-\theta)$ plus y.

Result: $\widehat{p}(y^{\mathrm{H}}_{1:T}, y^{\mathrm{IC}}_{1:T}|\boldsymbol{\theta})$
Input: $\boldsymbol{\theta}$, N, $y^{\mathrm{H}}_{1:T}$, $y^{\mathrm{IC}}_{1:T}$
compute
$p(y^{\mathrm{H}}_{1:T}|\boldsymbol{\theta}) = f\left(y^{\mathrm{H}}_{1:T}|\zeta^{\mathrm{H}}_t \cdot {}^0\theta^{\mathrm{H}} \cdot \sum_{d=0}^{D} \xi^0_{t-d} \cdot {}^0f^{\mathrm{H}}_d\right)$
with $f(\cdot)$ being a Poisson density
for $n = 1, \ldots, N$ **do**
  for $t = 0, 1, \ldots T$ **do**
    sample
    $x^{\mathrm{H}(n)}_t \sim \text{Pois}\left((1-\zeta^{\mathrm{H}}_t)^0\theta^{\mathrm{H}} \sum_{d=0}^{D} \xi^0_{t-d} \cdot {}^0f^{\mathrm{H}}_d\right) + y^{\mathrm{H}}_t$
    sample
    ${}^{\mathrm{H}}_t x^{\mathrm{IC}(n)}|x^{\mathrm{H}(n)}_t \sim \text{Bin}(x^{\mathrm{H}(n)}_t, {}^{\mathrm{H}}\theta^{\mathrm{IC}})$
    sample
    ${}^{\mathrm{H}}_t x^{\mathrm{IC}}_{1:S}{}^{(n)}|{}^{\mathrm{H}}_t x^{\mathrm{IC}(n)}, \boldsymbol{\theta} \sim \text{Multi}({}^{\mathrm{H}}_t x^{\mathrm{IC}(n)}, {}^{\mathrm{H}}f^{\mathrm{IC}}_{1:S})$
    compute
    $x^{\mathrm{IC}(n)}_t = \sum_{s=1}^{S} {}^{\mathrm{H}}_{t-s} x^{\mathrm{IC}(n)}_s$
  end
  $p(y^{\mathrm{IC}}_{1:T}|x^{\mathrm{IC}}_{1:T}{}^{(n)}, \boldsymbol{\theta}) = g(y^{\mathrm{IC}}_{1:T}|x^{\mathrm{IC}}_{1:T}{}^{(n)}, \zeta^{\mathrm{IC}}_t)$
  with $g(\cdot)$ being a Binomial density
end
$\widehat{p}(y^{\mathrm{H}}_{1:T}, y^{\mathrm{IC}}_{1:T}|\boldsymbol{\theta}) = p(y^{\mathrm{H}}_{1:T}|\boldsymbol{\theta}) \cdot \frac{1}{N}\sum_{n=1}^{N} p(y^{\mathrm{IC}}_{1:T}|x^{\mathrm{IC}}_{1:T}{}^{(n)}, \boldsymbol{\theta})$

Algorithm 8: Second approximation of the likelihood

Both the algorithms are very attractive since they use a simple vanilla Monte Carlo (MC) approximation. However, a better approximation can be obtained when the distribution from which the samples are drawn (in this case the distribution of the hidden states conditional on the first data included) matches well with the target distribution (in this case the distribution of the hidden states conditional on both) (Brooks et al., 2011). This matching improves substantially when the target distribution is more variable; on opposite, when the mass of the distribution is highly concentrated on a point, more simulations would result in low weights, hence they will be wasted. This principle motivates the choice between algorithm 7 and 8 as illustrated below.

For the case of USISS, as well as for many other observational collection schemes, more

severe cases are monitored more carefully. While only approximately 20% of the hospitalised cases are recorded in the dataset, almost all ICU cases are reported. The Binomial observational likelihood of $y^{\mathrm{H}}_{1:T}$ (Algorithm 7) is much more variable than the Binomial observational likelihood of $y^{\mathrm{IC}}_{1:T}$ (Algorithm 8). Moreover, the distribution from which Algorithm 7 samples, $x^{\mathrm{H}}_{1:T}|y^{\mathrm{IC}}_{1:T}$ results in 0 for all the sampled values $x^{\mathrm{H}(n)}_t < y^{\mathrm{H}}_t$ for $t = 1, \dots, T$. Similarly the distribution from which Algorithm 8 samples, $x^{\mathrm{IC}}_{1:T}|y^{\mathrm{H}}_{1:T}$ is equal to 0 for all the sampled values $x^{\mathrm{IC}(n)}_t < y^{\mathrm{IC}}_t$ for $t = 1, \dots, T$. The latter case takes place much more often, due to the high detection of ICU admissions: for this reason, obtaining a good sample for the MC approximation is much harder in this case. Hence Algorithm 7 is adopted throughout the chapter.

Both a Monte Carlo within Metropolis (MCWM) and a grouped independence Metropolis Hastings (GIMH) algorithm have been coded in order to perform parameter inference. While the former is affected by bias (which disappears for N large enough), the latter presents the inconvenient feature of bad mixing of the chains, especially when N is small (Andrieu and Roberts, 2009).

5.4 Relevance of the dependence

A full simulation study is set up to assess whether (and in which situations) accounting for the dependence makes any difference compared to assuming the two datasets to be independent. The simulation set-up and the results are showed below.

5.4.1 Simulation study set-up

The simulated data are formulated to reflect a situation similar to the motivating USISS data on a very small population, chosen in order to reduce the computation time. The datasets are generated with some common parameters and some scenario-specific parameters.

The common parameters are:

$$\left\{N = 10000, \beta = 0.63, \pi = 0.3, \iota = 0.0001, \sigma = \frac{1}{4}, \gamma = \frac{1}{3.5}, {}^{0}f^{\mathrm{H}} \sim \mathrm{Exp}(0.3), {}^{\mathrm{H}}f^{\mathrm{IC}} \sim \mathrm{Exp}(0.4)\right\}$$

while the scenario specific parameters are reported in Table 5.1. Each of the severity and detection parameters can take two values, a small and a large value. The smaller leads to situations where the probability of being observed in both datasets is low, therefore data are less dependent; while when parameters take bigger values, there is more overlap between datasets and therefore more dependence. The probability of hospitalization and the detection of hospitalization in the

	small dependence	big dependence
${}^{0}\theta^{\mathrm{H}}$	0.1	0.5
${}^{\mathrm{H}}\theta^{\mathrm{IC}}$	0.1	0.9
$\zeta^{\mathrm{H}}_t = \zeta^{\mathrm{H}}$	0.1	0.3
$\zeta^{\mathrm{IC}}_t = \zeta^{\mathrm{IC}}$	0.1	0.9

Table 5.1: Parameters used to generate the datasets.

big dependence case are smaller than the respective quantities for ICU admission. This choice is made (i) because usually more severe cases are better monitored, and hence have a higher

detection and (ii) because the algorithm chosen to approximate the joint likelihood performs better with smaller values of the hidden state.

The aim of the comparison is to assess whether a misspecified independent likelihood would affect the inference of the parameters showing different posterior distributions from the ones obtained with the MC approximation of the joint likelihood assuming dependent data. For this reason the results presented in the following pages are to be compared *within* each scenario: between the ones obtained with the independent misspecified model (often abbreviated with MISS IND) and the ones obtained using the dependent joint model (often abbreviated with JOINT DEP).

5.4.2 Comparison for transmission parameters

For the parameter inference with the approximated joint likelihood, a MCWM algorithm with $N = 2000$ was chosen. In fact, preliminary simulation experiments showed a much better mixing when the MCWM scheme was used in place of the GIMH and $N = 2000$ was observed to be large enough to make the simulation study feasible and the bias of the MCWM algorithm negligible (i.e. to provide posterior distributions that were indistinguishable from the GIMH ones).

Hence, a MCWM algorithm including approximation of the joint likelihood via Algorithm 7 and a blocked Metropolis Hastings (MH) algorithm using the misspecified independent Poisson likelihood are used on the same 1000 datasets and the results are compared. 500 datasets are simulated using the smallest values of the parameters (left column of Table 5.1) and 500 datasets using the largest values (right column of Table 5.1). The only parameters inferred are the transmission parameters β, ι and π with the severity parameters being fixed at their true, scenario-specific, value.

In the case of **small dependence**, the results show that the posterior distributions obtained with the misspecified independent likelihood are very similar to the ones obtained with the approximated joint likelihood. Figure 5.6 reports the posterior distributions of the 3 parameters estimated in 5 datasets randomly sampled from the 500 generated datasets. Moreover, Figure 5.7 displays the distribution of pairwise difference in the variance of the posterior sample:

$$\text{PWD}(\text{Var}(\alpha))^d = \text{Var}(\widehat{\alpha}^{\text{JOINT DEP}}|\boldsymbol{y}^d) - \text{Var}(\widehat{\alpha}^{\text{MISS IND}}|\boldsymbol{y}^d) \qquad \alpha = \beta, \pi, \ldots; d = 1, 2, \ldots, 500$$

and the distribution of pairwise difference in the length of the 95% credible interval (CrI) of the posterior sample:

$$\text{PWD}(\text{R}_{95}(\alpha))^d = \text{R}_{95}(\widehat{\alpha}^{\text{JOINT DEP}}|\boldsymbol{y}^d) - \text{R}_{95}(\widehat{\alpha}^{\text{MISS IND}}|\boldsymbol{y}^d) \qquad \alpha = \beta, \pi, \ldots; d = 1, 2, \ldots, 500$$

The figures show an imperceptible difference in the posterior distributions and their precision-summaries between the two models. This is confirmed by quantities such as the proportion of datasets in which the pairwise difference in variance is less than or equal to 0, for each of the parameters reported below. When this quantity is close to 0.5, the variances of the estimates obtained with the two methods are similar within datasets; when this quantity is close to 1 it suggests that the variance of the estimates obtained using the misspecified independent likelihood is systematically larger than the variance of the estimates obtained with the joint likelihood; and when this quantity is close to 0 it highlights that the former variance is systematically smaller

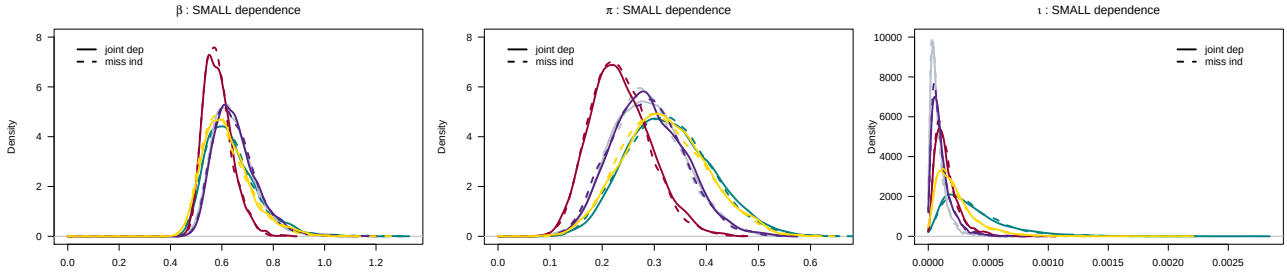


Figure 5.6: Posterior distribution of the transmission parameters β (left panel), ι (centre) and π (right panel) from 5 datasets. The colour of the posterior density identifies the dataset analysed while dashed lines refer to results from the misspecified independent model and filled lines to results from the model approximating the joint dependent likelihood.

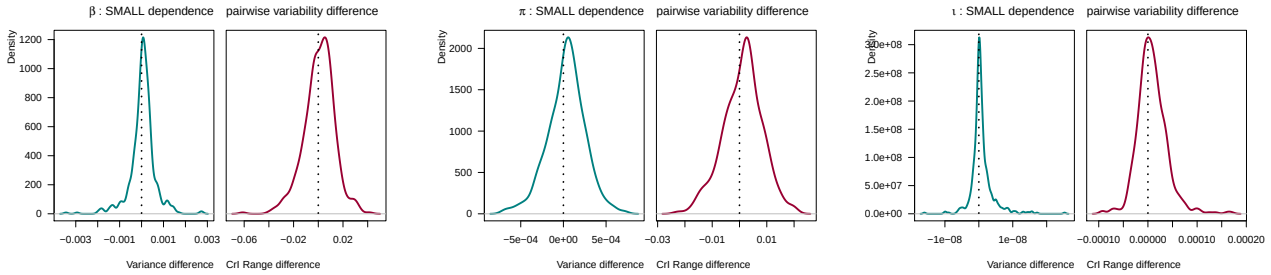


Figure 5.7: Distribution of the pairwise difference in variance (green plots) and in 95% CrI length (red plots) of the posterior distribution of the transmission parameters β (left panel), ι (centre) and π (right panel).

then the latter. This results is expected, above all in the case of high dependence scenario, signifying over-precision of the independent-likelihood-driven estimator.

Table 5.2 reports this quantity for each parameter estimated: here there is no evident signal of systematic difference between methods.

Parameter	Proportion of PWD(Var) ≤ 0
β	0.392
π	0.390
ι	0.378

Table 5.2: Proportion of datasets in which the pairwise difference of variance is smaller or equal to 0 for the three transmission parameters in the scenario with small dependence.

The same analysis is run on the 500 datasets with a **big dependence** with results reported in Figures 5.8 and 5.9.

Here there is a notable difference between the results from the two models: the posterior distributions from the misspecified model that assumes independent data are less variable than the ones derived using the MC approximation of the joint dependent likelihood.

This result was expected since the misspecified model, by assuming independent data, accounts for more information than is inherent in the data. This leads to an overconfidence that can be detected in the underestimation of the posterior variance. Results are confirmed by the proportion of differences less than or equal to 0 for all the parameters (Table 5.3), strongly suggesting a systematic difference in variability between the two methods. In all the simulated

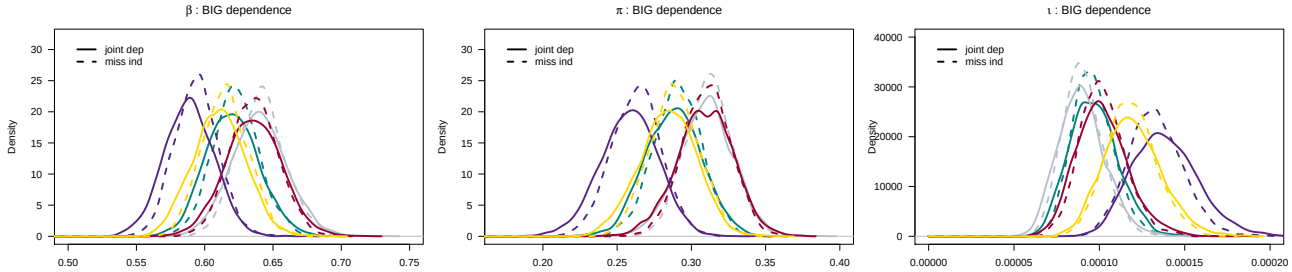


Figure 5.8: Posterior distribution of the transmission parameters β (left panel), ι (centre) and π (right panel) from 5 datasets. The colour of the posterior density identifies the dataset analysed while dashed lines refer to results from the misspecified independent model and filled lines to results from the model approximating the joint dependent likelihood.

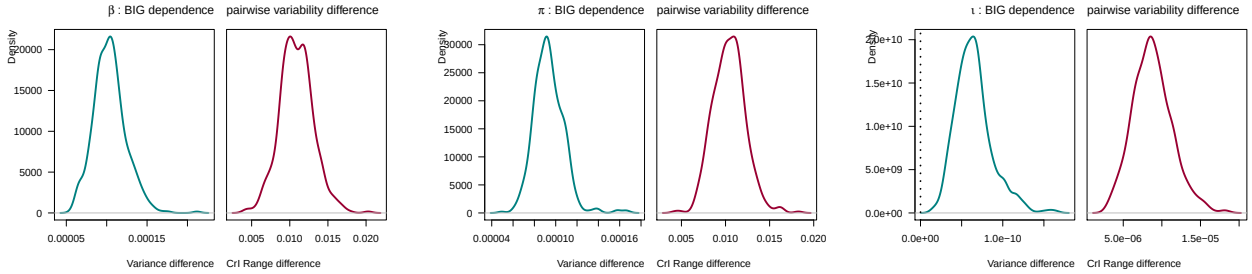


Figure 5.9: Distribution of the pairwise difference in variance (green plots) and in 95% CrI length (red plots) of the posterior distribution of the transmission parameters β (left panel), ι (centre) and π (right panel).

datasets the variance of the posterior distributions of the parameters are smaller in the analysis using the misspecified independent model than in the approximate joint model (Figure 5.9).

Parameter	Proportion of PWD(Var) ≤ 0
β	0
π	0
ι	0

Table 5.3: Proportion of datasets in which the pairwise difference of variance is smaller or equal to 0 for the three transmission parameters in the scenario with big dependence.

5.4.3 Results for transmission and severity parameters

The same kind of comparison is carried out in a context where inference is drawn both for the transmission and the severity parameters. Here, since more quantities are estimated and due to the high correlation of the parameters of epidemic models, a difference between the results from the two models may be more difficult to spot. Moreover, in this multi-parameter context, convergence is sometimes compromised, particularly in the big-dependence scenario.

The results within a scenario with **small dependence** are reported in Figure 5.10 and in Figure 5.11. Neither in the transmission parameters nor in the newly estimated severity parameters, can a large difference be seen.

The proportions of pairwise differences less than or equal to 0 confirm the non-difference in the variance of the posterior distributions (Table 5.4).

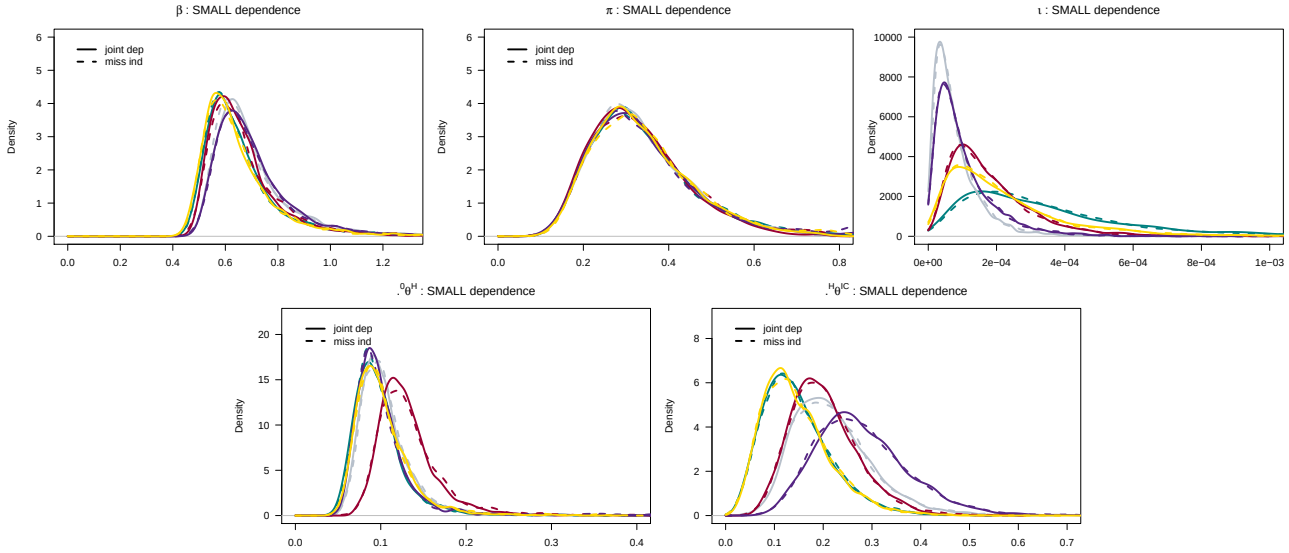


Figure 5.10: Posterior distribution of the transmission and severity parameters from 5 datasets. The colour of the posterior density identifies the dataset analysed while dashed lines refer to results from the misspecified independent model and filled lines to results from the model approximating the joint dependent likelihood.

Parameter	Proportion of PWD(Var) ≤ 0
β	0.498
π	0.464
ι	0.348
${}^{0}\theta^{\mathrm{H}}$	0.466
${}^{\mathrm{H}}\theta^{\mathrm{IC}}$	0.778

Table 5.4: Proportion of datasets in which the pairwise difference of variance is smaller or equal to 0 for the transmission and severity parameters in the scenario with small dependence.

The same results within a scenario with **big dependence** are plotted in Figures 5.12 and 5.13. The only notable difference can be seen in the distribution of ${}^{\mathrm{H}}\theta^{\mathrm{IC}}$: this parameter is what links the two datasets, since it defines the probability of ICU admission conditional on hospitalization. When the two datasets are jointly analysed, they both contribute to the estimation of ${}^{\mathrm{H}}\theta^{\mathrm{IC}}$, with hospital data informing the Binomial size in Equation 5.8 and ICU data informing the proportion of people in the more-severe state. When the two datasets are considered independently, the hospital data do not play any role in the inference of ${}^{\mathrm{H}}\theta^{\mathrm{IC}}$.

The proportions of pairwise differences less than or equal to 0 confirm the observations above: the variance of the posterior sample of the parameter ${}^{\mathrm{H}}\theta^{\mathrm{IC}}$ is always lower when inference is drawn with the approximation to the joint dependent likelihood compared to when the misspecified independent model is adopted (Table 5.5).

The remaining parameters, do not show any significant difference in the variance of their posterior distributions.

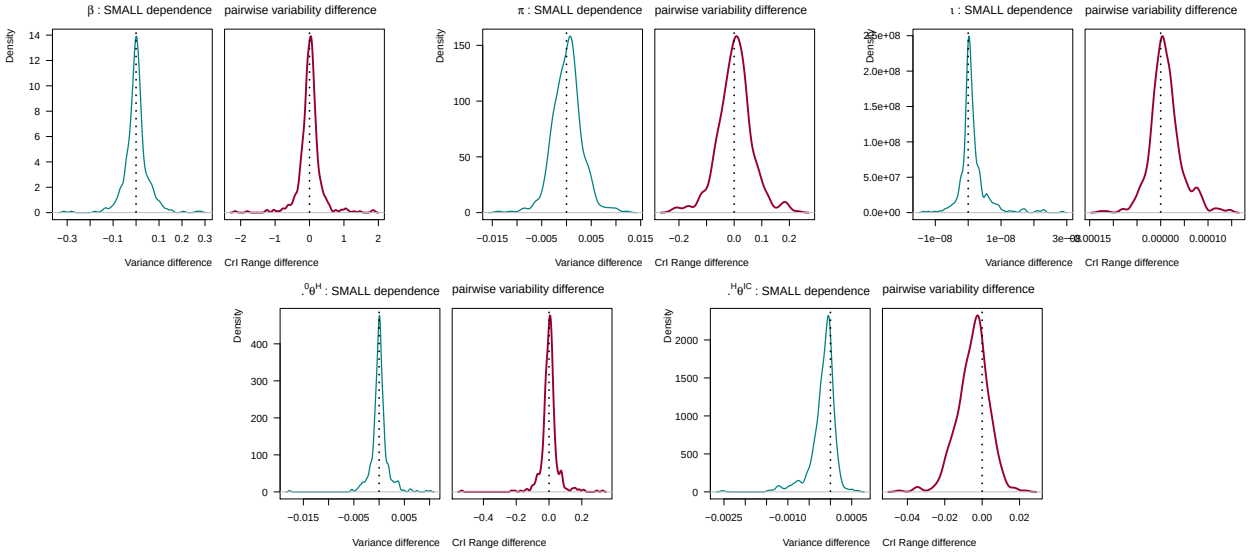


Figure 5.11: Distribution of the pairwise difference in variance (green plots) and in 95% CrI length (red plots) of the posterior distribution of the transmission and severity parameters.

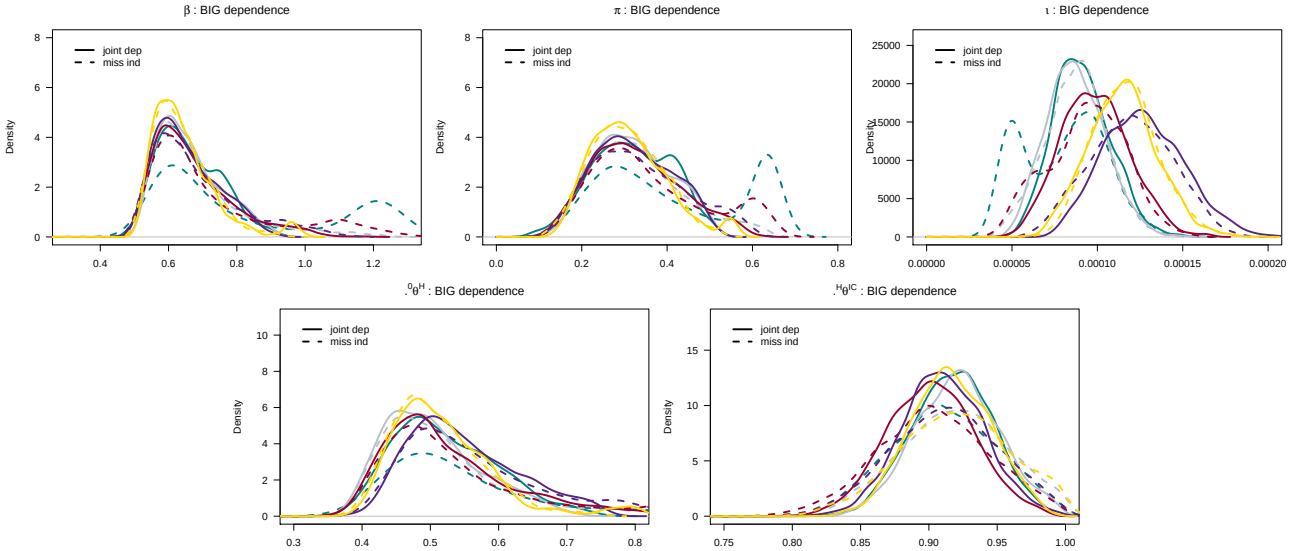


Figure 5.12: Posterior distribution of the transmission and severity parameters from 5 datasets. The colour of the posterior density identifies the dataset analysed while dashed lines refers to results from the misspecified independent model and filled lines to results from the model approximating the joint dependent likelihood.

Parameter	Proportion of PWD(Var) ≤ 0
β	0.554
π	0.546
ι	0.202
${}^{0}\theta^{\mathrm{H}}$	0.566
${}^{\mathrm{H}}\theta^{\mathrm{IC}}$	1

Table 5.5: Proportion of datasets in which the pairwise difference of variance is smaller or equal to 0 for the transmission and severity parameters in the scenario with big dependence.

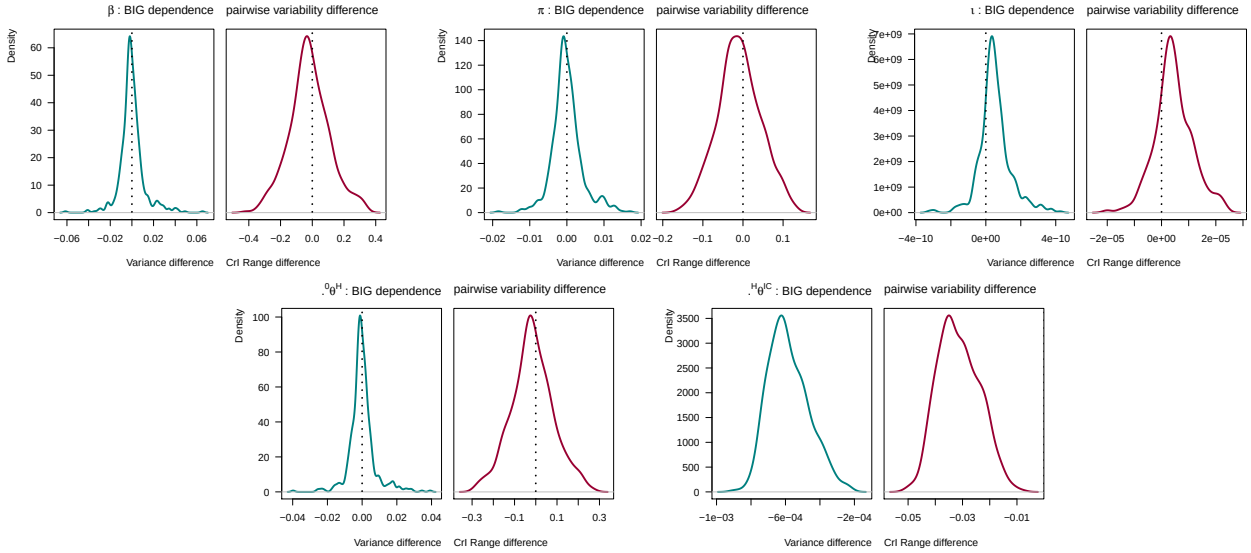


Figure 5.13: Distribution of the pairwise difference in variance (green plots) and in 95% CrI length (red plots) of the posterior distribution of the transmission and severity parameters from all the datasets.

5.4.4 Influential parameters

As a final comparison a further investigation into the main cause of the difference is undertaken. Starting from the small-dependence scenario, one at a time, each parameter of Table 5.1 is allowed to take the larger value in simulating the 500 datasets.

Estimates of the five parameters are then obtained according to the misspecified independent and the joint dependent model. The posterior distributions and the plots of the precision statistics are reported in Appendix E. While a detectable difference in the results is observed when all the parameters affecting the level of dependence vary, the same cannot be said when each parameter increases alone. Differences are less evident, with the probability of detection in ICU being the most influential parameter, as shown in Table 5.6, where each column corresponds to a scenario where all the parameters but the header of the column are assumed small.

Increased Parameter	${}^{0}\theta^{\mathrm{H}}$	${}^{\mathrm{H}}\theta^{\mathrm{IC}}$	ζ^{H}	ζ^{IC}
Parameter	Proportion of PWD(Var) ≤ 0			
β	0.468	0.454	0.296	0.490
π	0.450	0.454	0.214	0.496
ι	0.342	0.082	0.052	0.032
${}^{0}\theta^{\mathrm{H}}$	0.476	0.458	0.290	0.464
${}^{\mathrm{H}}\theta^{\mathrm{IC}}$	0.682	0.940	0.072	0.994

Table 5.6: Proportion of datasets in which the pairwise difference of variance is smaller or equal to 0 for the transmission and severity parameters in the scenario with small dependence except for the respective column-name parameter.

5.5 Conclusions

This chapter formulates a model for multiple correlated epidemic data and proposes an algorithm for its inference. The model proposed is innovative in that it merges a deterministic model for transmission, which is often a good approximation when seasonal epidemics are analysed, and a stochastic model for severity and detection.

The assumption of a deterministic transmission model breaks the dependence over time, with the exception of the delay between consequent severe events. This allows the formulation of a simple MC algorithm to approximate the likelihood. This method, is not only more intuitive, but also faster than a sequential importance procedure, for which weighting and sampling are to be computed at each time-step. Moreover, since it is possible to sample directly from the hidden distribution of interest, the problem of choosing an importance distribution is overcome.

A misspecified model that assumes independent data could be used in place of the proposed algorithm leading to a faster inference. However, a simulation study shows that the misspecified model leads to overly precise results in the estimation of the transmission parameters when the dependence is large. Furthermore, the parameter connecting the two severity states to which the data refer, is estimated with less precision when the misspecified model is assumed. The dependence between data instead is relevant in many situations, requiring the use of the algorithm proposed. These situations include the data on severe influenza cases collected in England, for which some individuals are likely to be detected twice, above all due to the high detection rate of the cases admitted to ICU.

While the model proposed is innovative and comprehensive, the estimation procedure might be extended in a number of ways, principally to enhance its efficiency since, as with many pseudo-marginal algorithms, the analysis takes a lot of time. The algorithm used to approximate the likelihood at each time step of the Markov chain Monte Carlo (MCMC) can be parallelised, potentially exploiting big computational resources such as high performance computers. The choice of using the MCWM algorithm is also dangerous: while preliminary simulation experiments have shown that the bias related to this methods should not affect the reported results, these are not highly generalizable. When applying this method to other scenarios, more particles might be needed in order to make the MCWM bias negligible; alternatively, a GIMH algorithm might be chosen in order to ensure unbiasdness of the likelihood approximation. Moreover, MCMC methods might not be the most appropriate match for MC-based approximations which require a big computational effort at every iteration. Alternative exploration of the parameter space (e.g. by sequential Monte Carlo (SMC) samplers (Del Moral, Doucet, and Jasra, 2006)) might be a more suitable match.

Chapter 6

Application to influenza in England

This chapter provides a comprehensive application of the methods illustrated in the previous chapters to multiple data on seasonal influenza in England.

The last season of influenza (during winter 2017/18) resulted in a moderate to high burden on health facilities (Public Health England, 2018). Many researchers at Public Health England (PHE) engaged in an estimation and prediction exercise exploiting a multiplicity of data to characterise the seasonal epidemic while it was ongoing. In light of the methods proposed in Chapter 5 for the joint analysis of multiple, possibly dependent, data sources, a model is proposed here, to synthesize information provided by several data streams.

The goal of the analysis is the joint estimation of transmission and severity parameters for a broad description of the epidemic. In what follows the available data are illustrated (Section 6.1); a model for transmission, severity and detection is proposed (Section 6.1) and results from real data are derived (Section 6.3). The chapter concludes with a discussion commenting both on the novelty of the model proposed and on the public-health value of the results provided.

6.1 Data

The annual monitoring of influenza activity in England is based on a number of data streams: consultations for influenza-like illness (ILI) at General Practitioners (GPs); virological testing of swabs from samples from GP consultations; serological data; and admissions to hospitals and Intensive Care Units (ICUs) from the UK Severe Influenza Surveillance System (USISS).

6.1.1 Hospital and ICU admissions

The USISS scheme, as described elsewhere in this thesis (Section 3.3.1), collects the weekly count of ICU admissions with confirmed influenza, in principle from all trusts in England, and the weekly count of hospitalizations with confirmed influenza in a stratified sentinel sample of the trusts.

Data on weekly ICU admissions are plotted in Figure 6.1, together with the proportion of the catchment population of the participating trusts over the population of England. The x axis refers to time, measured from week 40 of 2017, i.e. the beginning of the data-collection scheme. The observed number of ICU admissions peaks at week 15 (which corresponds to mid-January 2018), however the profile of the observations is peculiar: the number of admissions does not

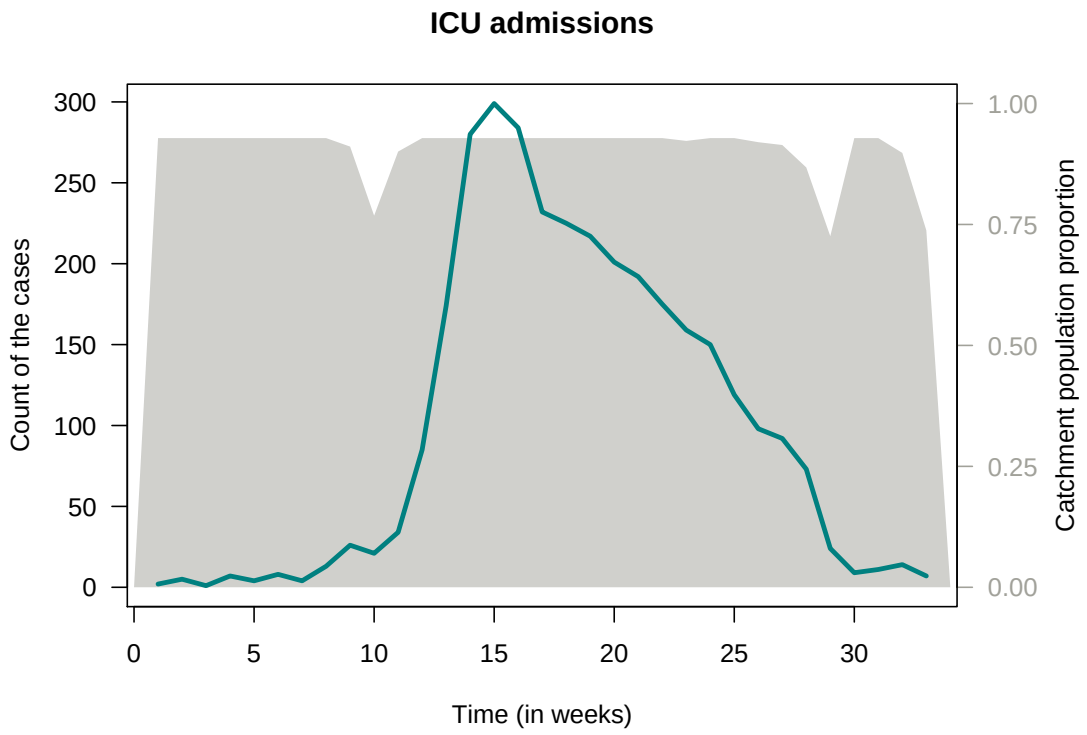


Figure 6.1: ICU data- Weekly number of ICU admissions (green) and proportion of the population that was monitored each week (shaded grey).

decrease steadily after the peak, but instead shows a plateau followed by a moderate decline in the cases.

The sentinel trusts participating to USISS cover only ≈20% of the population in England. The weekly number of reported hospitalizations at all level of care is plotted in Figure 6.2. The number of observations peaks again at week 15; with a second peak circa one month and a half later.

6.1.2 GP consultations for ILI

The daily GP ILI data come from the PHE's national influenza surveillance system which collects daily ILI consultations stratified by, amongst other things, age group and National Health Service (NHS) region. The PHE surveillance data result from the combination of databases owned by EMIS (Harcourt et al., 2012) and The Phoenix Partnership (The Phoenix Partnership, 2013) and cover on average 35% of the population of England.

The GP data are reported in Figure 6.3: the x axis refers to the time in days starting from the Monday of week 40 of 2017. Data are characterised by a weekly pattern with higher counts on Mondays and much lower counts during weekends. Nevertheless, the data still show an epidemic pattern, peaking during mid-January, similarly to the USISS data.

6.1.3 Virological positivity and serology

To identify the proportion of the reported ILI counts that are genuine cases of influenza infection, additional information collected by the Royal College of General Practitioners (RCGP) is used. RCGP primary-care surveillance is augmented with virological monitoring (Health Protection

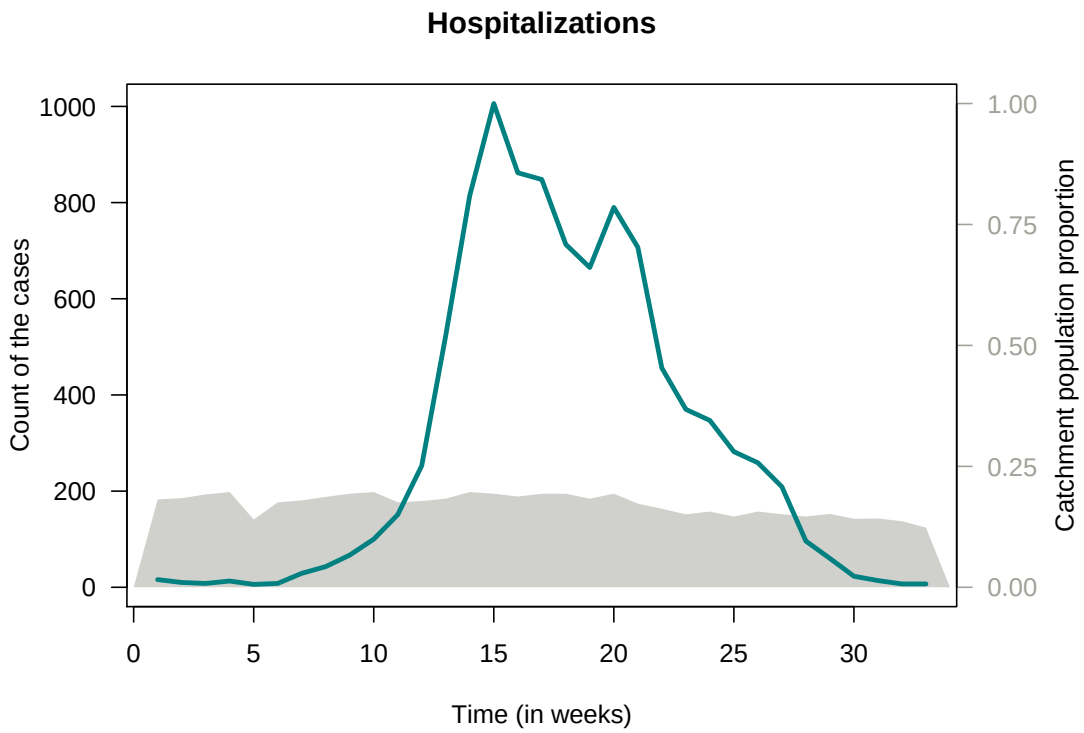


Figure 6.2: Hospital data- Weekly number of hospital admissions (green) and proportion of the population that was monitored each week (shaded grey).

Agency, 2014), which involves taking respiratory swabs from a randomly selected subset of the patients consulting for ILI. Polymerase chain reaction (PCR) testing identifies swabs positive for the presence of influenza strains as well as for other respiratory virus infections. These data, plotted in Figure 6.4, consist of the counts of the tests taken and the number of positive tests obtained.

An additional source of information is provided by a cross-sectional serological survey data on the presence of immunity-conferring antibodies in the general population. If available at the beginning of the season, this could inform the susceptibility of the population before seasonal virus circulation. These data have not been analysed directly in this thesis but estimates of the average level of immunity in the population was provided to all the people involved in the influenza-monitoring exercise (Charlett, 2018).

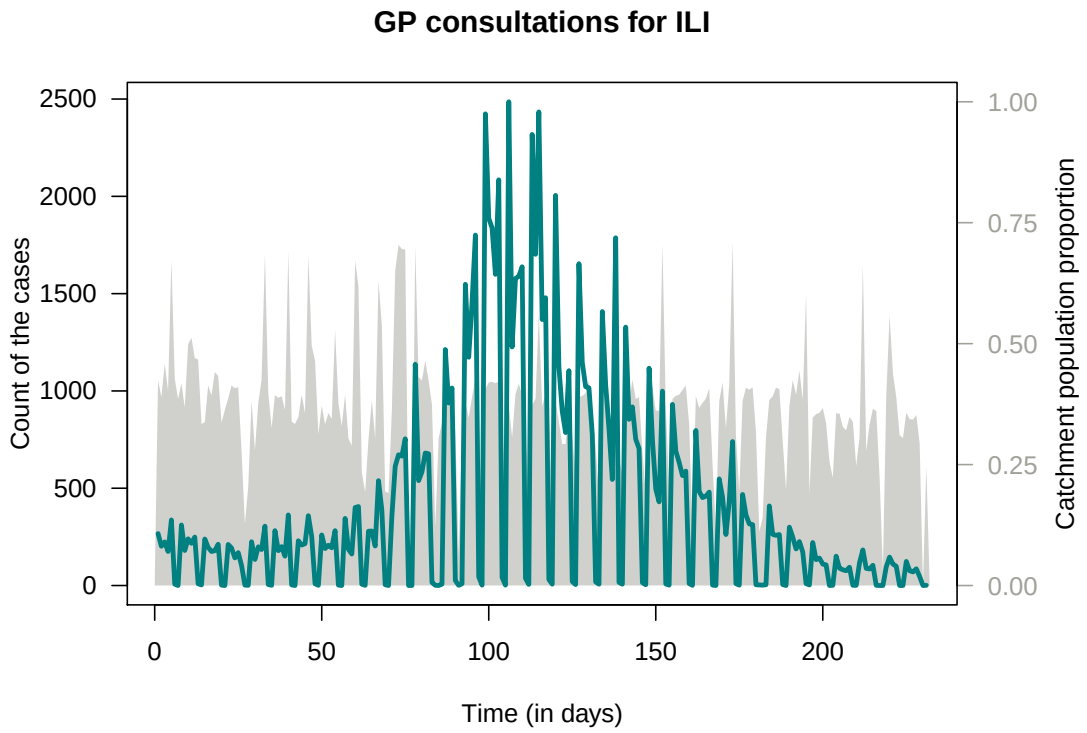


Figure 6.3: GP ILI data- Daily number of ILI GP consultations (green) and proportion of the population that was monitored each day (shaded grey).

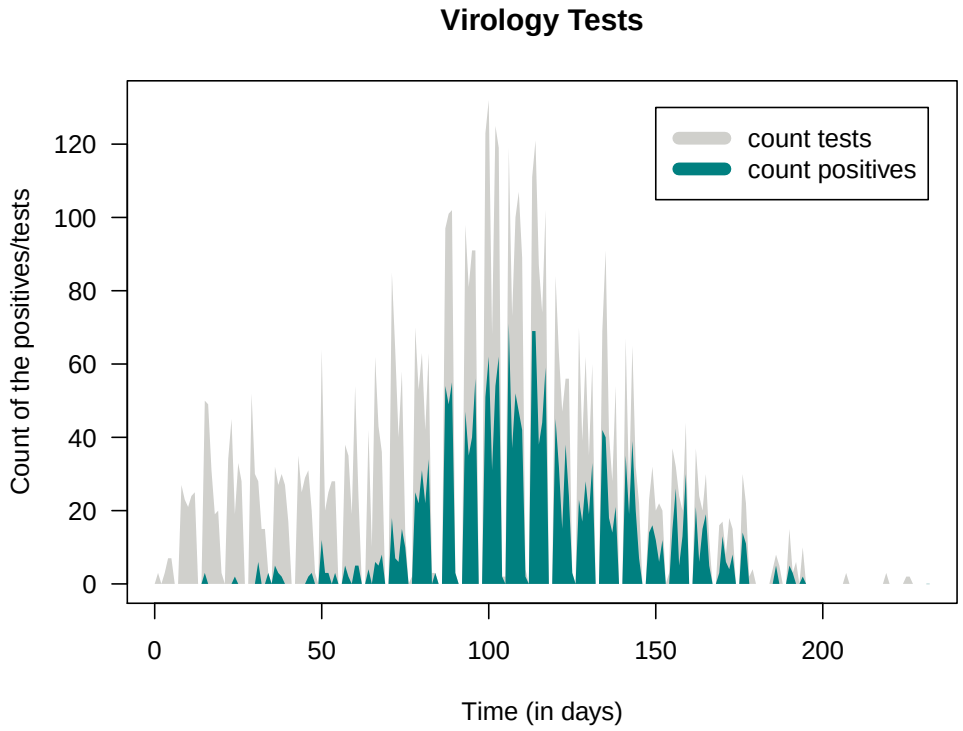


Figure 6.4: RCGP data - Daily number of positive tests for influenza (green) and number of tests sampled each day (grey).

6.2 Model formulation

A model for the joint analysis of GP consultations, virology, hospitalizations and ICU admissions data is here proposed. A few considerations are needed in order to motivate the model choices made in this section.

Firstly, GP-consultation data are highly affected by background noise (e.g. Birrell et al. (2011)) since they include both influenza- and non-influenza endemic ILI cases. These data need to be combined with data on virology that inform the proportion of influenza-positive ILI cases. In addition, these two components are related to the observed counts using an endemic/epidemic model (Held, Höhle, and Hofmann, 2005).

Secondly, while ICU and hospital data are highly dependent, elements that needs to be accounted in the inference, these two USISS datasets and the GP consultation data can be assumed independent: people admitted to hospital with confirmed influenza are likely to be emergency admissions and differ from mild cases.

Finally, the epidemic under analysis is a single wave of seasonal influenza. It is expected to take off every winter and to involve a large portion of the England population. For this reason, a deterministic transmission model is adopted to describe the underlying infection process. This deterministic model is then coupled with a fully-stochastic severity and detection model.

6.2.1 Notation and time

As described in Section 5.1, denote by ${}^{0}\theta^{\text{K}}$ the probability of experiencing event K conditionally on infection and by ${}^{\text{J}}\theta^{\text{K}}$ the probability of experiencing event K conditionally on having already experienced event J. Denote by ${}^{\text{J}}f_d^{\text{K}}$ the probability of experiencing a delay d between event J and K. Denote by X_t^{J} the number of people that have experienced event J at time t; by ${}_t^{\text{J}}X^{\text{K}}$ the number of people that will experience event K having had event J at time t; and by ${}_t^{\text{J}}X_s^{\text{K}}$ the number of people who have experienced event k at s having had event J at t.

Consider two different units of time, day and week, defining the discrete time of a process. Consider discrete time defined by intervals of length $\delta = 1$ day so that the u-th interval is defined as $[\delta u, \delta u + \delta)$ and the intervals are indexed by $u = 0, 1, 2, \ldots, U$. Consider discrete time defined by intervals of length $\delta = 7$ days so that the t-th interval is defined as $[\delta t, \delta t + \delta)$ and the intervals are indexed by $t = 0, 1, 2, \ldots, T$. The weekly quantities could be defined as sums of the daily quantities. However, in what follows both time scales are tracked, since both weekly and daily data are available.

6.2.2 Distributional assumptions

Transmission and first severity layer

Denote by ξ_u^0 the number of new infections generated during the uth interval. According to the assumption of a deterministic transmission model, this quantity is a function of the parameters $\pi, \iota, \beta, \sigma, \gamma, \kappa$, representing the proportion of individuals initially immune; the proportion of initially infected/infectious individuals;the transmission rate; the rate of becoming infectious; the recovery rate; and the school-closure effect, respectively. The number of new infections generated at week t from the beginning of the epidemic, denoted by φ_t^0, can be defined by

$\varphi_t^0 = \sum_{u=7(t-1)}^{7t} \xi_u^0$.

Denote by ${}_u^0X^{\mathrm{S}}$ the number of people that seek care (either going to the GP or going to hospital) and that have been infected on day u. These people are a Binomial sample of the new infections generated at u with probability ${}^0\theta^{\mathrm{S}}$. Since this probability is very small, the Binomial distribution can be approximated by a Poisson distribution. So that the number of new care-seekers infections results from a time non-homogeneous Poisson process, with rate varying according to ξ_u^0 and ${}^0\theta^{\mathrm{S}}$, i.e.:

$$\left({}_u^0X^{\mathrm{S}} | \xi_u^0, {}^0\theta^{\mathrm{S}}\right) \sim \mathrm{Pois}(\xi_u^0 \cdot {}^0\theta^{\mathrm{S}}) \qquad \text{for } u = 0, 1, \ldots, U$$

A number ${}_u^0X^{\mathrm{H}}$ of these people is admitted to hospital, while the remaining part, ${}_u^0X^{\mathrm{F}}$, seeks care at a GP clinic due to flu symptoms. This classification is assumed to happen according to a Binomial experiment from the total care seekers ${}_u^0X^{\mathrm{S}}$ with probability ${}^{\mathrm{S}}\theta^{\mathrm{H}}$. From the properties of the Poisson distribution, ${}_u^0X^{\mathrm{H}}$ and ${}_u^0X^{\mathrm{F}}$ are distributed according to two independent Poisson processes with respective rates $(1 - {}^{\mathrm{S}}\theta^{\mathrm{H}}) \cdot {}^0\theta^{\mathrm{S}} \cdot \xi_u^0$ and ${}^{\mathrm{S}}\theta^{\mathrm{H}} \cdot {}^0\theta^{\mathrm{S}} \cdot \xi_u^0$.

Since the parameters ${}^{\mathrm{S}}\theta^{\mathrm{H}}$ and ${}^0\theta^{\mathrm{S}}$ will always appear in such products, a re-parametrization can be made:

$$\begin{aligned} {}^0\theta^{\mathrm{H}} &= {}^0\theta^{\mathrm{S}} \cdot {}^{\mathrm{S}}\theta^{\mathrm{H}} \\ {}^0\theta^{\mathrm{F}} &= {}^0\theta^{\mathrm{S}} \cdot {}^{\mathrm{S}}\theta^{\mathrm{F}} \end{aligned}$$

where ${}^0\theta^{\mathrm{H}}$ and ${}^0\theta^{\mathrm{F}}$ represent the probability of hospitalization and flu-related GP visit given infection, respectively.

The daily processes of hospital admissions and flu-related GP consultations, are then described by the following equations:

$$\begin{aligned} \left({}_u^0X^{\mathrm{H}} \middle| \xi_u^0, {}^0\theta^{\mathrm{H}}\right) &\sim \mathrm{Pois}\left({}^0\theta^{\mathrm{H}} \cdot \xi_u^0\right) \qquad \text{for } u = 0, 1, \ldots, U \\ \left({}_u^0X^{\mathrm{F}} \middle| \xi_u^0, {}^0\theta^{\mathrm{F}}\right) &\sim \mathrm{Pois}\left({}^0\theta^{\mathrm{F}} \cdot \xi_u^0\right) \qquad \text{for } u = 0, 1, \ldots, U \end{aligned} \tag{6.1}$$

While the hospitalised cases might experience another level of severity (ICU admission), the flu-related GP consultations occur after some delay and are observed with some noise.

Flu related GP-consultation

Let ${}^0f_{0:C}^{\mathrm{F}} = \left({}^0f_0^{\mathrm{F}}, {}^0f_1^{\mathrm{F}}, \ldots {}^0f_C^{\mathrm{F}}\right)$ denote the vector of the probabilities of $0, 1, \ldots, C$ days elapsing between infection and the visit to the GP, respectively. The number of individuals ${}_u^0X_{u+c}^{\mathrm{F}}$ infected at u and visiting the GP after c days, with $c = 0, 1, \ldots C$, can then be thought as the realization of a Multinomial random variable (r.v.):

$$\left({}_u^0X_{u:u+C}^{\mathrm{F}} \middle| {}_u^0x^{\mathrm{F}}, {}^0f_{0:C}^{\mathrm{F}}\right) \sim \mathrm{Multi}({}_u^0x^{\mathrm{F}}, {}^0f_{0:C}^{\mathrm{F}}) \qquad \text{for } u = 0, 1, \ldots, U$$

From this, the number of flu-related GP visits at time u, X_u^{F}, can be expressed as a convolution of the Multinomial's components:

$$X_u^{\mathrm{F}} = \sum_{c=0}^{C} {}_{u-c}{}^0X_c^{\mathrm{F}} \qquad \text{for } u = 0, 1, \ldots, U$$

which, from the properties of the Poisson, is also distributed according to a time non-homogeneous Poisson process:

$$\left(X_u^{\mathrm{F}} \middle| {}^0\theta^{\mathrm{F}}, \xi_{1:U}^0, {}^0f_{0:C}^{\mathrm{H}}\right) \sim \mathrm{Pois}\left(\mu_u^{\mathrm{F}}\right) \qquad \text{for } u = 0, 1, \ldots, U$$

where $\mu_u^{\mathrm{F}} = {}^0\theta^{\mathrm{F}} \cdot \sum_{c=0}^{C} \xi_{u-c}^0 \cdot {}^0f_c^{\mathrm{F}}$. The variability of the flu-related GP cases is further affected by the uncertainty on the parameter ${}^0\theta^{\mathrm{F}}$, which can be described by a Gamma prior distribution with mean ι^{F} and variance $\frac{1}{\varepsilon^{\mathrm{F}}}$

$$\left({}^0\theta^{\mathrm{F}} \middle| \iota^{\mathrm{F}}, \varepsilon^{\mathrm{F}}\right) \sim \mathrm{Gamma}(\iota^{\mathrm{F}}\varepsilon^{\mathrm{F}}, \varepsilon^{\mathrm{F}}) \tag{6.2}$$

Background ILI GP consultation

GP-consultation data also include background cases which are cases not related to the influenza virus but that present symptoms similar to the flu-related cases. The background, endemic, cases are often assumed to follow a yearly seasonality, peaking around the same time as the seasonal influenza epidemic (Paul, Held, and Toschke, 2008).

The number of background cases at time u is here denoted by X_u^{B} and assumed to follow a Poisson distribution with time varying rate μ_u^{B}. This rate is assumed to be distributed according to a Gamma r.v. with mean b_u and variance $\frac{1}{\varepsilon^{\mathrm{B}}} = \frac{\sum_{c=0}^{C} \xi_{u-c}^0 \, {}^0f_c^{\mathrm{F}}}{\varepsilon^{\mathrm{F}}}$. The choice of a prior-variance increasing with the number of influenza cases, ξ_u^0, and with the variance of ${}^0\theta^{\mathrm{F}}$, $\frac{1}{\varepsilon^{\mathrm{F}}}$, is made for pragmatic reasons that will become clear in the formulation of the model for virological data. The above results in the following:

$$\begin{aligned} \left(X_u^{\mathrm{B}} \middle| \mu_u^{\mathrm{B}}\right) &\sim \mathrm{Pois}\left(\mu_u^{\mathrm{B}}\right) && \text{for } u = 0, 1, \ldots, U \\ \left(\mu_u^{\mathrm{B}} \middle| b_u, \varepsilon^{\mathrm{B}}\right) &\sim \mathrm{Gamma}(b_u\varepsilon^{\mathrm{B}}, \varepsilon^{\mathrm{B}}) && \text{for } u = 0, 1, \ldots, U \end{aligned} \tag{6.3}$$

The seasonality pattern of the endemic cases is modelled by a weekly-varying sine-cosine oscillation of the mean of the background process, b_u:

$$b_u = \exp\left\{\nu_1 + \nu_2 \cos\left(\frac{2\pi t_u}{52}\right) + \nu_3 \sin\left(\frac{2\pi t_u}{52}\right)\right\}$$

which is constant for all the days u in the same week t_u.

GP consultation data distribution

The total number of GP consultations includes both influenza-related cases and endemic background cases. The sum of these two processes is denoted here by X_u^{G}

$$X_u^{\mathrm{G}} = X_u^{\mathrm{F}} + X_u^{\mathrm{B}} \qquad \text{for } u = 0, 1, \ldots, U$$

This process is Poisson-distributed with time-varying rate $\mu_u^{\mathrm{F}} + \mu_u^{\mathrm{B}}$:

$$\left(X_u^{\mathrm{G}} \middle| \mu_u^{\mathrm{F}}, \mu_u^{\mathrm{B}}\right) \sim \mathrm{Pois}\left(\mu_u^{\mathrm{F}} + \mu_u^{\mathrm{B}}\right) \qquad \text{for } u = 0, 1, \ldots, U \tag{6.4}$$

The probability of detecting an ILI case, conditionally on visiting the GP is proportional to the day-specific catchment population of the practices participating in the collection scheme on day u, here denoted by ζ_u^{G}.

Moreover, the probability of attending a GP practice is affected by weekly fluctuations, caused by the weekend closure of GP practices, so that patients affected by ILI symptoms further delay their visit until the earliest days of the following week. This is a day-of-the-week distortion that can be modelled by a day-of-the-week factor shrinking the weekend rate and inflating the weekday rates as proposed in Birrell et al. (2016). This day-of the week fluctuation can be detected from the data as shown in Figure 6.5.

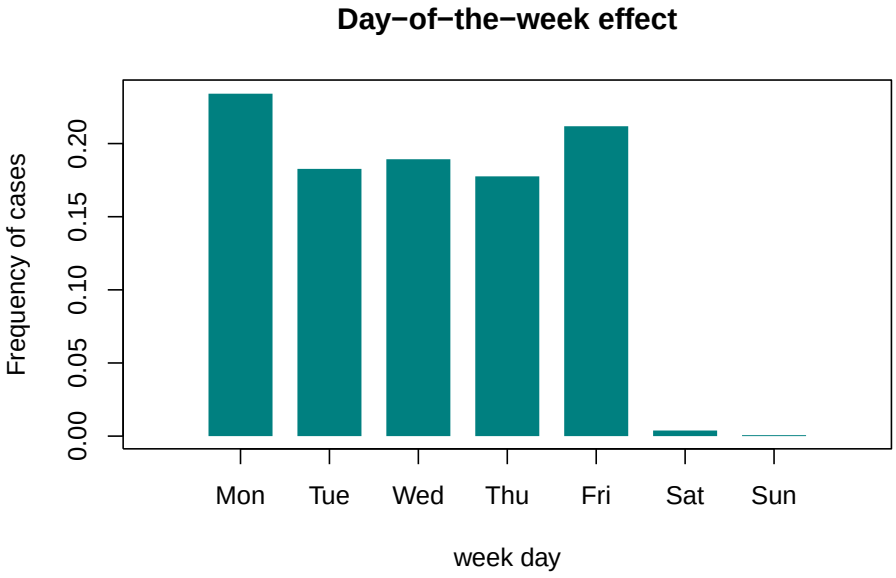


Figure 6.5: Observed of ILI GP consultations by day of the week.

Let $\lfloor \cdot \rfloor$ denote the day-of-the-week operator that assigns the week-day (1=Mon, 2=Tue, 3=Wed, 4=Thu, 5=Fri, 6=Sat, 7=Sun) to each day u. The day-of the week effect is then denoted by the parameter $\omega_{\lfloor u \rfloor}$, for $\lfloor u \rfloor = 1, 2, 3, 4, 5, 6, 7$. Interpreting $\omega_{\lfloor u \rfloor}$ as a distortion factor, for identifiability reason, it is useful to assume that its geometric mean over the 7 days is equal to 1, so that the rate of GP consultation is only re-distributed over the 7 week days according to the opening time of the GP practices, leading to:

$$\prod_{i \in \{1,2,3,4,5,6,7\}} \omega_i = 1$$

The reported number of GP consultations for ILI at time u, Y_u^{G} is therefore assumed to be distributed according to a Binomial r.v. having size equal to the cases X_u^{G} and with probability $\zeta_u^{\mathrm{G}} \cdot \omega_{\lfloor u \rfloor}$:

$$\left(Y_u^{\mathrm{G}} \middle| x_u^{\mathrm{G}}, \zeta_u^{\mathrm{G}}, \omega_{\lfloor u \rfloor} \right) \sim \mathrm{Bin}\left(x_u^{\mathrm{G}}; \zeta_u^{\mathrm{G}} \omega_{\lfloor u \rfloor} \right) \qquad \text{for } u = 0, 1, \ldots, U$$

Given Equation 6.4 and the properties of the Poisson distribution, Y_u^{G}, conditionally on the parameters composing μ_u^{F} and μ_u^{B}, is also Poisson-distributed around the overall mean μ_u^{G}, i.e.:

$$\begin{aligned} \left(Y_u^{\mathrm{G}} \middle| {}^0\theta^{\mathrm{F}}, \xi_{1:U}^0, {}^0 f_{0:C}^{\mathrm{F}}, \mu_u^{\mathrm{B}}, \zeta_u^{\mathrm{G}}, \omega_{\lfloor u \rfloor} \right) &\sim \mathrm{Pois}\left(\zeta_u^{\mathrm{G}} \omega_{\lfloor u \rfloor} \left(\mu_u^{\mathrm{F}} + \mu_u^{\mathrm{B}} \right) \right) \\ &\sim \mathrm{Pois}\left(\underbrace{\zeta_u^{\mathrm{G}} \omega_{\lfloor u \rfloor} \mu_u^{\mathrm{F}} + \zeta_u^{\mathrm{G}} \omega_{\lfloor u \rfloor} \mu_u^{\mathrm{B}}}_{\mu_u^{\mathrm{G}}} \right) \\ &\sim \mathrm{Pois}\left(\mu_u^{\mathrm{G}} \right) \end{aligned} \tag{6.5}$$

for $u = 0, 1, \ldots, U$. Given the two Gamma priors on the parameters ${}^0\theta^{\mathrm{F}}$ and μ_u^{B} (Equation 6.3 and 6.2), the rate of this Poisson, μ_u^{G}, can be approximated by a Gamma (Stewart et al., 2007) with the following parameters (derivation shown in Appendix F.1):

$$\left(\mu_u^{\mathrm{G}} \middle| \xi^0_{1:U}, {}^0f^{\mathrm{F}}_{0:C}, \zeta^{\mathrm{G}}_u, \omega_{\lfloor u \rfloor}, \iota^{\mathrm{F}}, \varepsilon^{\mathrm{F}}, g_u, \varepsilon^{\mathrm{B}}\right) \sim \text{Gamma}\left(\varepsilon^{\mathrm{F}}\iota^{\mathrm{F}} + \frac{b_u\varepsilon^{\mathrm{F}}}{\sum_{c=0}^{C}\xi^0_{u-c}{}^0f^{\mathrm{F}}_c}; \frac{\varepsilon^{\mathrm{F}}}{\zeta^{\mathrm{G}}_u\omega_{\lfloor u \rfloor}\sum_{c=0}^{C}\xi^0_{u-c}{}^0f^{\mathrm{F}}_c}\right)$$

for $u = 0, 1, \ldots, U$. A Poisson r.v. centred on a Gamma-distributed rate results in a Negative Binomial r.v. (see Appendix A). Therefore, a final formulation of the data distribution, depending on the hyper-parameters is:

$$\left(Y_u^{\mathrm{G}} \middle| \xi^0_{1:U}, {}^0f^{\mathrm{F}}_{0:C}, \zeta^{\mathrm{G}}_u, \omega_{\lfloor u \rfloor}, \iota^{\mathrm{F}}, \varepsilon^{\mathrm{F}}, g_u, \varepsilon^{\mathrm{B}}\right) \sim \text{NegBinom}\left(\varepsilon^{\mathrm{F}}\iota^{\mathrm{F}} + \frac{b_u\varepsilon^{\mathrm{F}}}{\sum_{c=0}^{C}\xi^0_{u-c}{}^0f^{\mathrm{F}}_c}; 1 + \frac{\zeta^{\mathrm{G}}_u\omega_{\lfloor u \rfloor}\sum_{c=0}^{C}\xi^0_{u-c}{}^0f^{\mathrm{F}}_c}{\varepsilon^{\mathrm{F}}}\right)$$

for $u = 0, 1, \ldots, U$; with parametrization $\text{NegBinom}(r, \eta)$ as in Appendix A.

Virology data distribution

The virology data are exploited to uncover the proportion of ILI cases actually affected by the influenza virus. The number of positive tests at each day u, Y_u^{V}, is assumed to be a Binomial sample from the number of tests taken N_u^{V}. The probability of each sample being positive is equal to $\frac{\mu_u^{\mathrm{F}}}{\mu_u^{\mathrm{F}}+\mu_u^{\mathrm{B}}}$, i.e. the rate of flu ILI during day u, over the rate of any ILI during day u. So:

$$\left(Y_u^{\mathrm{V}} \middle| N_u^{\mathrm{V}}, \mu_u^{\mathrm{F}}, \mu_u^{\mathrm{B}}\right) \sim \text{Binom}\left(N_u^{\mathrm{V}}, \frac{\mu_u^{\mathrm{F}}}{\mu_u^{\mathrm{F}} + \mu_u^{\mathrm{B}}}\right) \qquad \text{for } u = 0, 1, \ldots, U$$

As μ_u^{F} and μ_u^{B} are assumed Gamma-distributed with the same rate parameter $\frac{\varepsilon^{\mathrm{F}}}{\sum_{c=0}^{C}\xi^0_{u-c}{}^0f^{\mathrm{F}}_c}$ (this follows from Equations 6.2 and 6.3 as shown in Appendix F), the quantity $\frac{\mu_u^{\mathrm{F}}}{\mu_u^{\mathrm{F}}+\mu_u^{\mathrm{B}}}$ is distributed as a Beta r.v. (see Appendix A):

$$\left(\frac{\mu_u^{\mathrm{F}}}{\mu_u^{\mathrm{F}} + \mu_u^{\mathrm{B}}} \middle| b_u, \xi^0_{1:U}, {}^0f^{\mathrm{F}}_{0:C}, \iota^{\mathrm{F}}, \varepsilon^{\mathrm{F}}\right) \sim \text{Beta}\left(\iota^{\mathrm{F}}\varepsilon^{\mathrm{F}}, b_u\frac{\varepsilon^{\mathrm{F}}}{\sum_{c=0}^{C}\xi^0_{u-c}{}^0f^{\mathrm{F}}_c}\right) \qquad \text{for } u = 0, 1, \ldots, U$$

This results in the virology data distribution to be rewritten as the Beta-Binomial distribution:

$$\left(Y_u^{\mathrm{V}} \middle| N_u^{\mathrm{V}}, b_u, \xi^0_{1:U}, {}^0f^{\mathrm{F}}_{0:C}, \iota^{\mathrm{F}}, \varepsilon^{\mathrm{F}}\right) \sim \text{BetaBin}\left(N_u^{\mathrm{V}}, \iota^{\mathrm{F}}\varepsilon^{\mathrm{F}}, b_u\frac{\varepsilon^{\mathrm{F}}}{\sum_{c=0}^{C}\xi^0_{u-c}{}^0f^{\mathrm{F}}_c}\right)$$

for $u = 0, 1, \ldots, U$.

Hospitalization and ICU admissions data distribution

A model for the hospitalizations and ICU-admissions data is proposed in Chapter 5, with the two data streams analysed jointly, since they are intrinsically dependent: the model equations are reported again below. Denote by ${}^{\mathrm{H}}\theta^{\mathrm{IC}}$ the probability of ICU admission given hospitalization; by ${}^0f_d^{\mathrm{H}}$ and ${}^{\mathrm{H}}f_d^{\mathrm{IC}}$ the discrete probability of d weeks elapsing between infection and hospitalization and hospitalization and ICU admission, respectively; by ζ_t^{H} and ζ_t^{IC} the probability of being detected in hospital and in ICU, respectively.

The data-generating process can be expressed as a series of Binomial and Multinomial steps from the initial Poisson process of the infections ${}^0_tX^{\mathrm{H}}$ that will be hospitalised (Equation 6.1):

$$\begin{aligned}
\left({}^{0}_{t}X^{\mathrm{H}}_{t:t+D}\middle|{}^{0}_{t}x^{\mathrm{H}}, {}^{0}f^{\mathrm{H}}_{0:D}\right) &\sim \mathrm{Multi}({}^{0}_{t}x^{\mathrm{H}}, {}^{0}f^{\mathrm{H}}_{0:D}) \quad , \quad X^{\mathrm{H}}_{t} = \sum_{d=0}^{D} {}^{t-d}{}^{0}X^{\mathrm{H}}_{d} \\
\left({}^{\mathrm{H}}_{t}X^{\mathrm{IC}}\middle|x^{\mathrm{H}}_{t}, {}^{\mathrm{H}}\theta^{\mathrm{IC}}\right) &\sim \mathrm{Bin}(x^{\mathrm{H}}_{t}, {}^{\mathrm{H}}\theta^{\mathrm{IC}}) \\
\left({}^{\mathrm{H}}_{t}X^{\mathrm{IC}}_{t:t+D}\middle|{}^{\mathrm{H}}_{t}x^{\mathrm{IC}}, {}^{\mathrm{H}}f^{\mathrm{IC}}_{0:D}\right) &\sim \mathrm{Multi}({}^{\mathrm{H}}_{t}x^{\mathrm{IC}}, {}^{\mathrm{H}}f^{\mathrm{IC}}_{0:D}) \quad , \quad X^{\mathrm{IC}}_{t} = \sum_{d=0}^{D} {}^{t-d}{}^{\mathrm{H}}X^{\mathrm{IC}}_{d} \\
\left(Y^{\mathrm{H}}_{t}\middle|x^{\mathrm{H}}_{t}, \zeta^{\mathrm{H}}_{t}\right) &\sim \mathrm{Bin}(x^{\mathrm{H}}_{t}, \zeta^{\mathrm{H}}_{t}) \\
\left(Y^{\mathrm{IC}}_{t}\middle|x^{\mathrm{IC}}_{t}, \zeta^{\mathrm{IC}}_{t}\right) &\sim \mathrm{Bin}(x^{\mathrm{IC}}_{t}, \zeta^{\mathrm{IC}}_{t})
\end{aligned}$$

for time $t = 1, \ldots, T$. This model leads to the two marginal data distributions:

$$\left(Y^{\mathrm{H}}_{t}\middle|\zeta^{\mathrm{H}}_{t}, {}^{0}\theta^{\mathrm{H}}, \xi^{0}_{0:T}, {}^{0}f^{\mathrm{H}}_{0:D}\right) \sim \mathrm{Pois}\left(\zeta^{\mathrm{H}}_{t} \cdot {}^{0}\theta^{\mathrm{H}} \cdot \sum_{d=0}^{D} \varphi^{0}_{t-d} \cdot {}^{0}f^{\mathrm{H}}_{d}\right)$$

$$\left(Y^{\mathrm{IC}}_{t}\middle|\zeta^{\mathrm{H}}_{t}, {}^{0}\theta^{\mathrm{H}}, \xi^{0}_{0:T}, {}^{0}f^{\mathrm{H}}_{0:D}, {}^{\mathrm{H}}\theta^{\mathrm{IC}}, {}^{0}f^{\mathrm{IC}}_{0:D}\right) \sim \mathrm{Pois}\left(\zeta^{\mathrm{IC}}_{t} \cdot {}^{\mathrm{H}}\theta^{\mathrm{IC}} \cdot {}^{0}\theta^{\mathrm{H}} \cdot \sum_{d=0}^{D}\sum_{g=0}^{d} \varphi^{0}_{t-d-g} \cdot {}^{0}f^{\mathrm{H}}_{d} \cdot {}^{\mathrm{H}}f^{\mathrm{IC}}_{g}\right)$$

As in Chapter 5, to compute the joint distribution of $(Y^{\mathrm{H}}_{t}; Y^{\mathrm{IC}}_{t})$ for $t = 1, \ldots, T$, Algorithm 7 is adopted: the marginal distribution of the ICU admissions $Y^{\mathrm{IC}}_{1:T}$ is computed from the Poisson distribution above and the distribution of the hospitalizations $Y^{\mathrm{H}}_{1:T}$ conditionally on ICU data is approximated via Monte Carlo (MC) integration.

Note that hospitalization data tend to have a further level of variability compared to ICU data. Both datasets are affected by reporting noise (i.e. the hospitals that should report do not always comply), but hospitalization data are also affected by a further under ascertainment and noise due to the fact that only sentinel trusts are selected, possibly compromising the representativeness of the data. For this reason the parameter of detection of hospitalised cases, ζ^{H}_{t}, is assumed to be distributed according to a Beta r.v. around the observed proportion of the catchment population of the trusts reporting at time t over the population of England, denoted by d^{H}_{t} (Equation 6.6). The probability of detection in ICU admissions, instead, is assumed to be exactly equal to the observed proportion of the catchment population of the ICUs reporting at time t over the population of England: $\zeta^{\mathrm{IC}}_{t} = d^{\mathrm{IC}}_{t}$. Thus,

$$\left(\zeta^{\mathrm{H}}_{t}\middle|d^{\mathrm{H}}_{t}, \varepsilon^{\mathrm{H}}\right) \sim \mathrm{Beta}\left(\frac{d^{\mathrm{H}}_{t}}{1 - d^{\mathrm{H}}_{t}}\varepsilon^{\mathrm{H}}, \varepsilon^{\mathrm{H}}\right) \tag{6.6}$$

Whilst introducing the Beta prior on ζ^{H}_{t}, the data distribution of the hospitalizations data conditionally on the hospitalised cases, $\left(Y^{\mathrm{H}}_{t}\middle|x^{\mathrm{H}}_{t}, d^{\mathrm{H}}_{t}, \varepsilon^{\mathrm{H}}\right)$, can be rewritten as:

$$\left(Y^{\mathrm{H}}_{t}\middle|x^{\mathrm{H}}_{t}, d^{\mathrm{H}}_{t}, \varepsilon^{\mathrm{H}}\right) \sim \mathrm{BetaBin}\left(x^{\mathrm{H}}_{t}, \frac{d^{\mathrm{H}}_{t}}{1 - d^{\mathrm{H}}_{t}}\varepsilon^{\mathrm{H}}, \varepsilon^{\mathrm{H}}\right)$$

which can be included straightforwardly in the final approximation step of Algorithm 7 of Chapter 5.

6.2.3 Prior distributions

The prior distributions on the parameters involved in the model are described below and further summarised in Table 6.1.

- In the **transmission model**, uniform priors are assumed for the transmission rate, β, and the initial proportion of exposed/infectious individuals, ι. The initial proportion of immune people, π, is assumed to be distributed as a Beta r.v. centred in 0.375; this result was obtained from the analysis of serological data at the end of the previous season by collaborators at PHE (Charlett, 2018). The average latent period is assumed to be Log-Normally distributed with mean 2 days and the average infectious period is assumed known and equal to 3.14 days; these two values are taken from Birrell et al. (2011). The holidays factor, κ, models the increase or decrease in infection rate during school closure as follows:

$$S_{t+1} = S_t - (1+\kappa)\cdot\beta\delta S_t\frac{I_t}{N} \qquad \text{for } t \in \text{school holidays}$$

  κ can take any value between -1 and $+\infty$, with negative values indicating a decrease of infectiousness during school holidays and positive values indicating an increase of infectiousness during school closure. This parameter is assigned a shifted Log-Normal prior distribution centred on 0. Lastly, other parameters involved in the transmission process are: the population size, fixed to the latest available data from the Office of National Statistics (ONS), $N = 55268100$, i.e. the mid-2016 estimates (Office of National Statistics, 2017); and the length of the interval at which the discrete-time transmission model is evaluated, which is set to $\delta = 0.25$ day.

- The **main severity parameters** are ${}^0\theta^{\text{F}}$, ${}^0\theta^{\text{H}}$ and ${}^{\text{H}}\theta^{\text{IC}}$. The first probability of flu-related GP consultations, ${}^0\theta^{\text{F}}$, has hyper-parameters ι^{F} and ε^{F}. The former is given a Uniform prior between 0 and 1 and the precision ε^{F} is given a Uniform prior between 0 and 5000. Sampling directly from the priors leads to a Uniform-like distribution between 0 and 1 for ${}^0\theta^{\text{F}}$ (see plots in the result section). ${}^0\theta^{\text{H}}$ and ${}^{\text{H}}\theta^{\text{IC}}$ are both given Uniform priors between 0 and 1.

- The discrete **waiting-times** between consecutive severe events are assumed known. The distribution of the time from infection to flu-related GP consultation, ${}^0f^{\text{F}}_{0:C}$, is obtained by discretising in days the density of a Gamma (3.41,0.83) (taken from the sum of the prior distributions assumed in Birrell et al. (2011)). The distribution of the time from infection to hospitalization, ${}^0f^{\text{H}}_{0:D}$, and from hospitalization to ICU admission, ${}^hf^{\text{IC}}_{0:D}$, are obtained by discretising in days two Exponential distributions with rate 0.32 and 0.4, respectively (from the analysis of individual data from the USISS sentinel scheme).

- The mean of the **background ILI** consultations is centred on a sine-cosine transformation of the weeks. The parameters ν_1, ν_2 and ν_3 specifying this behaviour are assigned an informative prior obtained from fitting an *HHH model* (Held, Höhle, and Hofmann, 2005) to GP data from January 2015 to September 2017. This analysis is reported in Appendix F.2. The posterior samples of the endemic parameters are used to formulate independent Normal prior distributions on ν_1, ν_2 and ν_3.

- Six parameters for the **day of the week effect** are to be estimated (ω_4, the effect of Thursday, can be derived from the others thanks to the condition of the geometric mean being equal to 1). The parameter vector $(\omega_1, \omega_2, \omega_3, \omega_5, \omega_6, \omega_7)$ is assigned multivariate Log-Normal prior centred in 1, implicitly assuming no effect of the day of the week, and every $\log(\omega)$ is given variance equal to 0.4.

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_5 \\ \omega_6 \\ \omega_7 \end{pmatrix} \sim \text{Log-Normal}(\log(1), V_\omega)$$

  V_ω is formulated so that the equality of variances is preserved also for the parameter $\omega_4 = 1 \Big/ \prod_{i \in 1,2,3,5,6,7} \omega_i$. The derivation, reported in in Appendix F.3 and in Birrell et al. (2016), leads to:

$$V_\omega = 0.16 \begin{pmatrix} 1 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 \\ -1/6 & 1 & -1/6 & -1/6 & -1/6 & -1/6 \\ -1/6 & -1/6 & 1 & -1/6 & -1/6 & -1/6 \\ -1/6 & -1/6 & -1/6 & 1 & -1/6 & -1/6 \\ -1/6 & -1/6 & -1/6 & -1/6 & 1 & -1/6 \\ -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & 1 \end{pmatrix}$$

- The **detection parameters** are mainly informed by the catchment population of the reporting trusts or GP clinics, as a proportion of the population of England, denoted by d_u^{G} for GP data, d_t^{H} for hospital data and d_t^{IC} for ICU data. Specifically: $\zeta_u^{\text{G}} = d_u^{\text{G}}$ and $\zeta_t^{\text{IC}} = d_t^{\text{IC}}$, while ζ_t^{H}, as mentioned above, is given a Beta prior with hyper-parameters d_t^{H}, fixed, and ϵ^{H}, influencing the precision of the data, being assigned a uniform prior between 0 and 100.

Parameter	Name	Support	Prior
Transmission rate	β	$[0, \infty)$	Uniform(0,4)
Initial immunity	π	$[0, 1]$	Beta(37.5,62.5)
Initial exposed/infectious	ι	$[0, 1]$	Uniform(0, 0.1)
Average latent period	$d_L = 2/\sigma$	$[0.5, +\infty)$	Log-Normal(log(2),0.5)
Average infectious period	$d_I = 2/\gamma$	-	PP at 3.14
Total population	N	-	PP at 55268100
Factor for holidays	κ	$[-1, +\infty)$	shifted Log-Normal(0,1)
Mean of Case flu-symptoms risk	ι^{F}	$[0, 1]$	Uniform(0,1)
Precision of Case flu-symptoms risk	ε^{F}	$[0, +\infty]$	Uniform(0,5000)
Case hospitalization risk	${}^{0}\theta^{\mathrm{H}}$	$[0, 1]$	Uniform(0,1)
Hospital ICU admission risk	${}^{\mathrm{H}}\theta^{\mathrm{IC}}$	$[0, 1]$	Uniform(0,1)
Prob. of time infection $\rightarrow$ flu visit	${}^{0}f^{\mathrm{F}}_{0:C}$	-	PP Gamma (3.41,0.83)
Prob. of time infection $\rightarrow$ hospitalization	${}^{0}f^{\mathrm{H}}_{0:C}$	-	PP Exp (0.32)
Prob. of time hospitalization $\rightarrow$ ICU	${}^{\mathrm{H}}f^{\mathrm{IC}}_{0:C}$	-	PP Exp (0.4)
HHH model parameter 1	ν_1	$(-\infty, +\infty)$	Norm(4.66, 0.17)
HHH model parameter 2 (sin)	ν_2	$(-\infty, +\infty)$	Norm(-0.2, 0.11)
HHH model parameter 3 (cos)	ν_3	$(-\infty, +\infty)$	Norm(0.99, 0.08)
Monday, Tuesday, Wednesday, Friday, Saturday, Sunday effect	$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_5 \\ \omega_6 \\ \omega_7 \end{pmatrix}$	$[0, +\infty)$	Log-Normal(log(1),V_ω)
Thursday effect	ω_4	-	$1 \Big/ \prod_{i \in 1,2,3,5,6,7} \omega_i$
GP consultation detection	ζ^{G}_u	-	PP at observed d^{G}_u
Mean of hospitalization detection	-	-	PP at observed d^{H}_u
Shape parameter of hospitalization detection	ε^{H}	$[0, +\infty)$	Uniform(0,100)
ICU admission detection	ζ^{IC}_u	-	PP at observed d^{IC}_u

Table 6.1: Prior distributions of all the parameters of the model. The parameters in grey are assumed known and fixed while the others are estimated. *PP* stands for *Point Prior*.

6.3 Results

A retrospective analysis of all the available data is performed. The goal of this analysis is to obtain the joint posterior distribution for the unknown parameter vector: $\boldsymbol{\theta} = (\beta, \pi, \iota, d_L, \kappa, \iota^{\mathrm{F}}, \epsilon^{\mathrm{F}}, {}^{0}\theta^{\mathrm{H}}, {}^{\mathrm{H}}\theta^{\mathrm{IC}}, \nu_1, \nu_2, \nu_3, \varepsilon^{\mathrm{H}}, \omega_1, \omega_2, \omega_3, \omega_5, \omega_6, \omega_7)$.

Specifics of the algorithm used to sample from the joint posterior distribution, as well as results and model assessment are reported below.

6.3.1 Specifics of the inference

All the elements of $\boldsymbol{\theta}$ are transformed to lie in $(-\infty, +\infty)$, and a bespoke Metropolis Hastings (MH) sampler is coded to carry out the analysis.

The analysis is composed of two phases: firstly a Gibbs-sampler-like algorithm samples each element of the vector $\boldsymbol{\theta}$ conditional on the others and in the second block-update phase the whole parameter vector $\boldsymbol{\theta}$ is sampled jointly. In the Gibbs sampling, a Normal random walk is run on the transformed parameter space of $\boldsymbol{\theta}$; the proposal Normal distributions have parameter-specific standard deviations s_β, s_π, s_ι, and so on.

The first phase comprises: 10,000 adaptation iterations, where the values $s_\beta, s_\pi, s_\iota, \ldots, s_{\epsilon^{\mathrm{F}}}$ are adapted to lead to a desirable acceptance rate (between 0.2 and 0.3); 50,000 burn-in iterations, discarded; and 100,000 sampling iterations, saved with a thinning factor of 500. Three parallel chains are run, resulting in 6,000 samples for each parameter.

The samples are used to estimate $\mathcal{S}$, the variance-covariance matrix for the 19-variate Normal proposal for the second block-update phase: the observed variance-covariance matrix of the sample is multiplied by a factor which is adapted for the first 100,000 iterations of the block-update algorithm; 100,000 iterations are then discarded as burn-in; 1,000,000 iterations are sampled with a thinning factor of 200. Three parallel chains are run, resulting in 15,000 samples for each parameter.

The MH algorithm was coded in `R` (R Core Team, 2018). However the most computationally-expensive step is the evaluation of the joint likelihood of the hospital and ICU data, for which the MC algorithm proposed in Chapter 5 and coded in `Rcpp` (Eddelbuettel and François, 2018) is used.

The whole analysis ran on a low-performance laptop due to confidentiality constraints given by the data provider, taking in total twenty days.

6.3.2 Parameter estimation

In what follows, unless specified otherwise, prior distributions are represented in red and posterior distributions are represented in green. Summary statistics of the posterior distributions are reported in Table 6.2.

Transmission

Figure 6.6 reports the prior-to-posterior plots of the five parameters of the transmission model. As already shown in Chapter 3, some parameters are highly (or uniquely) informed by the prior distributions. This is the case for the initial immunity π, for which the posterior distribution

coincides with its prior, and the average latent period, that is also highly influenced by its prior. Data are characterised by a sudden increase in cases at the end of December and by a prolonged influenza-peak activity. These characteristics result in a shift of d_L towards higher values, which leads to a longer epidemic, and by posterior values for κ, the school holiday factor, higher than one, allowing transmission to increase during the Christmas break.

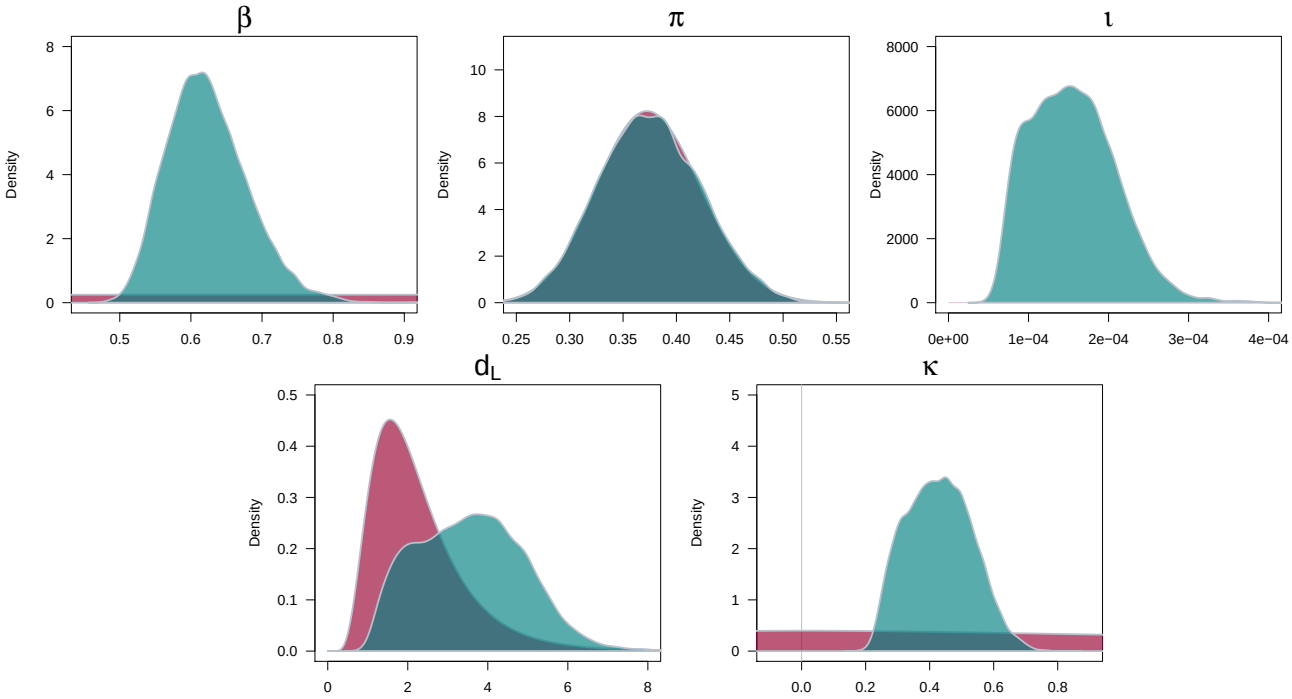


Figure 6.6: Prior (red) and posterior (green) distributions of the transmission parameters: the transmission rate β, the initial immunity π, the initial proportion of exposed/infectious ι, the average latent period d_L, the factor for school holidays κ.

The basic and effective reproduction numbers, R_0 and $R_e(0)$, are useful summaries of the transmission intensity over the season. Their prior-to-posterior plots are reported in Figure 6.7, both during school periods and school holidays. During school holidays, transmission becomes more intense, but also more variable, given the uncertainty around the parameter κ.

Finally, to have an overall picture of transmission over the course of the epidemic, the 95% credible intervals (CrIs) for the daily number of new infections are plotted in Figure 6.8. The trajectories clearly show the breakpoints in transmission.

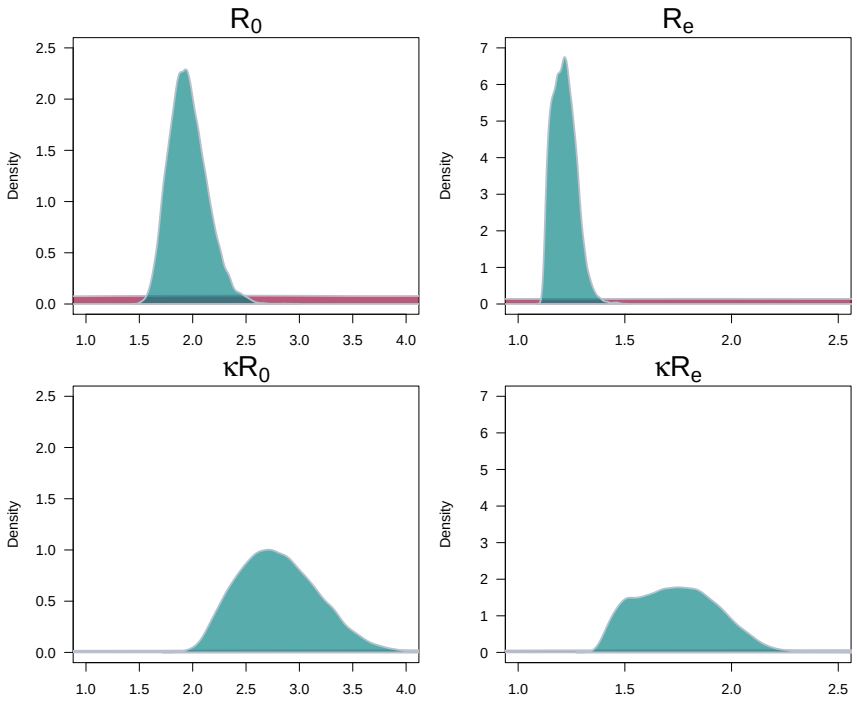


Figure 6.7: Prior (red) and posterior (green) distribution of the basic and effective reproduction number during school periods (top panels) and during holiday periods (bottom panels).

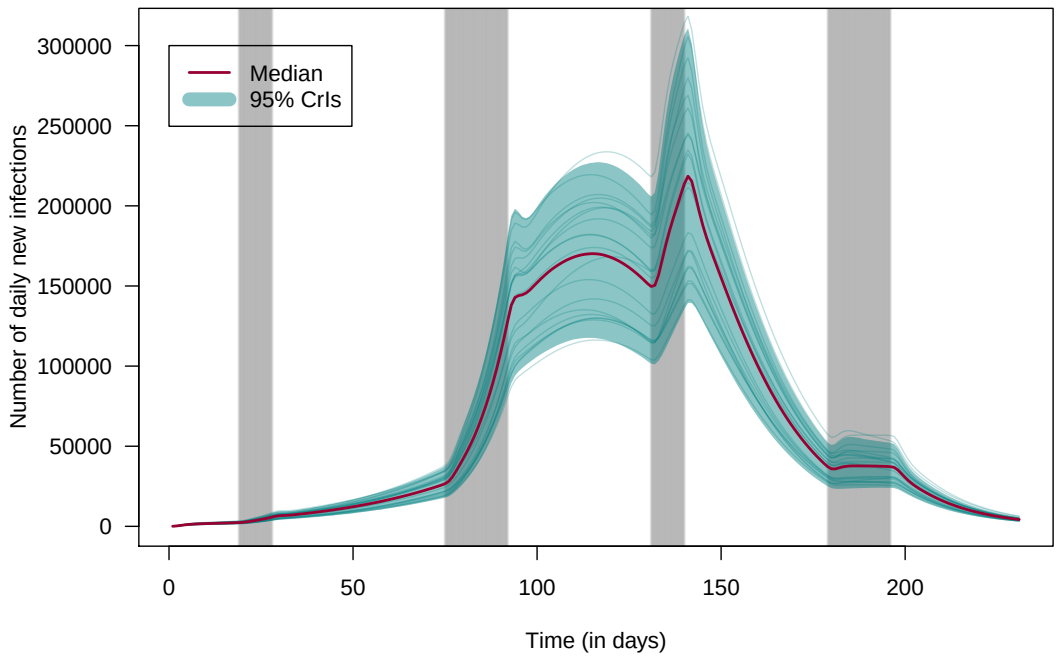


Figure 6.8: Median (red) and 95% CrIs (green) for the daily number of new infections. The grey areas represent the days corresponding to school holidays. 20 randomly selected trajectories are also computed and plotted as thin green lines.

Severity

The severity parameters include: ι^{F} and ε^{F}, i.e. the mean and the precision of the probability of GP consultation given infection, ${}^{0}\theta^{\text{F}}$; the probability of hospitalization given infection, ${}^{0}\theta^{\text{H}}$; and the probability of ICU admissions given hospitalization, ${}^{\text{H}}\theta^{\text{IC}}$.

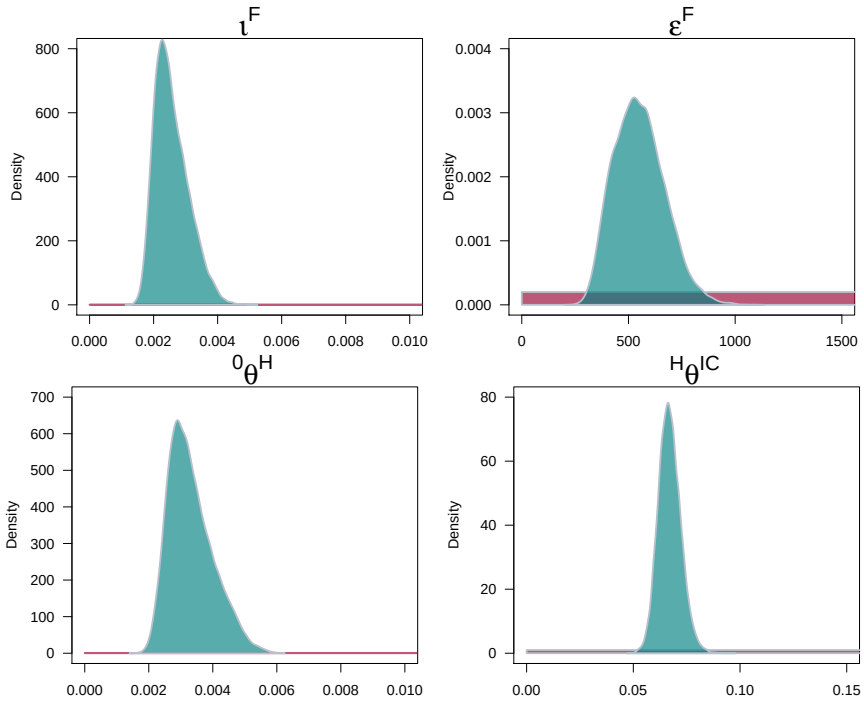


Figure 6.9: Prior (red) and posterior (green) distributions of the severity parameters: ι^{F}, the mean of the case-GP-consultation risk ${}^{0}\theta^{\text{F}}$; its precision ε^{F}; the case hospitalization risk ${}^{0}\theta^{\text{H}}$; and the probability of ICU admission given hospitalization ${}^{\text{H}}\theta^{\text{IC}}$.

The interpretation of the magnitude of ${}^{0}\theta^{\text{F}}$ (and the related ι^{F} and ε^{F}) necessitates some caution because GP visits are also affected by the day-of-the-week effect. As shown below in the results on the day-of-the-week effect, the risk of GP visit can become four times bigger or 20 times smaller according to the day of the week at which the visit takes place.

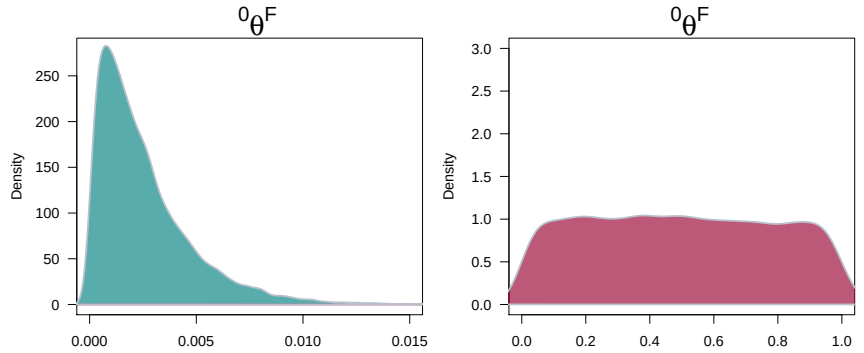


Figure 6.10: Distribution of ${}^{0}\theta^{\text{F}}$ under the posterior (left) and prior (right) distribution of the parameters ι^{F} and ε^{F}

The prior-to-posterior plots of the parameters describing severity are reported in Figure 6.9. All the parameters are highly informed by the data and the posterior distributions give a clear picture of the severity during the 2017/18 influenza epidemic. The probability of hospitaliza-

tion given infection is around 0.3% and the probability of ICU admission given hospitalization stretches between 6 and 7%.

To interpret the results on ι^{F} and ε^{F}, it is useful to draw the distribution of ${}^{0}\theta^{\mathrm{F}}$ under the prior and posterior distributions of these two parameters. Figure 6.10 shows that the prior knowledge on ${}^{0}\theta^{\mathrm{F}}$ was null compared to a well-informed posterior distribution.

Background ILI

The prior-to-posterior plots of the three parameters describing the background, non-flu-infected, ILI GP visits are reported in Figure 6.11.

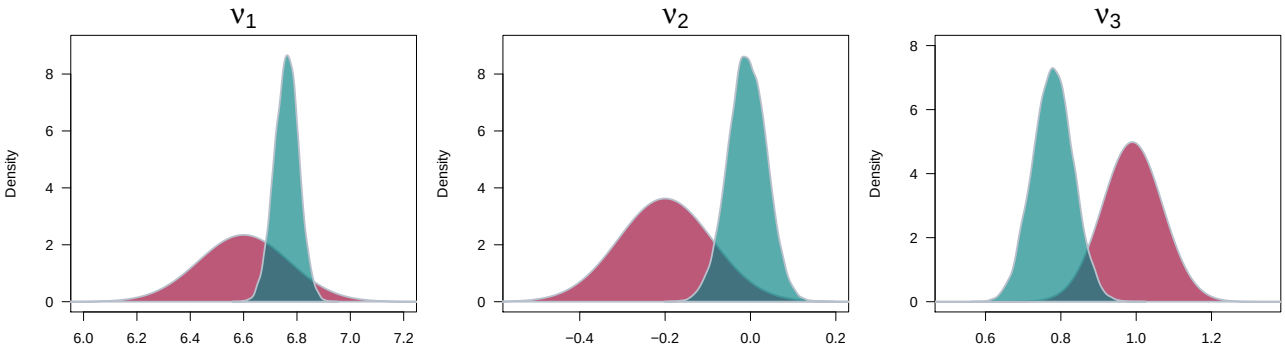


Figure 6.11: Prior (red) and posterior (green) distributions of the three parameters of the HHH model.

The resulting average of the background-ILI GP visit process is drawn in Figure 6.12. Comparing the posterior distribution with the prior, the incidence is slightly delayed but has similar magnitude.

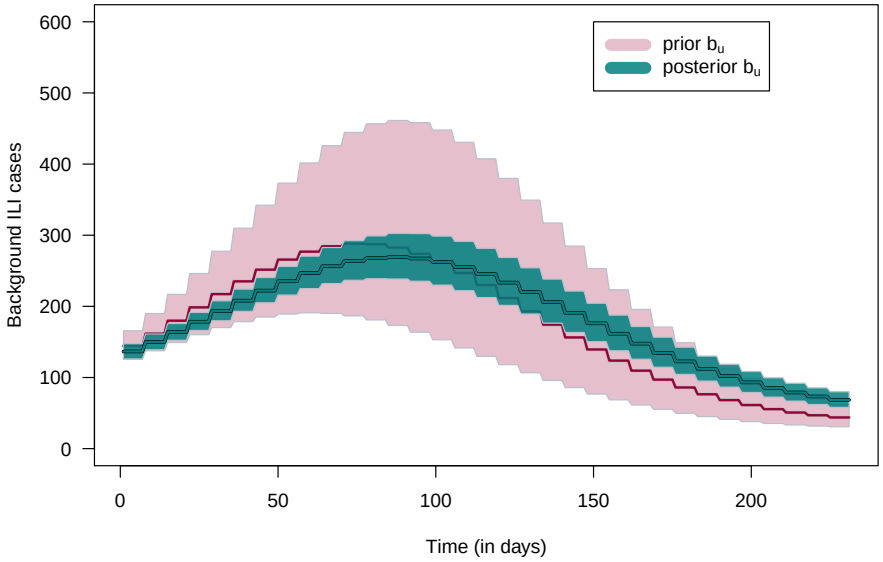


Figure 6.12: Median (solid line) and 95% CrIs (shaded area) of the prior (red) and posterior (green) for the mean of the rate of the daily number of non-influenza ILI GP consultations.

Day-of-the-week effect

The prior-to-posterior plots of the day-of-the-week parameters are reported in Figure 6.13. Monday is the day with highest consultation rate (4 times bigger than the weekly average), while Sunday has the smallest rate (20 times smaller than the average).

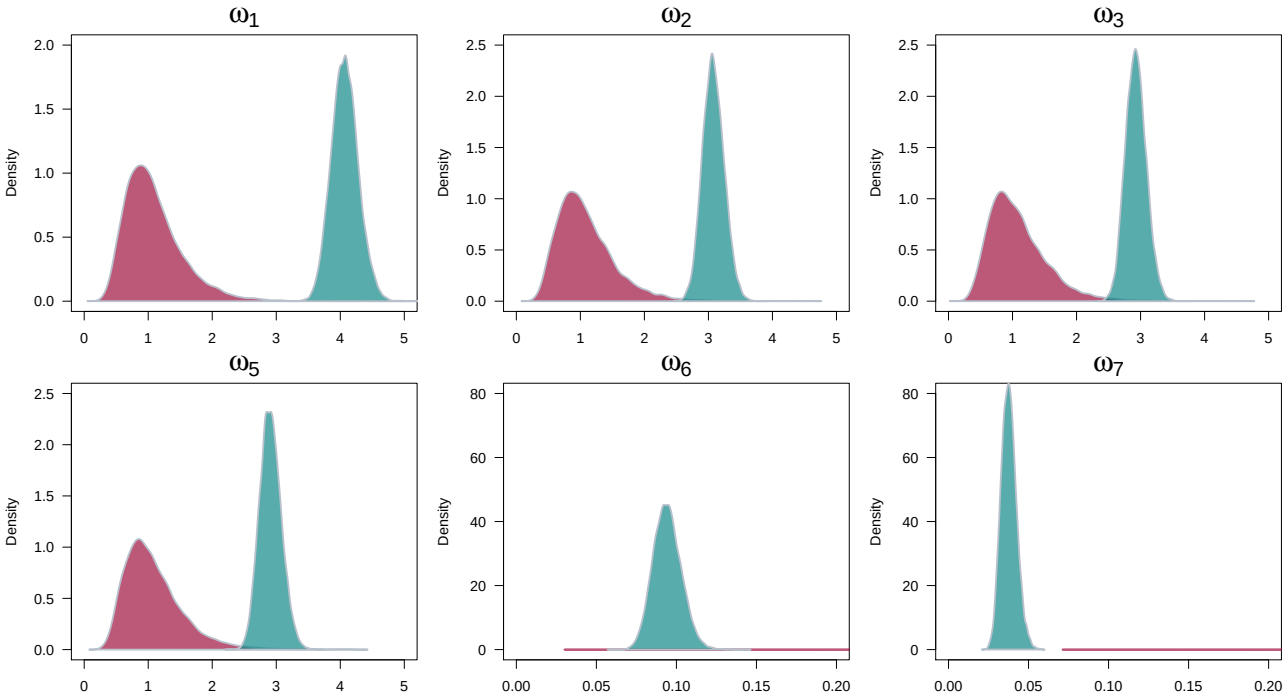


Figure 6.13: Prior (red) and posterior (green) distributions of the day-of-the-week parameters.

The posterior distributions of these parameters, together with the distribution plotted in Figure 6.10, allow to compute the day-specific risk of GP consultation for influenza cases, plotted in Figure 6.14.

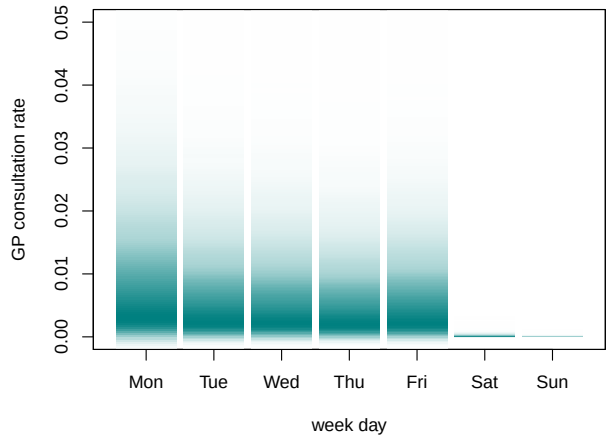


Figure 6.14: Density strips of the day-specific probability of GP consultation for influenza cases.

Parameter	Name	Median (95% CrI)
Transmission rate	β	0.6189 (0.5284-0.747)
Initial immunity	π	0.3744 (0.2836-0.4709)
Initial exposed/infectious	ι	$1.519{\cdot}10^{-4}$ ($7.1756{\cdot}10^{-5}$-$2.668{\cdot}10^{-4}$)
Average latent period	$d_L = 2/\sigma$	3.5437 (1.3458-6.2044)
Factor for holidays	κ	0.4273 (0.2523-0.6311)
Mean of Case flu-symptoms risk	ι^{F}	0.0025 (0.0018-0.0038)
Precision of Case flu-symptoms risk	ε^{F}	544.5252 (350.1071-804.3077)
Case hospitalization risk	${}^{0}\theta^{\mathrm{H}}$	0.0032 (0.0022-0.0049)
Hospital ICU admission risk	${}^{\mathrm{H}}\theta^{\mathrm{IC}}$	0.0667 (0.0574-0.078)
HHH model parameter 1	ν_1	6.7618 (6.6705-6.8495)
HHH model parameter 2 (sin)	ν_2	-0.008 (-0.0987-0.0792)
HHH model parameter 3 (cos)	ν_3	0.7789 (0.6742-0.8862)
Monday effect	ω_1	4.067 (3.6789-4.5025)
Tuesday effect	ω_2	3.0727 (2.7635-3.4208)
Wednesday effect	ω_3	2.9227 (2.6261-3.2603)
Friday effect	ω_5	2.9043 (2.5997-3.2598)
Saturday effect	ω_6	0.0939 (0.0781-0.1122)
Sunday effect	ω_7	0.0375 (0.0291-0.0479)
Shape parameter of hospitalizations detection	ε^{H}	25.2376 (13.863-44.1221)

Table 6.2: Medians and 95% CrIs of the posterior distribution of all the parameters.

Reporting

The posterior distribution of the shape parameter of the hospitalization detection is reported in Figure 6.15. The consequent posterior predictive distribution of the detection parameter $\zeta^{\mathrm{H}}_{1:T}$ is

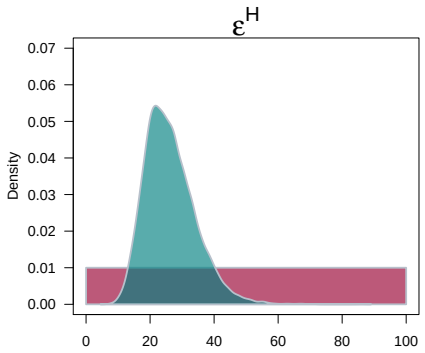


Figure 6.15: Prior (red) and posterior (green) distributions of the shape parameter of the detection of influenza hospitalizations.

reported in Figure 6.16, where the high variability of this parameter can be seen.

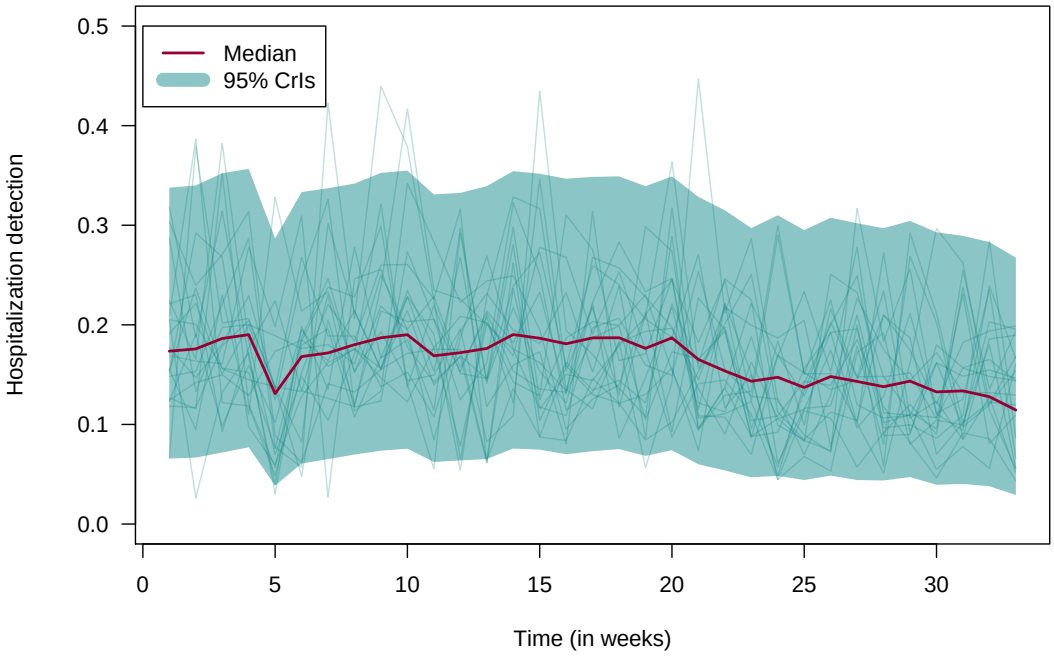


Figure 6.16: Median (red) and 95% CrIs (green) for the probability of detecting an hospitalised influenza case. 20 trajectories are also simulated and plotted as thin green lines.

6.3.3 Performance assessment

The goodness of fit of the model is evaluated graphically. Figure 6.17 plots the predictive posterior distribution of the observed GP consultations for ILI versus the data both on the natural and on the logarithmic scale.

The model describes well the epidemic during its beginning and its end, however the peak of the data is not well reproduced. The variability of the predictions is particularly high in the middle of the epidemic, reflecting the high variability assumed in the severity model and the uncertainty that characterizes the prior distributions of many parameters. The day of the week effect, instead, reflects well the shifts of the observations during weekends, compared to weekdays.

Virology data are much better modelled (Figure 6.18): the peak of the epidemic is reproduced well and the median predicted number of positives always lies close to the observed data.

Hospital and ICU data are to be considered together. While the model proposed above accounts for extra sources of variability for the hospital data (with the Beta prior distribution), the ICU data are assumed affected by Poisson noise only. Dropping this assumption will compromise the possibility of direct sampling of the hidden states in Algorithm 7. However, over-dispersion parameters have often been found useful, if not essential, to the modelling of infectious-disease data (Bretó, 2018). As a result, the ICU data are not reproduced well by the model (see Figure 6.20).

By contrast, the predictive distributions of the hospital data are more variable and the CrIs always include the observed value. Moreover, the median trajectory shows a second peak in correspondence with the observed second peak in the data.

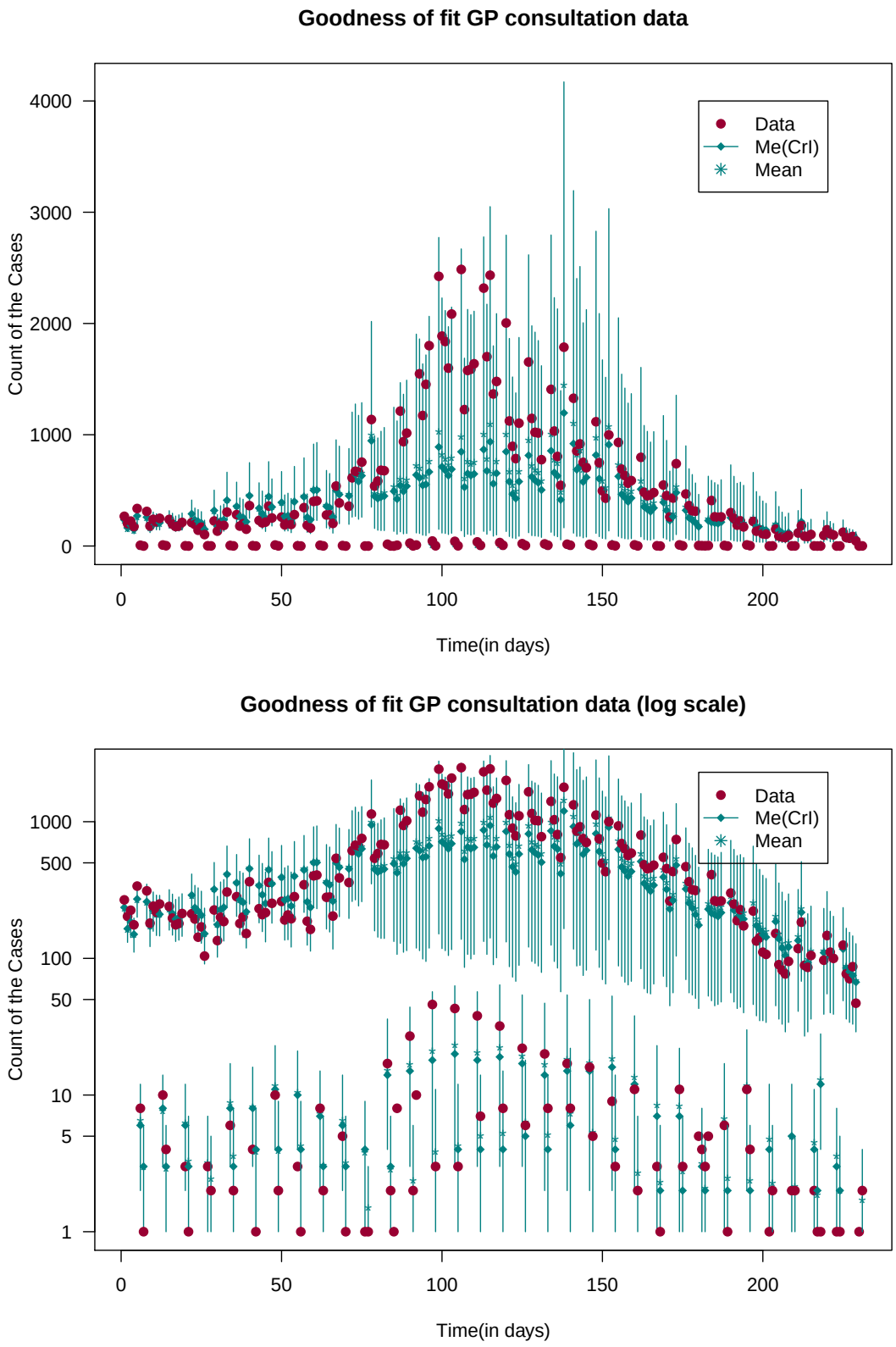


Figure 6.17: Median and 95% CrIs (green) for the posterior predicted distribution of the number of daily GP consultations on the natural (top panel) and logarithmic (bottom panel) scale. Red points are data.

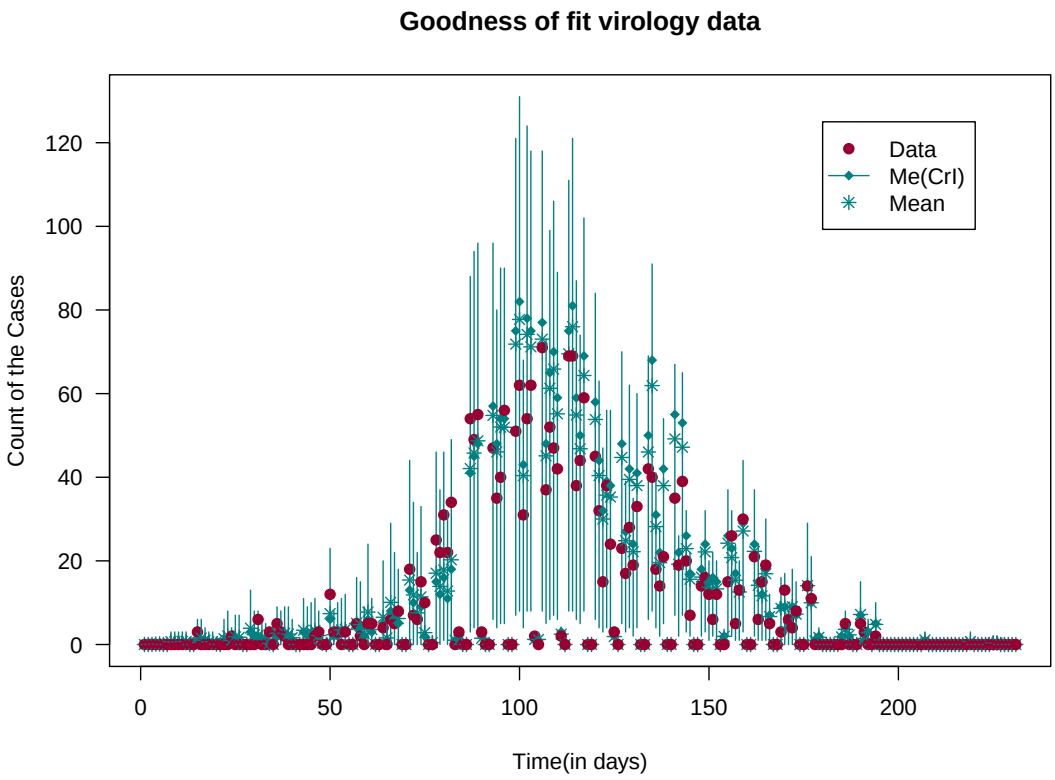


Figure 6.18: Median and 95% CrIs (green) for the posterior predicted distribution of the number of daily influenza-positive tests. Red points are data.

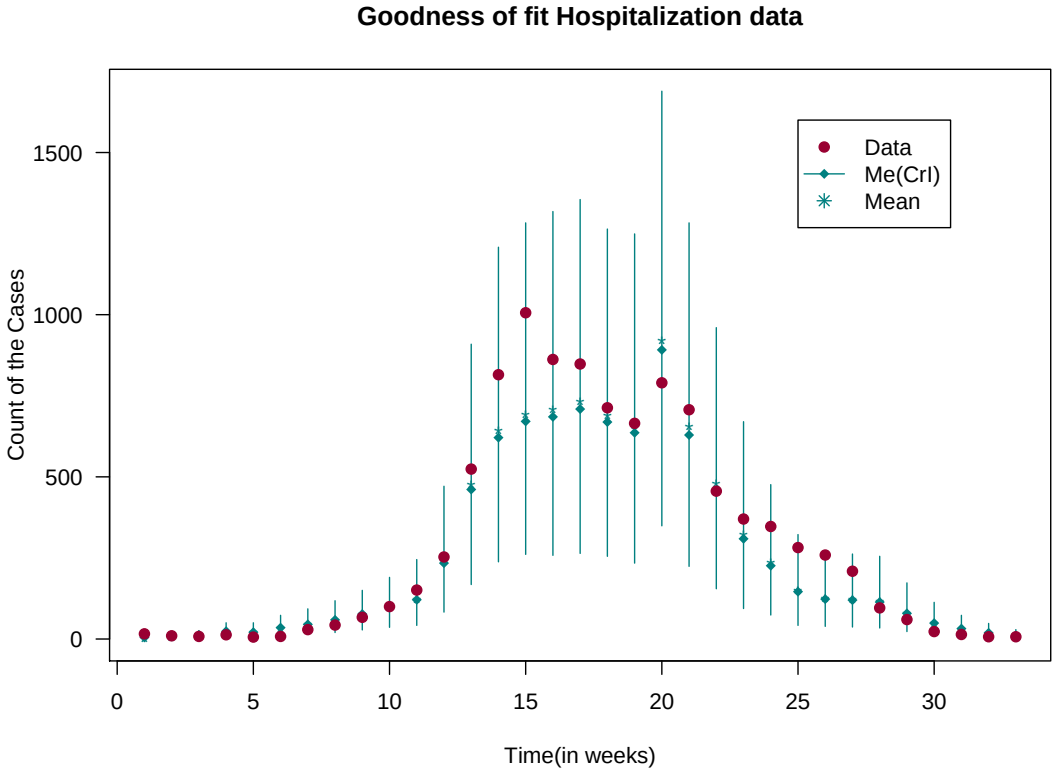


Figure 6.19: Median and 95% CrIs (green) for the posterior predicted distribution of the number of weekly hospital admissions. Red points are data.

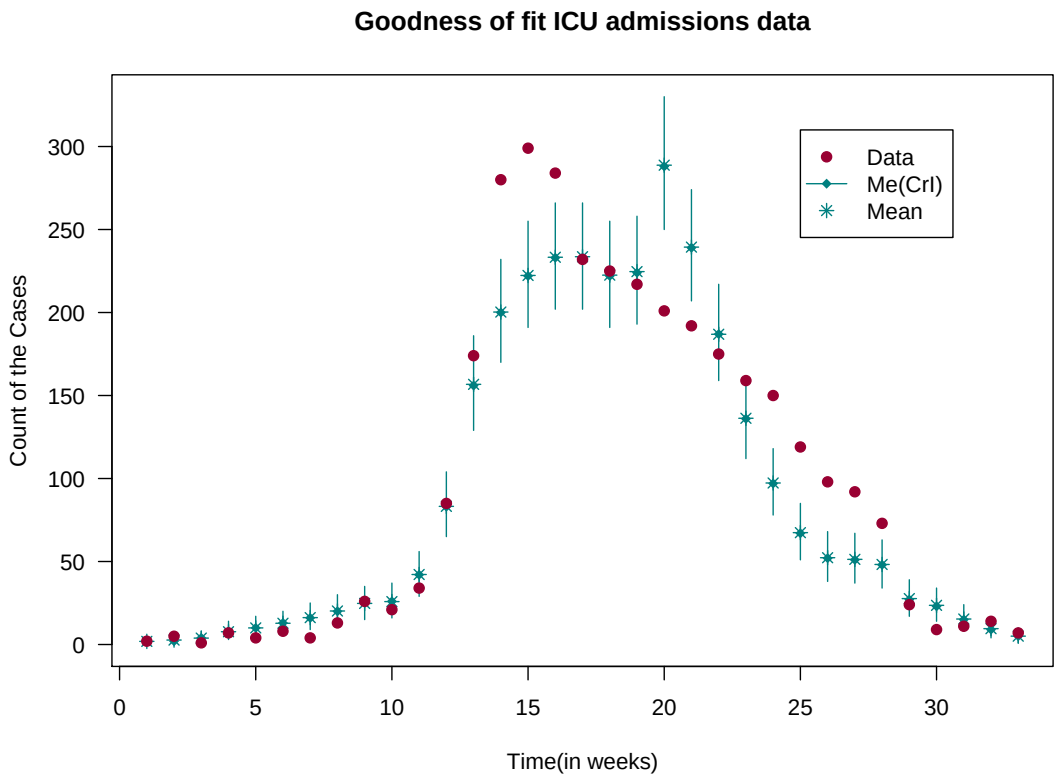


Figure 6.20: Median and 95% CrIs (green) for the posterior predicted distribution of the number of weekly ICU admissions. Red points are data.

6.4 Discussion

This chapter reports the analysis of the latest epidemic of seasonal influenza and the inference of its transmission and severity. Multiple data streams are used, each monitoring cases at different levels of severity. The model accounts for two important features of the data streams: counts can be dependent across different data streams and counts can include non-influenza cases.

The analysis provides sensible results. Firstly, the model fits well to most of the data analysed, with posterior predictive credible intervals often covering the observed data. When mismatch between predictions and data exists, this might be due either to the rigid temporal structure of the underlying transmission model or to conflicting information across the data sources or to both. Secondly, many transmission severity and detection parameters are well informed by the data. Despite epidemic models being often affected by identifiability issues, in this analysis a high number of parameters is investigated and only few of them are notably affected by unidentifiability: most of the posterior distributions, instead, are well identified and distinguishable from the prior.

To further improve fitting, the model could be extended in a number of ways. The first would be to consider a more flexible temporal structure for transmission, for example by assuming a Gaussian process for the rate β similarly to Shubin et al. (2016). Another important point would be to better model heterogeneity in contact and transmission by adopting an age-specific and strain-specific model. Under the inference perspective instead, the analysis could benefit from the use of more adequate hardware and computational-efficiency could be enhanced by parallelizing the algorithms (e.g. in the MC approximation for the hospital/ICU data).

To the knowledge of the author however, such a complex amount of data has never been synthesized by coupling a fully stochastic severity and detection model with a deterministic transmission model.

More often a fully stochastic state space model (SSM) is defined, with unnecessary waste of computational time for a stochastic epidemic model which might be easily approximated by a deterministic epidemic model (Marmara, Cook, and Kleczkowski, 2014; Ong et al., 2010; Shubin et al., 2016). Alternatively, people have used deterministic transmission and severity models attributing all the data noise to over-dispersion due to detection (Birrell et al., 2011), though this is not likely to be the only source of variability.

Hence, the model proposed here is very innovative because it merges multiple data, still accounting for many layers of variability and for plausible dependencies.

Moreover, in the future, this work can be approached under a value-of-information perspective (Jackson et al., 2017) and questions such as: *which data carries which information?* or *what amount of variability is attributable to the hidden state* x *?* can be addressed. Answers to these questions are key to future data collection planning.

Chapter 7

Conclusions

PIPPIN: I didn't think it would end this way.
GANDALF: End? No, the journey doesn't end here.

J. R. R. Tolkien, *The Lord of the Rings*

About 100 years ago the emergence of a new variant of the influenza virus, called influenza H1N1, led to the Spanish Flu pandemic reducing the world's population by at least 50 million and making life expectancy drop by, on average, 12 years (*The Influenza Epidemic of 1918*). Since then, pandemic and seasonal epidemics have threatened global health, so much so that in the UK, the National Risk Register of Civil Emergencies lists an outbreak of pandemic influenza as the greatest non-terrorism risk faced by the country (Cabinet Office, 2015).

To detect pandemics and, above all, to monitor seasonal influenza epidemics, many countries have developed surveillance schemes of influenza activity (WHO, 2013).

Exploiting the available data to quantify severity and transmission is, therefore, key for health policies aimed at allocating resources. This thesis focuses on the estimation of severity and transmission using surveillance data.

From a statistical point of view, this topic is particularly interesting: the data collected are only partial snapshots of a very complex unobserved phenomenon; this phenomenon evolves over time as a stochastic process; the data often come from multiple sources, possibly affected by dependencies; and further information might be available in the form of expert prior knowledge. Hence, the inference of influenza severity and transmission is here addressed within a Bayesian evidence synthesis framework.

7.1 Main Findings

This work is placed at the interface of public health, epidemic/dynamic models and Bayesian methods and computation. Advances are made in the three fields: from a public health point of view, relevant estimates of transmission and severity are provided to Public Health England (PHE); epidemic models with new features are formulated and applied to data never or only partially exploited before; and algorithms are extended and applied to new contexts.

7.1.1 Severity

Chapter 2 provides a detailed description of some of the existing methods for the estimation of the hospitalization case fatality risk (CFR). Rather than putting together a systematic and comprehensive review, the focus of the chapter is on showing the applicability of existing methods to the data collected in England.

A new model is proposed that allows the estimation of the time-varying case fatality risk (tvFR) from time-to-event data. However, like many of the other estimators reviewed, sensible estimates can be obtained only when large/complete time-to-event datasets are available. Time-to-event data on influenza severe events collected in England are rarely large and they often contain missing observations. By contrast, count data on the number of admissions and deaths alone are available but no estimator, apart from the one suggested by the World Health Organization (WHO) which is biased and over-precise, can infer the CFR from these data.

Aiming to exploit available count data, the process of severity is framed within the state space model (SSM) scheme in Section 4.4 and a sequential Monte Carlo (SMC) algorithm to approximate the likelihood is proposed. The SMC algorithm used works on a very constrained space so an extension is proposed to allow samples from the boundaries of the constrained space to be included. The proposed algorithm can be used to estimate the CFR in the early phases of an epidemic, and it is shown to provide better estimates than the ones derived with the WHO estimator.

7.1.2 Transmission from one source

Chapter 3 provides a review of epidemic models, their formulation and classification. The analysis of three seasons of influenza in England is presented as a case study. Transmission is inferred from data on the weekly counts of Intensive Care Unit (ICU) admissions.

Here the major obstacles of epidemic modelling arise: lack of parameter identifiability strongly compromises the inference and one data source alone might not be sufficient to estimate the whole process. Nevertheless, quantification of useful characteristics, such as the effective reproduction number and the probability of ICU admission given infection, are provided. Likewise, a prediction exercise is carried out assuming that counts become available in real time. Long term predictions can be drawn from soon after the epidemic peak.

7.1.3 Transmission from multiple sources

While exploring transmission from multiple sources, particular care should be taken in integrating information from possibly dependent data.

If a stochastic transmission model is used and combined with a stochastic severity model (as in Section 4.5), classical SMC techniques can be used to approximate the likelihood of the observed data. Conditional independence and Markovianity are explicitly modelled in the assumed SSM and therefore a classical bootstrap particle filter (BPF) can be used. The innovation of the model of Section 4.5 is that it approaches the severity process as an immigration-death process, accounting for the delay between events of different levels of severity, which distinguish this work from other applications of SMC methods to stochastic epidemic models (Marmara, Cook, and Kleczkowski, 2014).

If, as is realistic for the case of seasonal influenza, a deterministic epidemic model can be used, the stochasticity and the dependence between multiple data sources are embedded in the severity and observational process. Chapter 5 focusses on formulating a deterministic-transmission and stochastic-severity model that allows for dependent data. The relevance of the dependence is further investigated with a simulation study that shows that, for data that are assumed highly dependent, ignoring the dependence between data sources might cause significant over-precision in the estimates.

Lastly in Chapter 6 a model that jointly analyses four different data sources collected in England is formulated. This model combines a deterministic epidemic model with a stochastic severity and observational model. The key achievement of this chapter is the construction of a unique model that gives a comprehensive picture of influenza in England, simultaneously modelling severe cases (ICU and hospital admissions), mild cases (influenza-related General Practitioner (GP) consultations) and background, non influenza-related, GP consultations for influenza-like illness (ILI). Estimates of relevant quantities are obtained, including: the case-GP consultation risk, the case-hospitalization risk, the hospital to ICU admission risk, the basic and effective reproduction numbers. The model shows a fair fit to the data analysed, as well as some glimpse of possible conflict of information across data sources.

7.2 Future work

This thesis, while providing some answers to the questions posed at the beginning of the PhD, opens new research directions to be explored in the future.

Some of the future work should be aimed at extending and improving the models proposed in this thesis, as well as focussing on the wider methodological aspects that have arisen.

7.2.1 Model extensions

Multiple strains

In the introduction, the influenza virus was characterised as a highly-mutating virus; nevertheless none of the analyses presented in this thesis have included information on the type or sub-type of influenza. The counts of influenza cases used in Chapters 3 and 6 might also be available at the type or sub-type level but no clear formulation of how the different types of virus interact is available yet. Similarly, serology data are often collected for all the circulating viruses but whether a previous infection from a virus might grant immunity to a different type/sub-type is still to be explored (Wikramaratna et al., 2015).

Structured epidemics

As explained in Chapter 3, structures in the population might lead to inhomogeneous transmission dynamics. These structures might be due to spatial aggregation, age grouping, etc. (Birrell et al., 2016). Moreover, individuals belonging to different groups might have different propensity to acquire infection or to develop severe symptoms (Presanis et al., 2014). This aspect has not been addressed by the models presented in this thesis and constitutes a natural evolution of this work.

Time varying rates and probability

The only temporal structure present in the models used here is the difference in transmission during school opening and school closure and the day of the week effect. The transmission rate, as well as many other parameters, is likely to change over time following unknown trends and oscillation. Parameters have not been allowed to vary over time yet; this would be an interesting extension to make models more realistic (Shubin et al., 2016).

7.2.2 Model comparison and model choice

Both stochastic and deterministic models have been proposed. Combinations of the two aspects (e.g. stochastic transmission and stochastic severity, deterministic transmission and stochastic severity) are abundant and these models can achieve a high level of granularity and stochasticity. However, there is not a unique criteria to choose between one model and another: one model might result in a very good fit to the data but provide unreliable forecasts (Shubin et al., 2016); another could provide precise and unbiased forecasts at the expense of long computational time, which would prevent a timely use of the model; or again, a candidate model can fit the data very well but lead to non-interpretable parameters, preventing the comparison of the analysed epidemic with historical outbreaks (Brooks et al., 2015; Johnson et al., 2018).

A review of the statistical methodology for model evaluation/choice that can be applied to epidemics is presented in Gibson, Streftaris, and Thong (2018). Despite providing a considerable set of indexes and criteria, this paper does not deal with model evaluation based on predictive ability (Czado, Gneiting, and Held, 2009) and does not provide criteria to discriminate between models when the goal of the analysis is explicit (e.g. when there is a prediction goal compared to when the goal is retrospective fit of data).

Moreover, to the author's knowledge, these methods have never been applied in real time: as the data become available, a model chosen in an early stage of the epidemic might be replaced by a different, possibly more complex, model. Computationally-efficient models might be combined with concurrent model-choice criteria in order to achieve sequential adaptation of the whole analysis.

7.2.3 Value of information

A joint model for transmission and severity exploiting multiple data sources is proposed in Chapter 6. An interesting aspect that has not been addressed yet is the amount of information that is carried by each dataset and whether the introduction of a specific data source improves some aspects (and which aspects) of the inference. This requires a value-of-information perspective of the analysis (Jackson et al., 2017), which might suggest focussing resources on the collection of informative datasets and discontinuing less valuable collection schemes.

7.2.4 Modern computation methods in infectious disease inference

Advances in the field of computational methods, and more specifically Monte Carlo (MC) algorithms, are highly related to the possibility of performing faster and better inference. The inferential methods used in this thesis are classical Bayesian methods exploiting standard algorithms (e.g. Metropolis Hastings (MH)) or techniques that have already been around for

more than two decades (e.g. SMC algorithms such as the BPF). New MC methods should be experimented in the context of infectious disease analysis. Methods such as Hamiltonian Monte Carlo (HMC) (Neal, 2011) or algorithms based on piecewise deterministic Markov processes (Bierkens et al., 2018) have the promising potential of exploring a parameter space according to the geometry of the system (Betancourt, 2017) and this is particularly appealing for dynamical systems.

The application of these algorithms to the context of infectious disease seems an exciting follow-up to the work presented in this thesis.

Bibliography

Abbey, H (1952). "An examination of the Reed-Frost theory of epidemics." In: *Hum. Biol.* 24.3, pp. 201–33. ISSN: 0018-7143.

Ades, A. E., N. J. Welton, D. M. Caldwell, M. P. Price, A. Goubar, and G. Lu (2008). "Multiparameter evidence synthesis in epidemiology and medical decision-making". In: *J. Heal. Serv. Res. Policy* 13.SUPPL. 3, pp. 12–22. ISSN: 13558196. DOI: 10.1258/jhsrp.2008.008020. arXiv: 1108.1196.

Ajelli, M., B. Gonçalves, D. Balcan, V. Colizza, H. Hu, J. J. Ramasco, S. Merler, and A. Vespignani (2010). "Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models". In: *BMC Infect. Dis.* 10. ISSN: 14712334. DOI: 10.1186/1471-2334-10-190.

Akaike, H. (1987). "Factor Analysis and AIC". In: *Sel. Pap. Hirotugu Akaike*. Vol. 5. 2. Springer, New York, pp. 371–386. ISBN: 9788578110796. DOI: 10.1007/978-1-4612-1694-0_29. arXiv: NIHMS150003.

Andersen, P. K. and N. Keiding (2002). "Multi-state models for event history analysis." In: *Stat. Methods Med. Res.* 11.2, pp. 91–115. ISSN: 09622802. DOI: 10.1191/0962280202SM276ra.

Anderson, J. L. (1996). *A method for producing and evaluating probabilistic forecasts from ensemble model integrations*. DOI: 10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2.

Andersson, H. and T. Britton (2012). *Stochastic epidemic models and their statistical analysis*. Vol. 151. Springer Science & Business Media.

Andrieu, C., A. Doucet, and R. Holenstein (2010). "Particle Markov chain Monte Carlo methods". In: *J. R. Stat. Soc. B* 72.3, pp. 269–342. ISSN: 13697412. DOI: 10.1111/j.1467-9868.2009.00736.x.

Andrieu, C. and G. O. Roberts (2009). "The pseudo-marginal approach for efficient Monte Carlo computations". In: *Ann. Stat.* 37.2, pp. 697–725. ISSN: 00905364. DOI: 10.1214/07-AOS574. arXiv: 0903.5480.

Anon. (1978). "Influenza in a Boarding School". In: *The British Medical Journal* 1, p. 587.

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002). "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". In: *IEEE Trans. Signal Process.* 50.2, pp. 174–188. ISSN: 1053587X. DOI: 10.1109/78.978374. arXiv: arXiv:1011.1669v3.

Baguelin, M., A. J. V. Hoek, M. Jit, S. Flasche, P. J. White, and W. J. Edmunds (2010). "Vaccination against pandemic influenza A/H1N1v in England: A real-time economic evaluation". In: *Vaccine* 28.12, pp. 2370–2384. ISSN: 0264410X. DOI: 10.1016/j.vaccine.2010.01.002.

Baguelin, M., S. Flasche, A. Camacho, N. Demiris, E. Miller, and W. J. Edmunds (2013). "Assessing Optimal Target Populations for Influenza Vaccination Programmes: An Evidence

Synthesis and Modelling Study". In: *PLoS Med.* 10.10, pp. 1–19. ISSN: 15491277. DOI: 10.1371/journal.pmed.1001527.

Bailey, N. T. (1975). *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.

Baker, M. G., H Kelly, and N Wilson (2009). "Pandemic H1N1 Influenza Lessons from the Sourthern Hemisphere". In: *Eurosurveillance* 14.42, pp. 6–10. ISSN: 1560-7917.

Baker, M. G. et al. (2009). "Pandemic influenza A(H1N1)v in New Zealand: the experience from April to August 2009." In: *Eurosurveillance* 14.34, pp. 1–6. ISSN: 15607917.

Ball, F., T. Britton, T. House, V. Isham, D. Mollison, L. Pellis, and G. Scalia Tomba (2014). "Seven challenges for metapopulation models of epidemics, including households models". In: *Epidemics* 10, pp. 63–67. ISSN: 18780067. DOI: 10.1016/j.epidem.2014.08.001.

Beaumont, M. A. (2003). "Estimation of population growth or decline in genetically monitored populations". In: *Genetics* 164.3, pp. 1139–1160. ISSN: 00166731.

Becker, N (1979). "The uses of epidemic models". In: *Biometrics* 35.1, pp. 295–305. ISSN: 0006-341X. DOI: 10.2307/2529951.

Becker, N. (1989). *Analysis of Infectious Disease Data*. Vol. 33. CRC Press.

Becker, N. (1993). "Martingale methods for the analysis of epidemic data". In: *Stat. Methods Med. Res.* 2.1, pp. 93–112. ISSN: 0962-2802. DOI: 10.1177/096228029300200106.

Betancourt, M. (2017). "A conceptual introduction to Hamiltonian Monte Carlo". In: *arXiv preprint arXiv:1701.02434*.

Beyersmann, J., A. Latouche, A. Buchholz, and M. Schumacher (2009). "Simulating competing risks data in survival analysis". In: *Stat. Med.* 28.6, pp. 956–971. ISSN: 02776715. DOI: 10.1002/sim.3516.

Bierkens, J., A. Bouchard-Côté, A. Doucet, A. B. Duncan, P. Fearnhead, T. Lienart, G. Roberts, and S. J. Vollmer (2018). "Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains". In: *Stat. Probab. Lett.* 136, pp. 148–154. ISSN: 01677152. DOI: 10.1016/j.spl.2018.02.021.

Biggerstaff, M, S Cauchemez, C Reed, M Gambhir, and L Finelli (2014). "Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature". In: *BMC Infect Dis* 14, p. 480. ISSN: 1471-2334. DOI: 10.1186/1471-2334-14-480.

Birrell, P. et al. (2011). "Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London". In: *Proc. Natl. Acad. Sci.* 108.45, pp. 18238–18243. ISSN: 0027-8424. DOI: 10.1073/pnas.1103002108.

Birrell, P. J., D. De Angelis, and A. M. Presanis (2018). "Evidence Synthesis for Stochastic Epidemic Models". In: *Stat. Sci.* 33.1, pp. 34–43. ISSN: 0883-4237. DOI: 10.1214/17-STS631. arXiv: 1706.02624.

Birrell, P. J., T. R. Chadborn, O. N. Gill, V. C. Delpech, and D. De Angelis (2012). "Estimating Trends in Incidence, Time-to-Diagnosis and Undiagnosed Prevalence using a CD4-based Bayesian Back-calculation". In: *Stat. Commun. Infect. Dis.* 4.1. ISSN: 1948-4690. DOI: 10.1515/1948-4690.1055.

Birrell, P. J., X. S. Zhang, R. G. Pebody, N. J. Gay, and D. De Angelis (2016). "Reconstructing a spatially heterogeneous epidemic: Characterising the geographic spread of 2009

A/H1N1pdm infection in England". In: *Sci. Rep.* 6.June, pp. 1–10. ISSN: 20452322. DOI: 10.1038/srep29004.

Birrell, P. J., D. De Angelis, L. Wernisch, B. D. M. Tom, G. O. Roberts, and R. G. Pebody (2017). "Efficient real-time monitoring of an emerging influenza epidemic: how feasible?" In: *ArXiv*, pp. 1–36. arXiv: 1608.05292.

Bjornstad, O. N., B. F. Finkenstadt, and B. T. Grenfell (2002). "Dynamics of Measles Epidemics: Estimating Scaling of Transmission Rates Using a Time Series SIR Model". In: *Ecol. Monogr.* 72.2, p. 169. ISSN: 00129615. DOI: 10.2307/3100023.

Boddington, N. L., N. Q. Verlander, R. G. Pebody, and the UK severe influenza surveillance system Group (2017). "Developing a system to estimate the severity of influenza infection in England: findings from a hospital-based surveillance system between 2010/2011 and 2014/2015". In: *Epidemiol. Infect.* Pp. 1–10. ISSN: 0950-2688. DOI: 10.1017/S095026881700005X.

Bretó, C. et al. (2018). "Modeling and inference for infectious disease dynamics: a likelihood-based approach". In: *Statistical Science* 33.1, pp. 57–69.

Britton, T. (2010). "Stochastic epidemic models: A survey". In: *Math. Biosci.* 225.1, pp. 24–35. ISSN: 00255564. DOI: 10.1016/j.mbs.2010.01.006. arXiv: 0910.4443.

Brockwell, P. J. and R. A. Davis (2016). *Introduction to time series and forecasting*. Springer.

Brookmeyer, R., M. H. Gail, et al. (1994). *AIDS epidemiology: a quantitative approach*. Vol. 22. Oxford University Press on Demand.

Brooks, L. C., D. C. Farrow, S. Hyun, R. J. Tibshirani, and R. Rosenfeld (2015). "Flexible Modeling of Epidemics with an Empirical Bayes Framework". In: *PLoS Comput. Biol.* 11.8, pp. 1–18. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1004382.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

Brownlee, J. (1907). "Statistical Studies in Immunity: The Theory of an Epidemic". In: *Proc. R. Soc. Edinburgh* 26, pp. 484–521. ISSN: 0370-1646. DOI: 10.1017/S037016460002472X.

Cabinet Office (2015). *National Risk Register of Civil Emergencies. UK Cabinet Office*. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/419549/20150331_2015-NRR-WA_Final.pdf (visited on 02/20/2019).

Carrat, F., E. Vergu, N. M. Ferguson, M. Lemaitre, S. Cauchemez, S. Leach, and A. J. Valleron (2008). "Time lines of infection and disease in human influenza: A review of volunteer challenge studies". In: *Am. J. Epidemiol.* 167.7, pp. 775–785. ISSN: 00029262. DOI: 10.1093/aje/kwm375.

Carstensen, B. (2007). "Age–period–cohort models for the Lexis diagram". In: *Stat. Med.* 26.15, pp. 3018–3045. ISSN: 02776715. DOI: 10.1002/sim.2764.

Carstensen, B. (2016). *Statistical Analysis in the Lexis Diagram: Age-Period-Cohort models. Lecture notes for the Max Plack Insitute of Demographic Research, Rostock*. URL: https://bendixcarstensen.com/APC/MPIDR-2016/ (visited on 11/31/2018).

Carstensen, B., M. Plummer, E. Laara, and M. Hills (2018). *Epi: A Package for Statistical Analysis in Epidemiology*. R package version 2.32.

Cassini, A. et al. (2018). "Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to

2013". In: *Eurosurveillance* 23.16, pp. 1–20. ISSN: 1560-7917. DOI: 10.2807/1560-7917.ES.2018.23.16.17-00454.

Chang, I.-S. and C. A. Hsiung (2005). "Counting Process Methods in Survival Analysis". In: *Encyclopedia of Biostatistics* 4.

Charlett, A. (2018). "Personal communication on the level of immunity priori 2017/18 Influenza season". PHE.

Chiang, C. L. (1975). *An introduction to stochastic processes and their application*. Robert E. Krieger Pub. Com.

Chiang, C. L. (1980). *An introduction to Stochastic Processes and their applications*. Robert E. Krieger.

Chiang, C. L. (2007). "Stochastic processes". In: *Encyclopedia of Biostatistics*.

Chretien, J. P., D. George, J. Shaman, R. A. Chitale, and F. E. McKenzie (2014). "Influenza forecasting in human populations: A scoping review". In: *PLoS One* 9.4. ISSN: 19326203. DOI: 10.1371/journal.pone.0094130.

Churchill, G. A. (2005). "Hidden M arkov Models". In: *Encyclopedia of Biostatistics* 4.

Clayton, D. and E. Schifflers (1987a). "Models for temporal variation in cancer rates. I: Age–period and age–cohort models". In: *Stat. Med.* 6.4, pp. 449–467. ISSN: 02776715. DOI: 10.1002/sim.4780060405.

Clayton, D. and E. Schifflers (1987b). "Models for temporal variation in cancer rates. II: age-period-cohort models". In: *Stat. Med.* 6.4, pp. 469–481. ISSN: 02776715. DOI: 10.1002/sim.4780060406.

Commandeur, J., S. Koopman, and M Ooms (2011). "Statistical Software for State Space Methods". In: *J. Stat. Softw.* 41.1, pp. 1–18. ISSN: 1548-7660. DOI: 10.18637/jss.v041.i01.

Corbella, A., X. S. Zhang, P. J. Birrell, N. Boddington, R. G. Pebody, A. M. Presanis, and D. De Angelis (2018). "Exploiting routinely collected severe case data to monitor and predict influenza outbreaks". In: *BMC Public Health* 18.1, pp. 1–11. ISSN: 14712458. DOI: 10.1186/s12889-018-5671-7. arXiv: 1706.02527.

Cornuet, J.-M., J.-M. Marin, A. Mira, and C. P. Robert (2012). "Adaptive Multiple Importance Sampling". In: *Scand. J. Stat.* 39.4, pp. 798–812. ISSN: 03036898. DOI: 10.1111/j.1467-9469.2011.00756.x. arXiv: 0907.1254.

Czado, C., T. Gneiting, and L. Held (2009). "Predictive model assessment for count data". In: *Biometrics* 65.4, pp. 1254–1261. ISSN: 0006341X. DOI: 10.1111/j.1541-0420.2009.01191.x.

De Angelis, D. and A. M. Presanis (2018). "Analysing Multiple Epidemic Data Sources". In: *Handb. Infect. Dis. Data Anal.* Ed. by L. Held, N. Hens, P. O'Neill, and J. Wallinga. Chapman & Hall/CRC. arXiv: 1808.04312.

Del Moral, P., A. Doucet, and A. Jasra (2006). "Sequential Monte Carlo samplers". In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 68.3, pp. 411–436. ISSN: 13697412. DOI: 10.1111/j.1467-9868.2006.00553.x. arXiv: 0212648v1 [cond-mat].

Diekmann, O. and J. A. P. Heesterbeek (2000). *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Vol. 5. John Wiley & Sons.

Dietz, K (1988). "The first epidemic model: an historical note on P. D. En'ko". In: *Aust. J. Stat.* 30.A, pp. 56–65.

Donaldson, L. J., P. D. Rutter, B. M. Ellis, F. E. C. Greaves, O. T. Mytton, R. G. Pebody, and I. E. Yardley (2009). "Mortality from pandemic A/H1N1 2009 influenza in England: public health surveillance study." In: *BMJ* 339, b5213. ISSN: 0959-8138. DOI: 10.1136/bmj.b5213.

Donnelly, C. A. et al. (2003). "Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong". In: *Lancet* 361.9371, pp. 1761–1766. ISSN: 01406736. DOI: 10.1016/S0140-6736(03)13410-1.

Dorigatti, I., S. Cauchemez, and N. M. Ferguson (2013). "Increased transmissibility explains the third wave of infection by the 2009 H1N1 pandemic virus in England." In: *Proc. Natl. Acad. Sci. U. S. A.* 110.33, pp. 13422–7. ISSN: 1091-6490. DOI: 10.1073/pnas.1303117110.

Dorigatti, I., S. Cauchemez, A. Pugliese, and N. M. Ferguson (2012). "A new approach to characterising infectious disease transmission dynamics from sentinel surveillance: Application to the Italian 2009-2010 A/H1N1 influenza pandemic". In: *Epidemics* 4.1, pp. 9–21. ISSN: 17554365. DOI: 10.1016/j.epidem.2011.11.001.

Duan, W., Z. Fan, P. Zhang, G. Guo, and X. Qiu (2015). "Mathematical and computational approaches to epidemic modeling: a comprehensive review". In: *Front. Comput. Sci.* 9.5, pp. 806–826. ISSN: 20952236. DOI: 10.1007/s11704-014-3369-2.

Dukic, V., H. F. Lopes, and N. G. Polson (2012). "Tracking Epidemics With Google Flu Trends Data and a State-Space SEIR Model". In: *J. Am. Stat. Assoc.* 107.500, pp. 1410–1426. ISSN: 0162-1459. DOI: 10.1080/01621459.2012.713876.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Vol. 38. OUP Oxford.

Dureau, J., S. Ballesteros, and T. Bogich (2014). "SSM: Inference for time series analysis with State Space Models". In: *arXiv*, p. 1307.5626. arXiv: 1307.5626.

Dushoff, J., J. B. Plotkin, S. A. Levin, and D. J. D. Earn (2004). "Dynamical resonance can account for seasonality of influenza epidemics". In: *Proc. Natl. Acad. Sci.* 101.48, pp. 16915–16916. ISSN: 0027-8424. DOI: 10.1073/pnas.0407293101.

ECDC, E. (2018). *Factsheet about seasonal influenza*. URL: https://ecdc.europa.eu/en/seasonal-influenza/facts/factsheet (visited on 10/19/2018).

Eddelbuettel, D. and R. François (2018). "Rcpp syntactic sugar". In: *Rcpp Vignette*, pp. 1–12. ISSN: 09596526.

Farewell, V. T. (1982). "The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors". In: *Biometrics* 38.4, p. 1041. ISSN: 0006341X. DOI: 10.2307/2529885.

Farr, W (1840). "Progress of epidemics". In: *Second report of the Registrar General of England and Wales*, pp. 16–20.

Ferguson, N. M., D. A. T. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, and D. S. Burke (2006). "Strategies for mitigating an influenza pandemic". In: *Nature* 442.7101, pp. 448–452. ISSN: 0028-0836. DOI: 10.1038/nature04795.

Finkenstädt, B. F. and B. T. Grenfell (2000). "Time series modelling of childhood diseases: a dynamical systems approach". In: *J. R. Stat. Soc. Ser. C (Applied Stat.* 49.2, pp. 187–205. ISSN: 00359254. DOI: 10.1111/1467-9876.00187.

Garske, T., J. Legrand, C. A. Donnelly, H. Ward, S. Cauchemez, C. Fraser, N. M. Ferguson, and A. C. Ghani (2009). "Assessing the severity of the novel influenza A/H1N1 pandemic". In: *BMJ* 339, b2840.

Gelman, A., H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Ghani, A. C. et al. (2005). "Methods for estimating the case fatality ratio for a novel, emerging infectious disease". In: *Am. J. Epidemiol.* 162.5, pp. 479–486. ISSN: 00029262. DOI: 10.1093/aje/kwi230.

Gibson, G. J., G. Streftaris, and D. Thong (2018). "Comparison and Assessment of Epidemic Models". In: *Stat. Sci.* 33.1, pp. 19–33. ISSN: 0883-4237. DOI: 10.1214/17-STS615.

Gilks, W. R. and C Berzuini (2001). "Following a moving target - Monte Carlo inference for dynamic Bayesian models". In: *J. R. Stat. Soc. Ser. B-Statistical Methodol.* 63, pp. 127–146. ISSN: 1369-7412. DOI: 10.1111/1467-9868.00280.

Gillespie, D. T. (2001). "Approximate accelerated stochastic simulation of chemically reacting systems". In: *J. Chem. Phys.* 115.4, pp. 1716–1733. ISSN: 00219606. DOI: 10.1063/1.1378322.

Gillespie, D. T. (1976). "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *J. Comput. Phys.* 22.4, pp. 403–434. ISSN: 10902716. DOI: 10.1016/0021-9991(76)90041-3.

Gordon, N., D. Salmond, and a.F.M. Smith (1993). "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". In: *IEE Proc. F Radar Signal Process.* 140.2, p. 107. ISSN: 0956375X. DOI: 10.1049/ip-f-2.1993.0015.

Greenwood, M. (1931). "On the Statistical Measure of Infectiousness". In: *J. Hyg. (Lond).* 31.3, pp. 336–351. ISSN: 00221724. DOI: 10.1017/S002217240001086X.

Greenwood, P. E. and L. F. Gordillo (2009). "Stochastic epidemic modeling". In: *Mathematical and statistical estimation approaches in epidemiology*. Springer, pp. 31–52.

Gustafson, P. (2010). "Bayesian Inference for Partially Identified Models". In: *Int. J. Biostat.* 6.2, p. 17. ISSN: 1557-4679. DOI: 10.2202/1557-4679.1206.

Hall, I. M., R Gani, H. E. Hughes, and S Leach (2007). "Real-time epidemic forecasting for pandemic influenza." In: *Epidemiol. Infect.* 135, pp. 372–85. ISSN: 0950-2688. DOI: 10.1017/S0950268806007084.

Harcourt, S. E., G. E. Smith, A. J. Elliot, R Pebody, A Charlett, S Ibbotson, M Regan, and J Hippisley-Cox (2012). "Use of a large general practice syndromic surveillance system to monitor the progress of the influenza A(H1N1) pandemic 2009 in the UK." In: *Epidemiol. Infect.* 140.1, pp. 100–5. ISSN: 1469-4409. DOI: 10.1017/S095026881100046X.

Hastings, W. K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1, p. 97. ISSN: 00063444. DOI: 10.2307/2334940.

Health Protection Agency (2011a). *UK Severe Influenza Surveillance System (USISS) Protocol for all NHS Acute Trusts 2011-12*. URL: http://webarchive.nationalarchives.gov.uk/20140714043422/http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317132396353 (visited on 12/03/2018).

Health Protection Agency (2011b). *UK Severe Influenza Surveillance System (USISS) Protocol for sentinel Acute NHS Trusts 2011-12*. URL: http://webarchive.nationalarchives.gov.uk/20140714043432/http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317132396215 (visited on 12/03/2018).

Health Protection Agency (2014). *Sources of UK Flu Data: Influenza Surveillance in the United Kingdom*. URL: https://webarchive.nationalarchives.gov.uk/20140713201509/http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/SeasonalInfluenza/EpidemiologicalData/30influsSourcesofUKfludata/ (visited on 12/03/2018).

Heesterbeek, H. et al. (2015). "Modeling infectious disease dynamics in the complex landscape of global health". In: *Science (80-.).* 347.6227, aaa4339–aaa4339. ISSN: 10959203. DOI: 10.1126/science.aaa4339.

Held, L., M. Höhle, and M. Hofmann (2005). "A statistical framework for the analysis of multivariate infectious disease surveillance counts". In: *Stat. Model. An Int. J.* 5.3, pp. 187–199. ISSN: 1471-082X. DOI: 10.1191/1471082X05st098oa.

Hoschler, K., C. Thompson, N. Andrews, M. Galiano, R. Pebody, J. Ellis, E. Stanford, M. Baguelin, E. Miller, and M. Zambon (2012). "Seroprevalence of influenza A(H1N1) pdm09 virus antibody, England, 2010 and 2011". In: *Emerg. Infect. Dis.* 18.11, pp. 1894–1897. ISSN: 10806040. DOI: 10.3201/eid1811.120720.

House, T. et al. (2011). "Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic." In: *Proc. Biol. Sci.* 278.1719, pp. 2753–2760. ISSN: 0962-8452. DOI: 10.1098/rspb.2010.2688.

Ionides, E. L., C Bretó, and A. A. King (2006). "Inference for nonlinear dynamical systems." In: *Proc. Natl. Acad. Sci. U. S. A.* 103.49, pp. 18438–43. ISSN: 0027-8424. DOI: 10.1073/pnas.0603181103.

Jackson, C. H. (2015). "flexsurv: A Platform for Parametric Survival Modelling in R". In: *Compr. R Arch. Netw.*

Jackson, C., A. Presanis, S. Conti, and D. De Angelis (2017). "Value of Information: Sensitivity Analysis and Research Design in Bayesian Evidence Synthesis". In: arXiv: 1703.08994.

Jewell, N. P., X. Lei, A. C. Ghani, C. A. Donnelly, G. M. Leung, L.-M. Ho, B. J. Cowling, and A. J. Hedley (2007). "Non-parametric estimation of the case fatality ratio with competing risks data: an application to Severe Acute Respiratory Syndrome (SARS)". In: *Stat. Med.* 26.9, pp. 1982–1998. ISSN: 02776715. DOI: 10.1002/sim.2691.

Johnson, L. R., R. B. Gramacy, J. Cohen, E. Mordecai, C. Murdock, J. Rohr, S. J. Ryan, A. M. Stewart-Ibarra, and D. Weikel (2018). "Phenomenological forecasting of disease incidence using heteroskedastic gaussian processes: A dengue case study". In: *Ann. Appl. Stat.* 12.1, pp. 27–66. ISSN: 19417330. DOI: 10.1214/17-AOAS1090. arXiv: 1702.00261.

Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *J. Basic Eng.* 82.1, p. 35. ISSN: 00219223. DOI: 10.1115/1.3662552.

Kaplan, E. L. and P. Meier (1958). "Nonparametric Estimation from Incomplete Observations". In: *J. Am. Stat. Assoc.* 53.282, p. 457. ISSN: 01621459. DOI: 10.2307/2281868.

Katzfuss, M., J. R. Stroud, and C. K. Wikle (2016). "Understanding the Ensemble Kalman Filter". In: *Am. Stat.* 70.4, pp. 350–357. ISSN: 0003-1305. DOI: 10.1080/00031305.2016.1141709.

Keeling, M. J. and K. T. Eames (2005). "Networks and epidemic models". In: *J. R. Soc. Interface* 2.4, pp. 295–307. ISSN: 17425662. DOI: 10.1098/rsif.2005.0051.

Keeling, M. J. and P. Rohani (2011). *Modeling infectious diseases in humans and animals*. Princeton University Press.

Kermack, W. O. and A. G. McKendrick (1927). "Contributions to the Mathematical Theory of Epidemics. I." In: *Proc. R. Soc. A Math. Phys. Eng. Sci.* 115, pp. 700–721. ISSN: 00928240. DOI: 10.1007/BF02464423.

Kermack, W. O. and A. G. McKendrick (1932). "Contributions to the Mathematical Theory of Epidemics. II. The Problem of Endemicity". In: *Proc. R. Soc. A Math. Phys. Eng. Sci.* 138.834, pp. 55–83. ISSN: 1364-5021. DOI: 10.1098/rspa.1932.0171.

Kermack, W. O. and A. G. McKendrick (1933). "Contributions to the Mathematical Theory of Epidemics. III. Further Studies of the Problem of Endemicity". In: *Proc. R. Soc. A Math. Phys. Eng. Sci.* 141.843, pp. 94–122. ISSN: 1364-5021. DOI: 10.1098/rspa.1933.0106.

Killingley, B. and J. Nguyen-Van-Tam (2013). "Routes of influenza transmission". In: *Influenza Other Respi. Viruses* 7.SUPPL.2, pp. 42–51. ISSN: 17502640. DOI: 10.1111/irv.12080.

King, A. A., D. Nguyen, and E. L. Ionides (2016). "Statistical Inference for Partially Observed Markov Processes via the R Package pomp". In: *J. Stat. Softw.* 69.12, pp. 1–43. ISSN: 15487660. DOI: 10.18637/jss.v069.i12. arXiv: arXiv:1509.00503v1.

Kingman, J. F. C. (1992). *Poisson processes*. Vol. 3. Clarendon Press.

Kypraios, T., P. Neal, and D. Prangle (2017). "A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation". In: *Math. Biosci.* 287, pp. 42–53. ISSN: 00255564. DOI: 10.1016/j.mbs.2016.07.001.

Lam, K. F., J. V. Deshpande, E. H. Y. Lau, U. V. Naik-Nimbalkar, P. S. F. Yip, and Y. Xu (2008). "A test for constant fatality rate of an emerging epidemic: With applications to severe acute respiratory syndrome in Hong Kong and Beijing". In: *Biometrics* 64.3, pp. 869–876. ISSN: 0006341X. DOI: 10.1111/j.1541-0420.2007.00935.x.

Lindsten, F. (2013). "Backward Simulation Methods for Monte Carlo Statistical Inference". In: *Found. Trends® Mach. Learn.* 6.1, pp. 1–143. ISSN: 1935-8237. DOI: 10.1561/2200000045.

Lipsitch, M. and C. Viboud (2009). "Influenza seasonality: Lifting the fog". In: *Proc. Natl. Acad. Sci.* 106.10, pp. 3645–3646. ISSN: 0027-8424. DOI: 10.1073/pnas.0900933106.

Lipsitch, M. et al. (2015). "Potential Biases in Estimating Absolute and Relative Case-Fatality Risks during Outbreaks." In: *PLoS Negl. Trop. Dis.* 9.7, e0003846. ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0003846.

Longini, I. M. (1986). "The generalized discrete-time epidemic model with immunity: a synthesis". In: *Math. Biosci.* 82.1, pp. 19–41. ISSN: 00255564. DOI: 10.1016/0025-5564(86)90003-9.

Lowen, A. C., S. Mubareka, J. Steel, and P. Palese (2007). "Influenza virus transmission is dependent on relative humidity and temperature". In: *PLoS Pathog.* 3.10, pp. 1470–1476. ISSN: 15537366. DOI: 10.1371/journal.ppat.0030151.

Macdonald, G. (1952). "The analysis of equilibrium in malaria." In: *Tropical diseases bulletin* 49.9, p. 813.

Magpantay, F. M. G., M. Domenech De Cellès, P. Rohani, and a. a. King (2015). "Pertussis immunity and epidemiology: mode and duration of vaccine-induced immunity". In: *Parasitology*, pp. 1–15. ISSN: 0031-1820. DOI: 10.1017/S0031182015000979. arXiv: 15334406.

Marmara, V, A Cook, and A Kleczkowski (2014). "Estimation of force of infection based on different epidemiological proxies : 2009 / 2010 Influenza epidemic in Malta". In: *Epidemics* 9, pp. 52–61. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2014.09.010.

Marshall, A. W. (1954). *The use of multi-stage sampling schemes in Monte Carlo computations*. Tech. rep. RAND CORP SANTA MONICA CALIF.

Matis, J. H., T. E. Wehrly, and T. R. Kiffe (2005). "Compartment Models". In: *Encyclopedia of Biostatistics*. American Cancer Society. ISBN: 9780470011812. DOI: 10.1002/0470011815.b2a07011. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470011815.b2a07011.

McKinley, T. J., J. V. Ross, R. Deardon, and A. R. Cook (2014). "Simulation-based Bayesian inference for epidemic models". In: *Comput. Stat. Data Anal.* 71, pp. 434–447. ISSN: 01679473. DOI: 10.1016/j.csda.2012.12.012.

Merl, D., L. R. Johnson, R. B. Gramacy, and M. Mangel (2009). "A statistical framework for the adaptive management of epidemiological interventions". In: *PLoS One* 4.6, pp. 1–9. ISSN: 19326203. DOI: 10.1371/journal.pone.0005807.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). "Equation of state calculations by fast computing machines". In: *J. Chem. Phys.* 21.6, pp. 1087–1092. ISSN: 00219606. DOI: 10.1063/1.1699114. arXiv: 5744249209.

Murray, C. J., A. D. Lopez, B. Chin, D. Feehan, and K. H. Hill (2006). "Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918-20 pandemic: a quantitative analysis". In: *Lancet* 368.9554, pp. 2211–2218. ISSN: 01406736. DOI: 10.1016/S0140-6736(06)69895-4.

NIAID, N. (2012). *Influenza: Get the (Antigenic) Drift*. URL: https://www.youtube.com/watch?v=ug-M1nIhfIA (visited on 10/19/2018).

National Archives and Records Administration. *The Influenza Epidemic of 1918*. Ed. by Archives.gov. URL: https://www.archives.gov/exhibits/influenza-epidemic/ (visited on 02/20/2019).

Neal, R. M. et al. (2011). "MCMC using Hamiltonian dynamics". In: *Handbook of markov chain monte carlo* 2.11, p. 2.

Neumann, G., T. Noda, and Y. Kawaoka (2009). "Emergence and pandemic potential of swine-origin H1N1 influenza virus". In: *Nature* 459.7249, pp. 931–939. ISSN: 00280836. DOI: 10.1038/nature08157.

Nishiura, H., D. Klinkenberg, M. Roberts, and J. A. P. Heesterbeek (2009). "Early epidemiological assessment of the virulence of emerging infectious diseases: A case study of an influenza pandemic". In: *PLoS One* 4.8. ISSN: 19326203. DOI: 10.1371/journal.pone.0006852.

O'Neill, P. D. (2010). "Introduction and snapshot review: Relating infectious disease transmission models to data". In: *Stat. Med.* 29.20, pp. 2069–2077. ISSN: 02776715. DOI: 10.1002/sim.3968.

O'Neill, P. D. and G. O. Roberts (1999). "Bayesian inference for partially observed stochastic epidemics". In: *J. R. Stat. Soc. Ser. A Stat. Soc.* 162.1, pp. 121–129. ISSN: 09641998. DOI: 10.1111/1467-985X.00125.

Office of National Statistics (2012-2015). *Population Estimates for UK, England and Wales, Scotland and Northern Ireland*. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandand (visited on 06/23/2016).

Office of National Statistics (2017). *Population Estimates for UK, England and Wales, Scotland and Northern Ireland, mid 2016*. URL: https://www.ons.gov.uk/file?uri=

/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland/mid2016/ukmidyearestimate xls (visited on 02/02/2018).

Ong, J. B. S., M. I. cheng Chen, A. R. Cook, H. C. Lee, V. J. Lee, R. T. P. Lin, P. A. Tambyah, and L. G. Goh (2010). "Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in singapore". In: *PLoS One* 5.4, pp. 1–11. ISSN: 19326203. DOI: 10.1371/journal.pone.0010036.

Owen, A. and Y. Zhou (1998). "Safe and Effective Importance Sampling". In: *J. Am. Stat. Assoc.* 95.December, pp. 135–143. ISSN: 0162-1459. DOI: 10.1080/01621459.2000.10473909.

Palese, P. (2004). "Influenza: Old and new threats". In: *Nat. Med.* 10.12S, S82–S87. ISSN: 1546170X. DOI: 10.1038/nm1141.

Patterson, K. D. and G. F. Pyle (1991). "The geography and mortality of the 1918 influenza pandemic". In: *Bulletin of the History of Medicine* 65.1, pp. 4–21.

Paul, M., L. Held, and A. M. Toschke (2008). "Multivariate modelling of infectious disease surveillance data". In: *Stat. Med.* 27.29, pp. 6250–6267. ISSN: 02776715. DOI: 10.1002/sim.3440.

Pell, B., Y. Kuang, C. Viboud, and G. Chowell (2018). "Using phenomenological models for forecasting the 2015 Ebola challenge". In: *Epidemics* 22, pp. 62–70. ISSN: 18780067. DOI: 10.1016/j.epidem.2016.11.002.

Pitt, M. K. and N. Shephard (1999). "Filtering via Simulation: Auxiliary Particle Filters". In: *J. Am. Stat. Assoc.* 94.446, pp. 590–599. ISSN: 0162-1459. DOI: 10.1080/01621459.1999.10474153.

Porta, M. (2008). *A dictionary of epidemiology*. Oxford university press.

Prentice, R. L., J. D. Kalbfleisch, A. V. Peterson, N. Flournoy, V. T. Farewell, and N. E. Breslow (1978). "The Analysis of Failure Times in the Presence of Competing Risks". In: *Biometrics* 34.4, p. 541. ISSN: 0006341X. DOI: 10.2307/2530374.

Presanis, A. M., R. G. Pebody, B. J. Paterson, B. D. M. Tom, P. J. Birrell, A Charlett, M Lipsitch, and D De Angelis (2011). "Changes in severity of 2009 pandemic A/H1N1 influenza in England: a Bayesian evidence synthesis." In: *BMJ* 343.April 2009, p. d5408. ISSN: 1756-1833. DOI: 10.1136/bmj.d5408.

Presanis, A. M. et al. (2009). "The severity of pandemic H1N1 influenza in the United States, from April to July 2009: A Bayesian analysis". In: *PLoS Med.* 6.12. ISSN: 15491277. DOI: 10.1371/journal.pmed.1000207.

Presanis, A. M., R. G. Pebody, P. J. Birrell, B. D. M. Tom, R. K. Green, H. Durnall, D. Fleming, and D. De Angelis (2014). "Synthesising evidence to estimate pandemic (2009) A/H1N1 influenza severity in 2009-2011". In: *Ann. Appl. Stat.* 8.4, pp. 2378–2403. ISSN: 19417330. DOI: 10.1214/14-AOAS775. arXiv: 1408.7025.

Public Health England (2014). *Sources of UK flu data: influenza surveillance in the UK*. URL: https://www.gov.uk/guidance/sources-of-uk-flu-data-influenza-surveillance-in-the-uk (visited on 08/14/2017).

Public Health England (2013). *Surveillance of influenza and other respiratory viruses, including novel respiratory viruses, in the United Kingdom: Winter 2012/13*. URL: https://www.gov.

uk/government/uploads/system/uploads/attachment_data/file/325217/Annual_flu_report_winter_2012_to_2013.pdf (visited on 03/29/2017).

Public Health England (2014). *Surveillance of influenza and other respiratory viruses in the United Kingdom: Winter 2013/14*. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/325203/Flu_annual_report_June_2014.pdf (visited on 03/29/2017).

Public Health England (2015). *Surveillance of influenza and other respiratory viruses in the United Kingdom: winter 2014 to 2015*. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/429617/Annualreport_March2015_ver4.pdf (visited on 03/29/2017).

Public Health England (2018). *Surveillance of influenza and other respiratory viruses in the UK: Winter 2017 to 2018*. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/740606/Surveillance_of_influenza_and_other_respiratory_viruses_in_the_UK_2017_to_2018.pdf (visited on 12/13/2018).

Putter, H, M. Fiocco, and R. B. Geskus (2007). "Tutorial in biostatistics: competing risks and multi-state models". In: *Stat. Med.* 26.11, pp. 2389–2430. DOI: 10.1002/sim.2712.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Riley, S et al. (2003). "Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions". In: *Science (80-.).* 300.5627, pp. 1961–1966. ISSN: 0036-8075. DOI: 10.1126/science.1086478.

Robert, C. P., G. Casella, and G. Casella (2010). *Introducing Monte Carlo Methods with R*. Vol. 18. Springer.

Robert, C. and G. Casella (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.

Rohani, P. (2016). "Personal communication at SISMID 2016".

Ross, R. (1911). *The prevention of malaria*. John Murray; London.

Schön, T. B. and F. Lindsten (2017). "Learning of dynamical systems. Particle filters and Markov chain methods".

Schön, T. B., A. Svensson, L. Murray, and F. Lindsten (2018). "Probabilistic learning of nonlinear dynamical systems using sequential Monte Carlo". In: *Mech. Syst. Signal Process.* 104, pp. 866–883. ISSN: 10961216. DOI: 10.1016/j.ymssp.2017.10.033. arXiv: 1703.02419.

Serfling, R. E. (1963). "Methods for current statistical analysis of excess pneumonia-influenza deaths." In: *Public Health Rep.* 78.6, pp. 494–506. ISSN: 0094-6214.

Shaman, J., A. Karspeck, W. Yang, J. Tamerius, and M. Lipsitch (2013). "Real-time influenza forecasts during the 2012–2013 season". In: *Nat. Commun.* 4.1, p. 2837. ISSN: 2041-1723. DOI: 10.1038/ncomms3837.

Sherlock, C., P. Fearnhead, and G. O. Roberts (2010). "The Random Walk Metropolis: Linking Theory and Practice Through a Case Study". In: *Stat. Sci.* 25.2, pp. 172–190. ISSN: 0883-4237. DOI: 10.1214/10-STS327. arXiv: arXiv:1011.6217v1.

Shubin, M., M. Virtanen, S. Toikkanen, O. Lyytikäinen, and K. Auranen (2014). "Estimating the burden of A(H1N1)pdm09 influenza in Finland during two seasons". In: *Epidemiol. Infect.* 142.5, pp. 964–974. ISSN: 14694409. DOI: 10.1017/S0950268813002537.

Shubin, M., A. Lebedev, O. Lyytikäinen, and K. Auranen (2016). "Revealing the True Incidence of Pandemic A(H1N1)pdm09 Influenza in Finland during the First Two Seasons - An Analysis Based on a Dynamic Transmission Model". In: *PLoS Comput. Biol.* 12.3, pp. 1–19. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1004803.

Soetaert, K., T. Petzoldt, and R. W. Setzer (2010). "Package deSolve : Solving Initial Value Differential Equations in R". In: *J. Stat. Softw.* 33.9, pp. 1–25. ISSN: 15487660. DOI: 10.18637/jss.v033.i09.

Solin, A., M. Kok, N. Wahlstrom, T. B. Schon, and S. Sarkka (2018). "Modeling and Interpolation of the Ambient Magnetic Field by Gaussian Processes". In: *IEEE Trans. Robot.* 34.4, pp. 1112–1127. ISSN: 15523098. DOI: 10.1109/TRO.2018.2830326. arXiv: 1509.04634.

Stewart, L and P McCarthy Jr (1992). "The use of Bayesian belief networks to fuse continuous and discrete information for target recognition, tracking and situation assessment". In: *Proc SPIE 1699 Signal Process. Sens. Fusion Target Recognit.* July 1992, pp. 177–184. ISSN: 00396028. DOI: 10.1117/12.138224.

Stewart, T., L. Strijbosch, H. Moors, and P. van Batenburg (2007). "A Simple Approximation to the Convolution of Gamma Distributions (Revision of DP 2006-27)". In: *Cent. Discuss. Pap.* 2007-70.

Streftaris, G. and G. J. Gibson (2002). "Statistical inference for stochastic epidemic models". In: *Proc. 17th Int. Work. Stat. Model. Chania* January 2002, pp. 1–8.

Taubenberger, J. K. and D. M. Morens (2006). "1918 Influenza: The mother of all pandemics". In: *Emerg. Infect. Dis.* 12.1, pp. 15–22. ISSN: 10806059. DOI: 10.3201/eid1209.05-0979.

Taubenberger, J. K. and D. M. Morens (2008). "The Pathology of Influenza Virus Infections". In: *Annu. Rev. Pathol. Mech. Dis.* 3.1, pp. 499–522. ISSN: 1553-4006. DOI: 10.1146/annurev.pathmechdis.3.121806.154316.

Te Beest, D. E., P. J. Birrell, J. Wallinga, D. De Angelis, and M. van Boven (2015). "Joint modelling of serological and hospitalization data reveals that high levels of pre-existing immunity and school holidays shaped the influenza A pandemic of 2009 in the Netherlands." In: *J. R. Soc. Interface* 12.103, p. 20141244. ISSN: 1742-5662. DOI: 10.1098/rsif.2014.1244.

The Phoenix Partnership (2013). *Real-time Syndromic Surveillance*. URL: www.researchone.org/public-health-monitoring/ (visited on 12/03/2018).

Tom, B. D. M., A. J. van Hoek, R Pebody, J McMenamin, C Robertson, M Catchpole, and D De Angelis (2011). "Estimating time to onset of swine influenza symptoms after initial novel A(H1N1v) viral infection." In: *Epidemiol. Infect.* 139.9, pp. 1418–1424. ISSN: 1469-4409. DOI: 10.1017/S0950268810002566.

Treanor, J. (2004). "Influenza Vaccine — Outmaneuvering Antigenic Shift and Drift". In: *N. Engl. J. Med.* 350.3, pp. 218–220. ISSN: 0028-4793. DOI: 10.1056/NEJMp038238.

Tuite, A. R. and D. N. Fisman (2018). "The IDEA model: A single equation approach to the Ebola forecasting challenge". In: *Epidemics* 22, pp. 71–77. ISSN: 18780067. DOI: 10.1016/j.epidem.2016.09.001.

UNAIDS (2018). *Fact sheet - Latest statistics on the status of the AIDS epidemic. 2017 GLOBAL HIV STATISTICS*. URL: http://www.unaids.org/en/resources/fact-sheet (visited on 07/30/2018).

Van Houwelingen, H. and H. Putter (2011). *Dynamic prediction in clinical survival analysis*. CRC Press.

Van Kerkhove, M. D. et al. (2010). "Studies Needed to Address Public Health Challenges of the 2009 H1N1 Influenza Pandemic : Insights from Modeling". In: *PLoS Med.* 7.6, pp. 1–6. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1000275.

Veach, E. and L. J. Guibas (1995). "Optimally combining sampling techniques for Monte Carlo rendering". In: *Proc. 22nd Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '95*, pp. 419–428. ISSN: 00978930. DOI: 10.1145/218380.218498.

Viboud, C., P. Y. Boëlle, F. Carrat, A. J. Valleron, and A. Flahault (2003). "Prediction of the Spread of Influenza Epidemics by the Method of Analogues". In: *Am. J. Epidemiol.* 158.10, pp. 996–1006. ISSN: 00029262. DOI: 10.1093/aje/kwg239.

Vynnycky, E and W. J. Edmunds (2008). "Analyses of the 1957 (Asian) influenza pandemic in the United Kingdom and the impact of school closures". In: *Epidemiol. Infect.* 136.2, pp. 166–179. ISSN: 0950-2688. DOI: 10.1017/s0950268807008369.

Vynnycky, E. and R. White (2010). *An introduction to infectious disease modelling*. Oxford University Press.

WHO (2010). *H1N1 in post-pandemic period*. URL: http://www.who.int/mediacentre/news/statements/2010/h1n1_vpc_20100810/en/.

WHO (2013). *Global Epidemiological Surveillance Standards for Influenza*. Tech. rep.

WHO (2014). *Metrics: Disability-Adjusted Life Year (DALY). Quantifying the Burden of Disease from mortality and morbidity*. URL: http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/ (visited on 10/20/2018).

WHO (2015). *A manual for estimating disease burden associated with seasonal influenza*. World Health Organization.

WHO (2016). *Ebola Situation Report - 30 March 2016*. URL: http://apps.who.int/ebola/current-situation/ebola-situation-report-30-march-2016 (visited on 07/30/2018).

WHO (2018a). *Health topics. Infectious diseases*. URL: http://www.who.int/topics/infectious_diseases/en/ (visited on 07/31/2018).

WHO (2018b). *Influenza (Seasonal)*. URL: http://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal) (visited on 10/19/2018).

Wearing, H. J., P. Rohani, and M. J. Keeling (2005). "Appropriate models for the management of infectious diseases". In: *PLoS Med.* 2.7, pp. 0621–0627. ISSN: 15491277. DOI: 10.1371/journal.pmed.0020174.

Webster, R. G., A. S. Monto, T. J. Braciale, and R. A. Lamb, eds. (2013). *Textbook of Influenza*. Oxford, UK: John Wiley & Sons, Ltd. ISBN: 9781118636817. DOI: 10.1002/9781118636817.

Wikramaratna, P. S., A. Kucharski, S. Gupta, V. Andreasen, A. R. McLean, and J. R. Gog (2015). "Five challenges in modelling interacting strain dynamics". In: *Epidemics* 10, pp. 31–34. ISSN: 18780067. DOI: 10.1016/j.epidem.2014.07.005.

Wong, J. Y., H. Kelly, D. K. M. Ip, J. T. Wu, G. M. Leung, and B. J. Cowling (2013a). "Case Fatality Risk of Influenza A (H1N1pdm09): A Systematic Review Jessica". In: *Epidemiology* 24.6, pp. 830–841. ISSN: 1044-3983. DOI: 10.1097/EDE.0b013e3182a67448.

Wong, J. Y. et al. (2013b). "Infection fatality risk of the pandemic a(H1N1)2009 virus in Hong Kong". In: *Am. J. Epidemiol.* 177.8, pp. 834–840. ISSN: 00029262. DOI: 10.1093/aje/kws314.

Yaari, R., G. Katriel, L. Stone, E. Mendelson, M. Mandelboim, and A. Huppert (2016). "Model-based reconstruction of an epidemic using multiple datasets: understanding influenza A/H1N1 pandemic dynamics in Israel." In: *J R. Soc Interface* 13.116, pp. 92–92. ISSN: 17425662. DOI: 10.1098/rsif.2016.0099.

Yang, W., A. Karspeck, and J. Shaman (2014). "Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics". In: *PLoS Comput. Biol.* 10.4. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1003583.

Yang, W., M. Lipsitch, and J. Shaman (2015). "Inference of seasonal and pandemic influenza transmission dynamics". In: *Proc. Natl. Acad. Sci.* 112.9, pp. 2723–2728. ISSN: 0027-8424. DOI: 10.1073/pnas.1415012112.

Yip, P. S. F., E. H. Y. Lau, K. F. Lam, and R. M. Huggins (2005a). "A chain multinomial model for estimating the real-time fatality rate of a disease, with an application to severe acute respiratory syndrome". In: *Am. J. Epidemiol.* 161.7, pp. 700–706. ISSN: 00029262. DOI: 10.1093/aje/kwi088.

Yip, P. S. F., K. F. Lam, E. H. Y. Lau, P. H. Chau, K. W. Tsang, and A. Chao (2005b). "A comparison study of realtime fatality rates: Severe acute respiratory syndrome in Hong Kong, Singapore, Taiwan, Toronto and Beijing, China". In: *J. R. Stat. Soc. Ser. A Stat. Soc.* 168.1, pp. 233–243. ISSN: 09641998. DOI: 10.1111/j.1467-985X.2004.00345.x.

Appendix A

Some parametric distributions

Several parametric distributions benefit of multiple formulations and parametrizations. To avoid misunderstanding, the density function of the distributions used in this thesis are listed below.

A.1 Gamma

A Gamma random variable (r.v.) $X > 0$ is a continuous r.v.. The Gamma distribution is parametrised in terms of:

$\alpha > 0$, the shape parameter

$\beta > 0$, the rate parameter

and is denoted by

$$X \sim \text{Gamma}(\alpha, \beta)$$

The probability density function of X is

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

with

$$\mathbb{E}(X) = \frac{\alpha}{\beta}$$
$$\text{Var}(X) = \frac{\alpha}{\beta^2}$$

A property of the Gamma distribution is that if $X \sim \text{Gamma}(\alpha, \beta)$ and given a positive real number c, then the r.v. $Y = cX$ is also Gamma-distributed with parameters α and $\frac{\beta}{c}$:

$$Y \sim \text{Gamma}\left(\alpha, \frac{\beta}{c}\right)$$

A.2 Negative Binomial

A Negative Binomial r.v. $X > 0$ is a discrete r.v. that models the number of successes in a sequence of independent and identically distributed (iid) Bernoulli trials before a specified (non-random) number of failures occurs. The Negative Binomial distribution is parametrised in terms of:

$r > 0$, the number of failures

$p \in (0, 1)$, the probability of success

and is denoted by

$$X \sim \text{NegBin}(r, p)$$

The probability mass function of X is

$$\begin{aligned} f(x|r,p) =& \binom{x+r-1}{x} p^x (1-p)^r \\ =& \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} p^x (1-p)^r \end{aligned}$$

with

$$\begin{aligned} \mathbb{E}(X) &= \frac{rp}{(1-p)} \\ \text{Var}(X) &= \frac{rp}{(1-p)^2} \end{aligned}$$

A.2.1 Re-parametrization

In this thesis, as well as in the literature on Negative Binomial regression, a re-parametrization is adopted. Denote by η the over-dispersion parameter and let:

$$\eta = \frac{1}{1-p}$$

The Negative Binomial distribution is then denoted by

$$X \sim \text{NegBin}(r, \eta)$$

with density function

$$f(x|r,\eta) = \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} \left(1 - \frac{1}{\eta}\right)^x \left(\frac{1}{\eta}\right)^r$$

This parametrization allows r to be expressed as a function of the mean, denoted here by $\mu = \frac{rp}{(1-p)}$, and the over-dispersion parameter η:

$$r = \frac{\mu}{\eta - 1}$$

having $\text{Var}(X) = \eta\mu$. This parametrization can also be expressed as: $X \sim \text{NegBin}(\mu, \eta)$.

A.2.2 Poisson-Gamma formulation

The Negative Binomial r.v. can also be obtained from a Poisson r.v. with Gamma-distributed rate. Let

$$\begin{aligned} X|(\Lambda = \lambda) &\sim \text{Pois}(\lambda) \\ \Lambda|\alpha, \beta &\sim \text{Gamma}(\alpha, \beta) \end{aligned}$$

The distribution of $X|\alpha,\beta$ can be derived as follows

$$\begin{aligned}
f(x|\alpha,\beta) &= \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\mathrm{d}\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)x!}\int_0^\infty e^{-\lambda}\lambda^x\lambda^{\alpha-1}e^{-\beta\lambda}\mathrm{d}\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)\Gamma(x+1)}\int_0^\infty e^{-\lambda(\beta+1)}\lambda^{x+\alpha-1}\mathrm{d}\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)\Gamma(x+1)}\frac{\Gamma(x+\alpha)}{(\beta+1)^{x+\alpha}} \\
&= \frac{\Gamma(x+\alpha)}{\Gamma(x+1)\Gamma(\alpha)}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^x
\end{aligned}$$

which can be recognised as a Negative Binomial probability distribution:

$$X \sim \text{NegBin}\left(r=\alpha, p=\frac{1}{\beta+1}\right)$$

or, in the re-parametrization above:

$$X \sim \text{NegBin}\left(r=\alpha, \eta=1+\frac{1}{\beta}\right)$$

A.3 Beta

The Beta r.v. $X \in [0,1]$ is a continuous r.v.. The Beta distribution is parametrised in terms of:

$\alpha > 0$, the first shape parameter

$\beta > 0$, the second shape parameter

and is denoted by

$$X \sim \text{Beta}(\alpha,\beta)$$

The probability density function of X is

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

with

$$\begin{aligned}
\mathbb{E}(X) &= \frac{\alpha}{\alpha+\beta} \\
\text{Var}(X) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

The probability density function can be rewritten by defining the beta function:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

as:

$$f(x|\alpha,\beta) = B(\alpha,\beta)x^{\alpha-1}(1-x)^{\beta-1}$$

A property of the Beta distribution is that if $X \sim \text{Gamma}(\alpha,\beta)$ and $Y \sim \text{Gamma}(\theta,\beta)$ then the r.v. $Z = \frac{X}{X+Y}$ has the following distribution

$$\frac{X}{X+Y} \sim \text{Beta}\left(\alpha,\theta\right)$$

A.4 Beta Binomial

The Beta Binomial r.v. $X \in \{0, 1, \ldots, n\}$ is a discrete r.v.. The Beta Binomial distribution is parametrised in terms of:

$n \in \mathbb{N}$ the number of trials

$\alpha > 0$, the first shape parameter

$\beta > 0$, the second shape parameter

and is denoted by

$$X \sim \text{BetaBin}(\alpha, \beta, n)$$

The probability mass function of X is

$$\begin{aligned} f(x|n, \alpha, \beta) &= \binom{n}{x} \frac{B(x + \alpha, n - x + \beta)}{B(\alpha, \beta)} \\ &= \frac{\Gamma(\alpha + n + \beta)}{\Gamma(x + \alpha)\Gamma(n - x + \beta)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \end{aligned}$$

with

$$\begin{aligned} \mathbb{E}(X) &= \frac{n\alpha}{\alpha + \beta} \\ \text{Var}(X) &= \frac{n\alpha\beta(\alpha + \beta + n)}{(\alpha + \beta)^2(\alpha + \beta + 1)} \end{aligned}$$

The Beta Binomial distribution can be obtained as the conjugate distribution of a Binomial with Beta-distributed probability parameter:

$$\begin{aligned} X|n, \theta &\sim \text{Bin}(n, \theta) \\ \theta|\alpha, \beta &\sim \text{Beta}(\alpha, \beta) \\ X|n, \alpha, \beta &\sim \text{BetaBin}(\alpha, \beta, n) \end{aligned}$$

Appendix B

Additional results to the analysis of Severity

This appendix reports a further exploration of the application of parametric survival analysis to the context of case fatality risk (CFR) estimates.

B.1 Other distributional assumptions

The `R` package `flexsurv` (Jackson, 2015) allows the fit of several parametric distributions to observed data within a competing-risks setting.

The package works by fitting, via maximum likelihood, a chosen parametric model to cause-specific times to event.

Many parametric distributions can be fitted to the data, namely: Exponential, Weibull, Gamma, Log-Normal, Gompertz, Log-Logistic, and Generalised Gamma. Each of these distributions has its own parametrization.

As explained in the the introduction to competing-risks models in Section 2.3.1, the most useful quantity is the cumulative intensity function, $I_h(t)$, defined as:

$$P(T \leq t; D = h) = I_h(t) = \int_0^t \alpha_h(s) S(s) \mathrm{d}s \tag{B.1}$$

Two challenges arise when attempting to estimate this quantity in a parametric setting.

(i) The hazard function $\alpha_h(s)$ must have an explicit parametric form.
This is true for only a few of the parametric functions considered, among which Exponential and Weibull; however, for most of the distributions used for survival data, a specific formula for
$$\alpha_h(t) = \lim_{\Delta t \to 0} \frac{P(T \leq t + \Delta t; D = h | T \geq t)}{\Delta t} = -\frac{\mathrm{d} \log S_h(t)}{\mathrm{d}t}$$
cannot be derived.

(ii) The overall survival function, which is defined as:
$$S(t) = e^{-\sum_{h=1}^{k} A_h(t)}, \qquad \text{with} \qquad A_h(t) = \int_0^t \alpha_h(s) \mathrm{d}s \tag{B.2}$$
cannot be simplified into a solvable function.

These challenges prevent the computation of the exact limit of the cumulative intensity function of death for $t \to \infty$ and hence the application of other parametric distributions for the inference of the CFR.

Appendix C

Appendix to Section 3.3

This appendix contains further information to the study presented in Section 3.3. Most of the content of this appendix is published as Supplementary Information to Corbella et al. (2018). Section C.1 contains the derivation of the observational model (Equation 3.11 of the main text). Section C.2 reports all the results obtained on the analysis of real data. Lastly, in Section C.3 additional results from the analysis of simulated pandemic data are reported.

C.1 Estimation of the observational model

The density function of the time from infection to Intensive Care Unit (ICU) admission plays a crucial role in the computation of the likelihood of the UK Severe Influenza Surveillance System (USISS) data, as can be seen from Equation 3.11 of the main text:

$$\mu_w = \sum_{v=0}^{w} f_{w-v}^{\mathrm{ICU|I}} \cdot \Delta I_v p_{ICU}.$$

$f_w^{\mathrm{ICU|I}}$ denotes the probability of waiting w weeks between infection and ICU admission. This information, combined with the number of new infections ΔI_v and the probability of ICU admission given contact p_{ICU} allows the computation of the average number of ICU admission at a given week w, denoted by μ_w. The process is represented in figure C.1 and comprises the random variable (r.v.) Y, representing the incubation time, and the r.v. Z representing the time from symptom onset to ICU admission.

The distribution of Y was estimated by Tom et al., 2011 via parametric survival analysis of individual-level data, and resulted in the following approximation:

$$Y \sim \mathrm{Gamma}(\alpha = 0.678, \beta = 0.417). \tag{C.1}$$

Data form the USISS sentinel scheme were used to approximate the distribution of Z, the time from symptom onset to ICU admission. The patients admitted to ICU in the sentinel trusts were asked to report the day of their first symptoms. Several parametric survival curves are fitted to the data on 120 patients admitted to ICU during influenza seasons 2011/12 and 2012/13 and the model with lowest Akaike information criterion (AIC) (Akaike, 1987) was selected.

The fitted distributions are reported in Figure C.2 and the results are reported in Table C.1.

In conclusion, Z was approximated by:

$$Z \sim \mathrm{LogNormal}(\mu_{\log} = 0.607, \sigma_{\log} = 1.022) \tag{C.2}$$

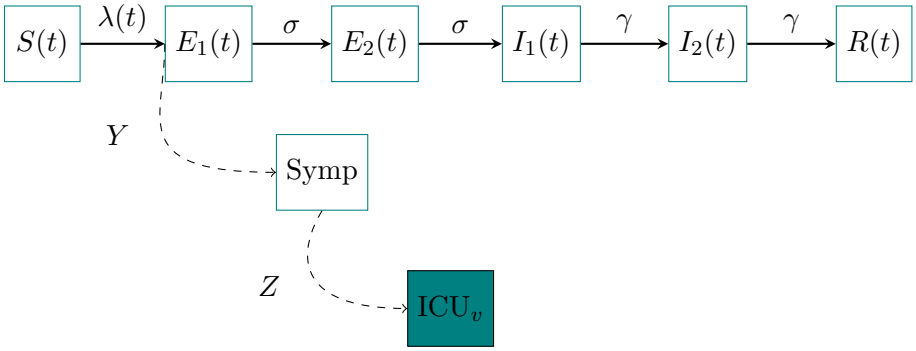


Figure C.1: Extended observational model of ICU/HDU admissions. Transmission model is linked to ICU/HDU admissions thought the r.v.s Y, the incubation time, and Z, the time from symptom to ICU admission.

model	par1(SE)	par2(SE)	logL	AIC
Exponential	rate = 0.32 (se= 0.028)	/	-280.3059	562.6118
Weibull	shape = 0.976 (se= 0.062)	scale = 3.087 (se= 0.293)	-280.2274	564.4549
LogNormal	meanlog = 0.607 (se= 0.089)	sdlog = 1.022 (se= 0.063)	-268.2563	540.5126
Gompertz	shape = -0.047 (se= 0.023)	rate = 0.377 (se= 0.043)	-277.837	559.6741
Gamma	shape = 1.074 (se= 0.118)	rate = 0.344 (se= 0.047)	-280.0966	564.1933

Table C.1: Estimates of the parameters and AICs of the parametric models used to fit data on the time from symptom onset to ICU admission.

The probability density functions of the two r.v.s X and Y are reported in Figure C.3.

To compute $f_{ICU|I}(w)$ for week $w = 0, 1, 2, \ldots$ the cumulative distribution function of the r.v. T, the time in days from infection to ICU admission, is defined as the convolution of Y and Z. Denoted by $F_X(x)$ the cumulative distribution function of a general r.v. X ($F_X(x) = \mathrm{P}(X \leq x)$) and by $f_X(x)$ its probability density function ($f_X(x) = \mathrm{P}(X = x)$). The convolution for T can be obtained by

$$F_T(t) = P(T \leq t) = \int_{y=0}^{t} f_Y(y) \cdot F_z(t-y)dy \tag{C.3}$$

$F_T(t)$ is computed by numerical integration for $t = 7, 14, 21, \ldots$ so that the probability of w week elapsing from infection to ICU admission can be simply calculated using equation C.4 and its values for small number of weeks are reported in Table C.2.

$$f_w^{\mathrm{ICU|I}} = F_T(w \cdot 7) - F_T((w-1) \cdot 7) \tag{C.4}$$

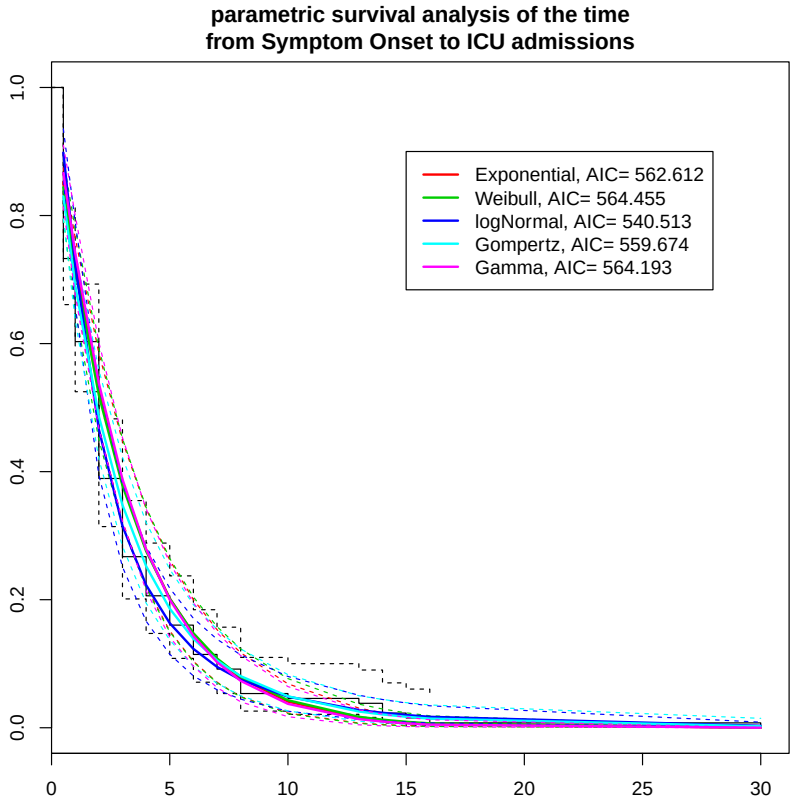


Figure C.2: Fitted parametric distributions to the time form symptom onset to ICU admission. Dotted lines represent 95% CI.

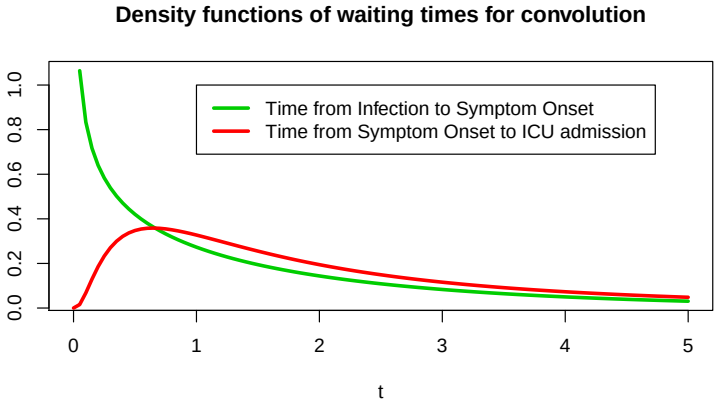


Figure C.3: Estimated density function for the incubation time Y and the time from symptoms to ICU admission Z.

week	$f_{ICU\|I}(w)$
1	0.81128
2	0.15180
3	0.02546
4	0.00670
5	0.00242
6	0.00106
7	0.00052
8	0.00028
9	0.00016
...	...

Table C.2: Estimated probability function of the convolution over weeks of the incubation time Y and the time from symptom onset to ICU admission Z.

C.2 Supplementary results

This section reports the results from all the analyses. There is significant overlap with the main text but repetitions were allowed for the sake of completeness of this appendix.

C.2.1 Results on the full datasets

In this section all the results obtained from the analysis of real data are reported. Results are divided according to the priors used: firstly, results from the uninformative scenario, with Uniform priors for all the parameters, are given. Table C.3 lists the priors and the values assumed known and fixed. Secondly, the results of the analyses within the informative scenario are shown, all the parameters except for p_{ICU} and π are assumed to be Uniformly distributed. p_{ICU} and π are distributed as specified in the main text and reported in Table C.4.

Unknown parameters definition	Parameter	Distribution
Susceptibility	π	$\sim$ Unif$(0, 1)$
Initial infection rate	λ_0	$\sim$ Unif$(0, 0.000241)$
Exponential growth rate	ψ	$\sim$ Unif$(0, 0.39)$
Over-dispersion	η	$\sim$ Unif$(1, 100)$
Probability of ICU admission given infection	p_{ICU}	$\sim$ Unif$(0, 1)$
Parameters assumed known	**Parameter**	**Value**
Rate of becoming infectious	σ	1
Rate of recovery	γ	0.5797
Mean time from infection to ICU admission	$\mu_{ICU\|E}$	5.708
Variance of time from infection to ICU admission	$\sigma^2_{ICU\|E}$	18.24
Population of 2012/13	$N_{2012/13}$	53,679,750
Population of 2013/14	$N_{2013/14}$	54,091,200
Population of 2014/15	$N_{2014/15}$	54,551,450

Table C.3: Flat prior distributions and fixed parameters.

Parameter	Distribution
π	$\sim$ LogNorm$(\log \mu = \log(0.401), \log \sigma = 0.2)$ Hoschler et al., 2012
p_{ICU}	$\sim$ LogNorm$(\log \mu = \log(0.000239), \log \sigma = 1)$ Presanis et al., 2014

Table C.4: Informative prior distributions from previous findings.

Results with uniform priors

To obtain these results 3 independent chains of 1.1 million iterations of the blocked Metropolis Hastings (MH) algorithm were run. Of these, the first 100,000 iterations were used adaptively to tune the algorithm (i.e. to estimate the mixing parameter ν) and the following 200,000 were discarded as burn-in period. Moreover, a thinning factor of 100 was used, saving only one iteration every 100.

Figure C.4 displays the chains of the 6 parameters of the system. Despite all the parameters showing convergence, the chains of some of them (namely π and ψ) are moving over a uniform distribution and they are severely correlated. The plots of the prior and the posterior distribu-

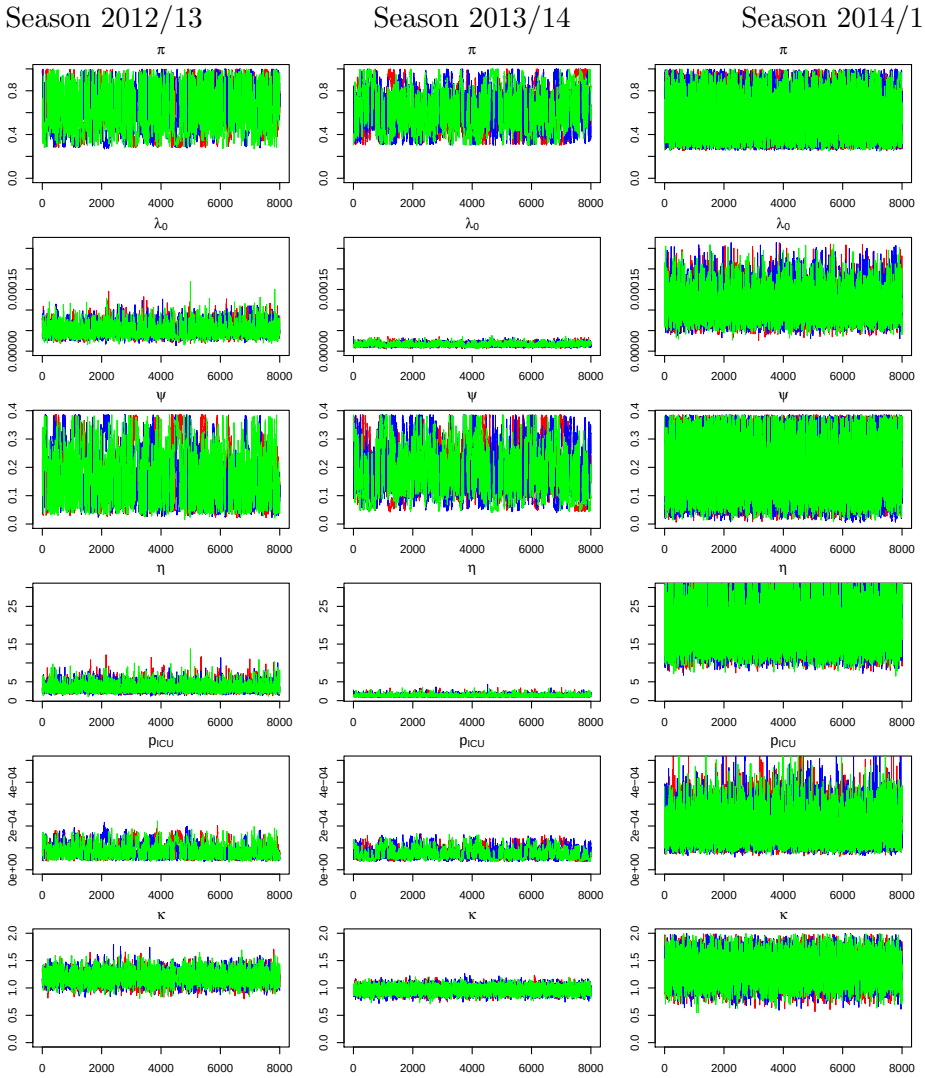


Figure C.4: Flat scenario - The three independent chains are plotted here with three different colours (red, blue and green). The parameters are (in order, from top to bottom): the initial susceptibility π, the initial infection rate λ_0, the exponential growth rate ψ, the over-dispersion parameter η, the probability of ICU admission given infection p_{ICU}, and the scaling parameter κ. The results are derived from season 2012/13 (left column), season 2013/14 (centre) and season 2014/15 (right column).

tions of all the parameters are reported in Figure C.5. Similarly, Figure C.6 contains the plots of the prior and posterior distributions of the other functional quantities of interest commented in the main text.

Table C.5 reports median and 95% credible intervals (CrIs) of all the parameters.

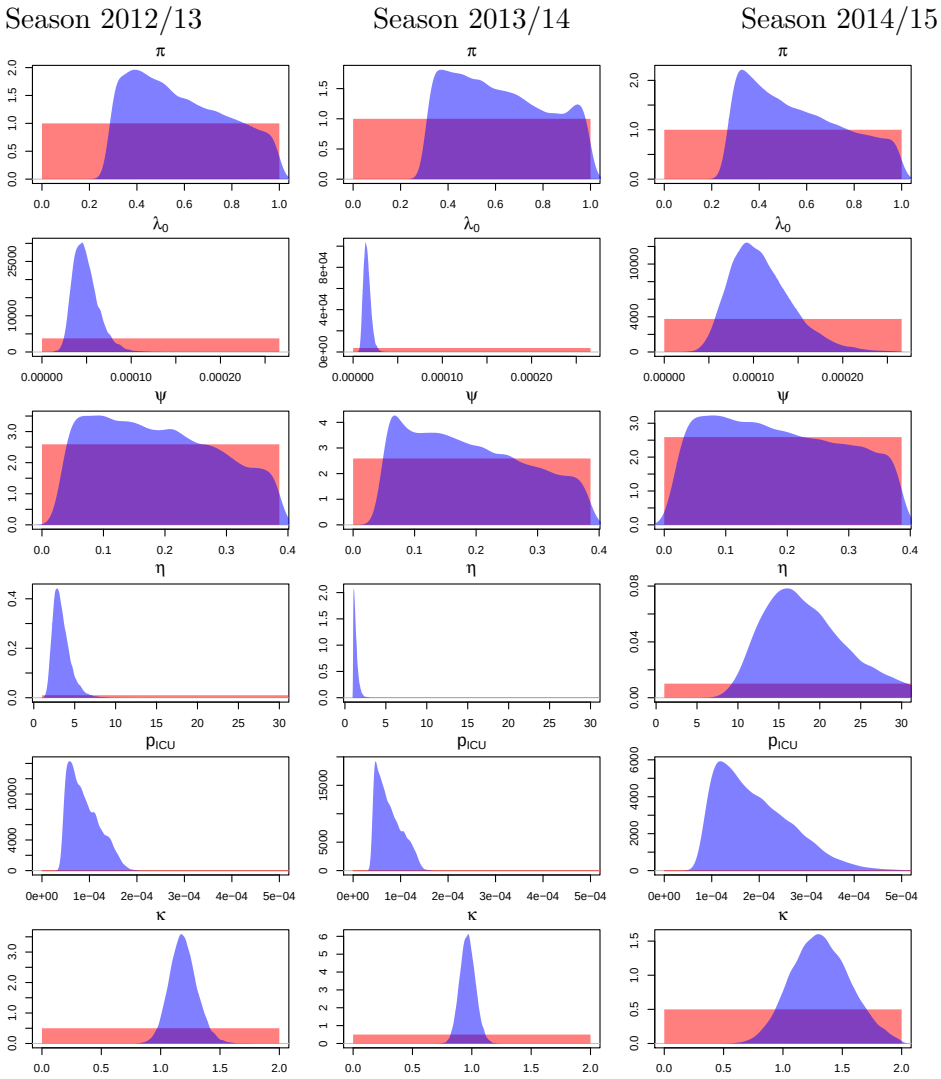


Figure C.5: Flat scenario - Prior (red) and posterior (blue) distributions of the parameters. The parameters are (in order, from top to bottom): π, λ_0, ψ, η, p_{ICU} and κ. The results are derived from season 2012/13 (left column), season 2013/14 (centre) and season 2014/15 (right column).

	2012/13	2013/14	2014/15
π	0.557 (0.298 - 0.969)	0.599 (0.323 - 0.978)	0.53 (0.28 - 0.968)
ψ	0.471 (0.271 - 0.842)	0.152 (0.091 - 0.248)	1.046 (0.549 - 1.921)
λ_0	0.18 (0.039 - 0.374)	0.179 (0.053 - 0.371)	0.178 (0.024 - 0.374)
η	3.204 (1.886 - 6.129)	1.251 (1.011 - 2.088)	17.947 (10.429 - 35.711)
p_{ICU}	0.083 (0.046 - 0.16)	0.07 (0.042 - 0.133)	0.175 (0.085 - 0.375)
κ	1.186 (0.974 - 1.437)	0.964 (0.84 - 1.099)	1.312 (0.864 - 1.82)
I_0^{tot}	4219.9 (1482.5 - 11510.8)	1386.2 (493 - 3328.4)	9568.6 (3052.3 - 28529)
β	0.6 (0.345 - 1.121)	0.598 (0.366 - 1.111)	0.597 (0.324 - 1.119)
R_0	2.071 (1.189 - 3.868)	2.063 (1.263 - 3.834)	2.058 (1.116 - 3.859)
R_n	1.152 (1.093 - 1.209)	1.235 (1.196 - 1.275)	1.089 (0.997 - 1.195)

Table C.5: Flat scenario - Posterior medians and 95% CrIs form the retrospective analysis of the ICU admissions over the three years considered.

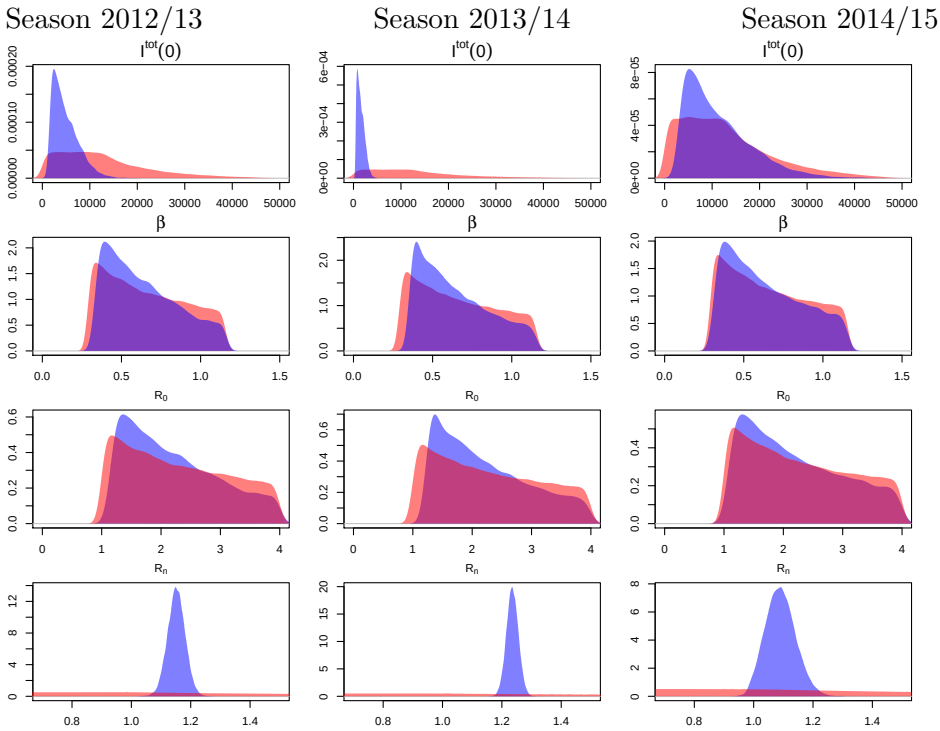


Figure C.6: Flat scenario - Prior (red) and posterior (blue) distributions of: the total number of initial infectious I_0^{tot}, the basic transmission rate β, the basic reproduction number R_0, and the effective reproduction number R_n.

Results with informative priors

The same analysis was performed assuming prior information on some of the parameters as reported in Table C.4.

To obtain these results, 3 independent chains of 1.1 million iterations of the blocked MH algorithm were run. Of these, the first 100,000 iterations were used adaptively to tune the algorithm (i.e. to estimate the mixing parameter ν) and the following 200,000 were discarded as burn-in period. Moreover, a thinning factor of 100 was used, saving only one iteration every 100.

Figure C.7 displays the chains of the 6 parameters of the system. All the chains have converged but correlation persists.

The plots of the prior and the posterior distributions of all the parameters are reported in figure C.8. Similarly, Figure C.9 contains the plots of the prior and posterior distributions of the functional quantities of interest.

The lack of information of the data has been compensated by the introduction of prior of information. Due to the very high correlation that is affecting the system, by putting a prior on π and p_{ICU} other parameters, such as ψ, β and R_0 are also affected. The posterior distributions of these parameters are mainly driven by the prior distribution on π and p_{ICU}. As a signal of this problem the posterior distributions looks almost identical across the different seasons (i.e. with different data).

Table C.6 reports median and CrIs of all the parameters.

Lastly, Figure C.10, reports the posterior predictive distribution of the number of ICU admissions. There is no significant improvement compared to the same result within the uninformative

	2012/13	2013/14	2014/15
π	0.4 (0.3 - 0.581)	0.403 (0.316 - 0.58)	0.406 (0.292 - 0.588)
λ_0	0.405 (0.242 - 0.681)	0.128 (0.082 - 0.201)	0.916 (0.508 - 1.601)
ψ	0.278 (0.167 - 0.373)	0.297 (0.188 - 0.379)	0.256 (0.147 - 0.364)
η	3.188 (1.889 - 6.039)	1.248 (1.012 - 2.113)	17.503 (10.323 - 33.765)
p_{ICU}	0.114 (0.077 - 0.162)	0.104 (0.071 - 0.137)	0.22 (0.138 - 0.358)
κ	1.184 (0.97 - 1.433)	0.966 (0.842 - 1.097)	1.318 (0.883 - 1.825)
I_0^{tot}	2603.2 (1318.1 - 5533.9)	787 (431 - 1558.3)	6408.2 (2928.1 - 14205.3)
β	0.835 (0.572 - 1.117)	0.888 (0.617 - 1.135)	0.778 (0.532 - 1.089)
R_0	2.882 (1.974 - 3.853)	3.065 (2.129 - 3.916)	2.684 (1.834 - 3.757)
R_n	1.153 (1.093 - 1.213)	1.235 (1.196 - 1.276)	1.089 (0.997 - 1.194)

Table C.6: Informative scenario - Posterior medians and 95% CrIs form the retrospective analysis of the ICU admissions over the three years considered.

scenario (Figure 3.13 of the main text).

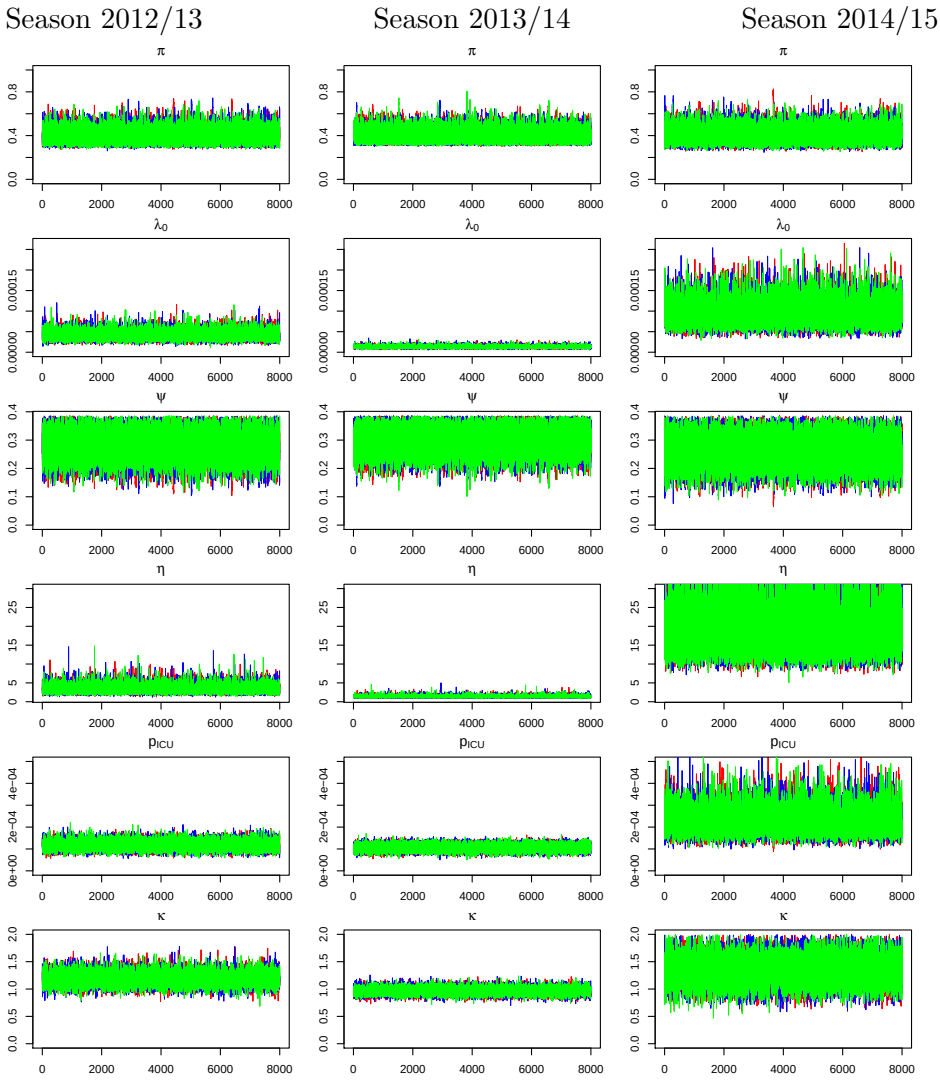


Figure C.7: Informative scenario - The three independent chains are plotted here with three different colours (red, blue and green). The parameters are (in order, from top to bottom): π, λ_0, ψ, η, p_{ICU} and the scaling parameter, κ. The results are derived from season 2012/13 (left column), season 2013/14 (centre) and season 2014/15 (right column).

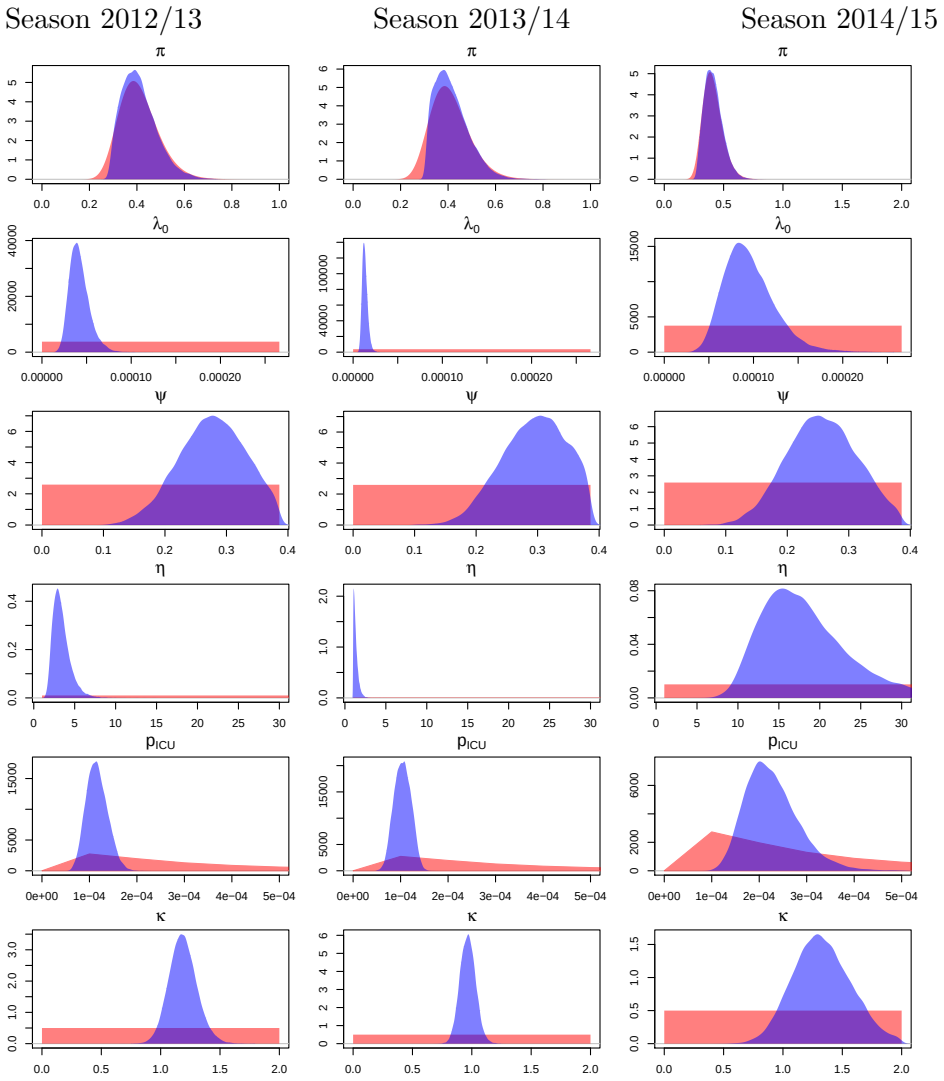


Figure C.8: Informative scenario - Prior (red) and posterior (blue) distributions of the parameters. The parameters are (in order, from top to bottom): π, λ_0, ψ, η, p_{ICU} and κ. The results are derived from season 2012/13 (left column), season 2013/14 (centre) and season 2014/15 (right column).

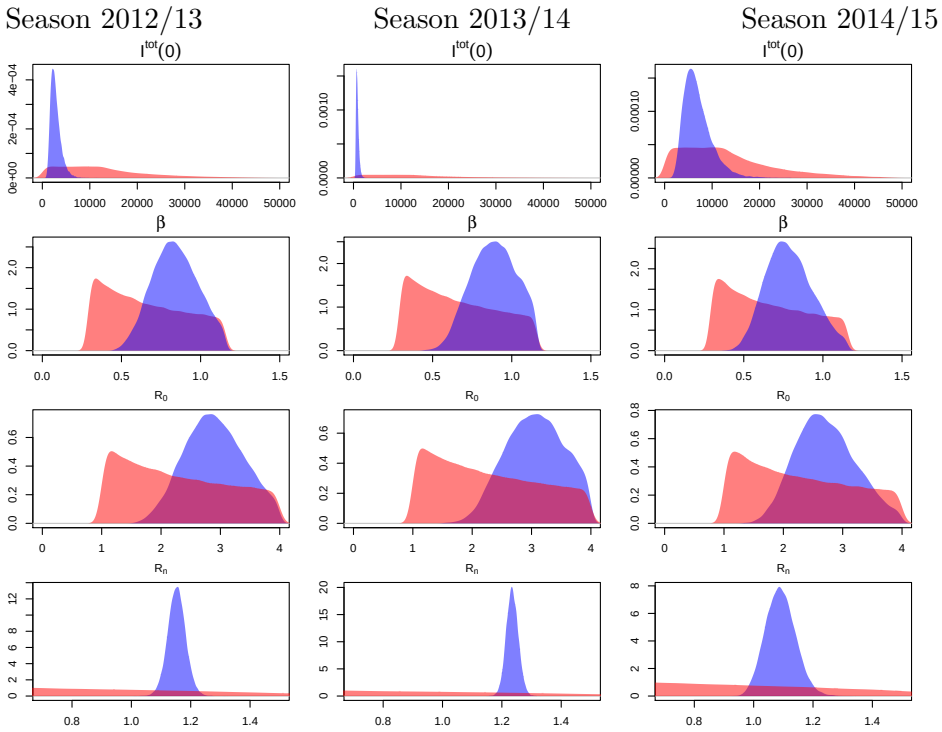


Figure C.9: Informative scenario - Prior (red) and posterior (blue) distributions of (from top to bottom): the total number of initial infectious I_0^{tot}, the basic transmission rate β, the basic reproduction number R_0, and the effective reproduction number R_n.

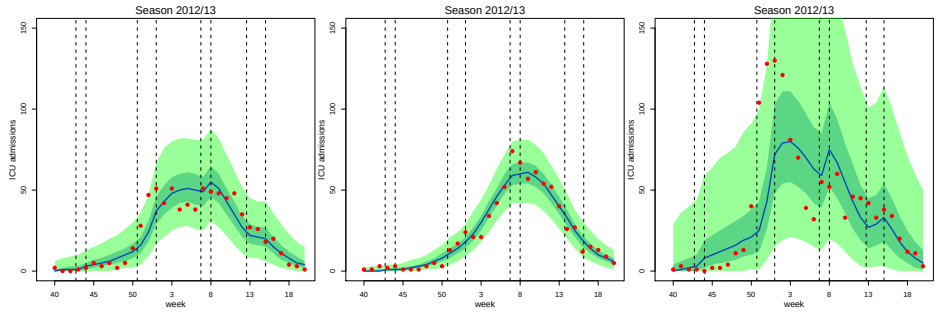


Figure C.10: Informative scenario - Median (blue), 95 % CrI (light green) and quartile (dark green) of the posterior predictive distributions and observed values (red) of the weekly ICU/HDU admissions across seasons. The vertical dashed lines represent the breakpoints for the piecewise-constant transmissibility β^*.

C.2.2 Results on the datasets updated every five weeks

The sequential learning of the parameter via inclusion of more and more data has already been discussed in the main text. For this reason, only the predictive performance of the model in the uninformative scenario is reported here, for the sake of comparison with the results presented in Figure 3.15 of the main text. This is reported in Figure C.11 and it highlights the extreme need of informative priors in order to obtain useful predictions.

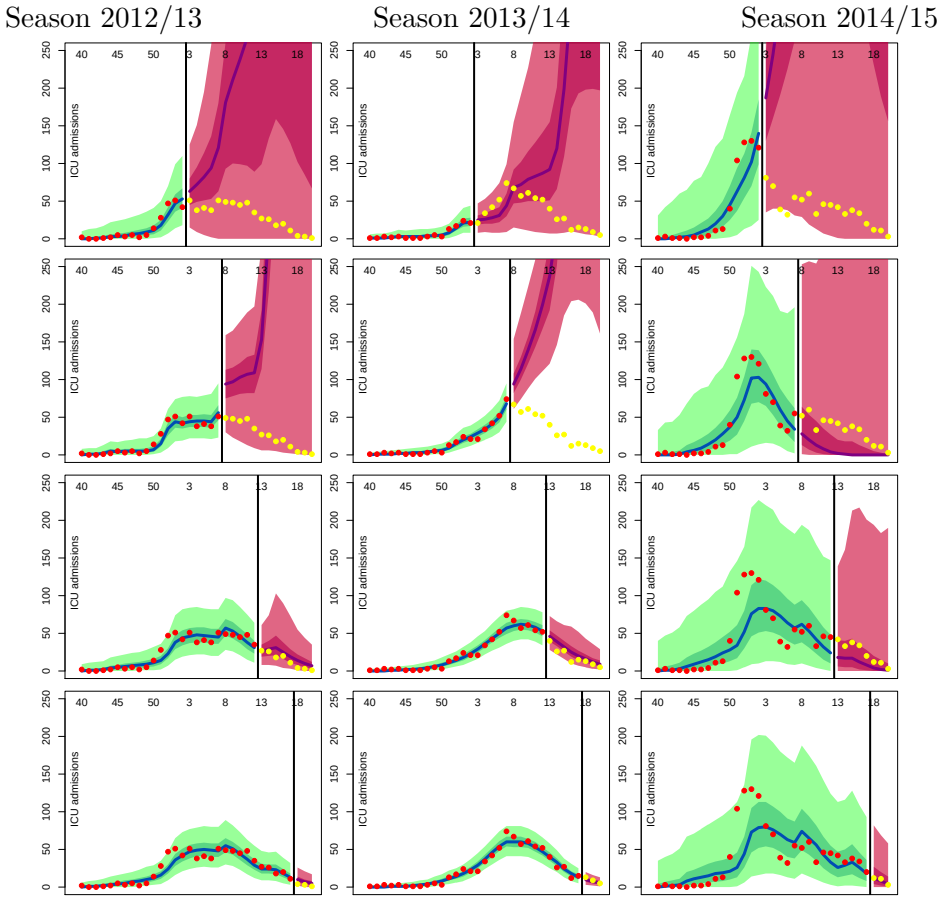


Figure C.11: Flat scenario - The black line displays the analysis time; the blue line and green shaded area represent median, quartile (dark green) and 95% CrIs (light green) of the posterior predictive distribution for the training weeks. The pink area displays posterior quartiles (deep pink) and 95% CrIs (light pink) for the predicted future observations, and the purple line displays the median; the red dots are the training data and the yellow dots are the objects of prediction.

C.3 Simulation of a pandemic

The simulated data on the number of hospital admissions during an epidemic described in the main text are plotted in Figure C.12.

C.3.1 Results on the full datasets

Results with uniform priors

3 independent chains of 1.1 million iterations of the blocked MH algorithm were run. Of these, the first 100,000 iterations were used adaptively to tune the algorithm (i.e. to estimate the

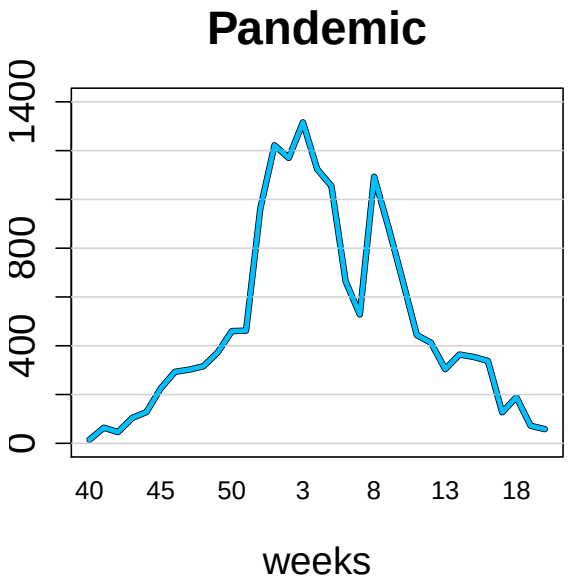


Figure C.12: Simulated data in the case of a pandemic.

mixing parameter ν) and the following 200,000 were discarded as burn-in period. Moreover, a thinning factor of 100 was used, saving only one iteration every 100.

Figure C.13 displays the chains of the 6 parameters of the system. Despite all the parameters showing convergence, the chains of some of them (namely π and ψ) are moving over a uniform distribution and their chains are particularly correlated. This shows that, despite the numbers being higher in the case of a pandemic, the identifiability issue is not solved

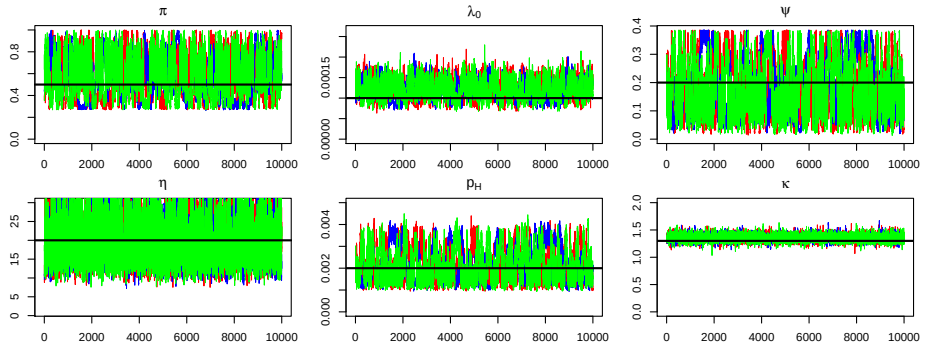


Figure C.13: Flat scenario - The three independent chains are plotted here with three different colours (red, blue and green). The parameters are: the initial susceptibility π, the initial infection rate λ_0, the exponential growth rate ψ, the over-dispersion parameter η, the probability of hospital admission given infection p_H, and the scaling parameter κ. The black horizontal lines denote the values used to simulate the dataset.

The plots of the prior and the posterior distributions of all the parameters are reported in figure C.14. Figure C.15 contains the plots of the prior and posterior distributions of the other quantities of interest.

Table C.7 reports median and 95% CrIs of all the parameters.

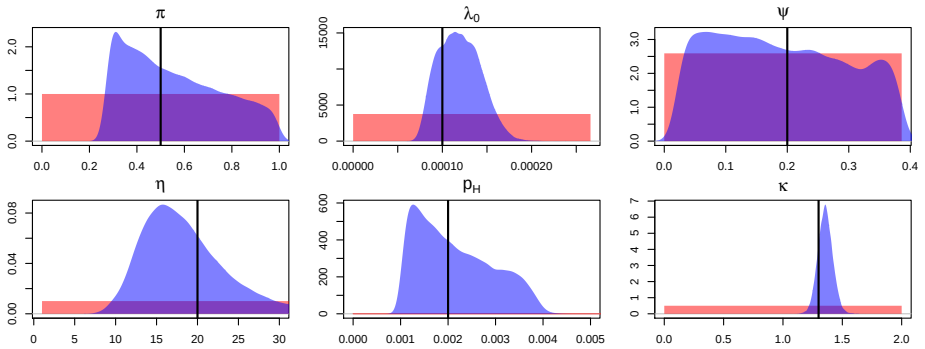


Figure C.14: Flat scenario - Prior (red) and posterior (blue) distributions of the parameters. The parameters are: π, λ_0, ψ, η, p_H and κ. The black vertical lines denote the values used to simulate the dataset.

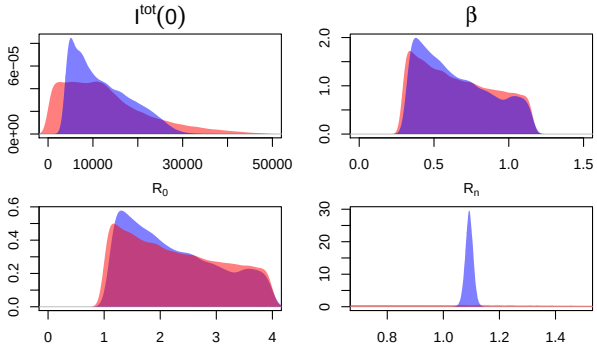


Figure C.15: Flat scenario - Prior (red) and posterior (blue) distributions of: the total number of initial infectious I_0^{tot}, the transmission rate β, the basic reproduction number R_0, and the effective reproduction number R_n.

	Median (95% CrI)
π	0.523 (0.282 - 0.961)
λ_0	1.183 (0.804 - 1.662)
ψ	0.182 (0.028 - 0.375)
η	17.665 (10.677 - 32.4)
p_H	1.996 (1.072 - 3.747)
κ	1.356 (1.238 - 1.481)
I_0^{tot}	10721 (4104 - 25973.6)
β	0.605 (0.329 - 1.123)
R_0	2.087 (1.136 - 3.873)
R_n	1.092 (1.064 - 1.12)

Table C.7: Flat scenario - Posterior medians and 95% CrIs form the retrospective analysis of the hospital admissions.

Results with informative priors

Figure C.16 displays the chains of the 6 parameters of the system from the analysis with informative priors. Despite all the parameters showing convergence, the chains of some of them (namely π and ψ) are still very correlated.

The plots of the prior and the posterior distributions of all the parameters are reported in

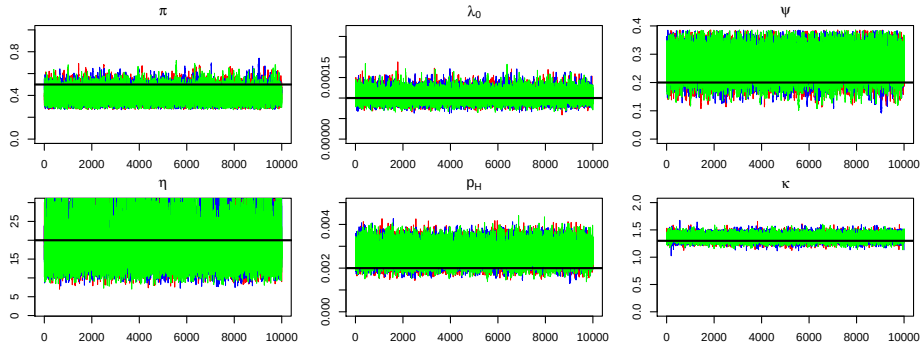


Figure C.16: Informative scenario - The three independent chains are plotted here with three different colours (red, blue and green). The parameters are: π, λ_0, ψ, η, p_H and κ. The black vertical lines denote the values used to simulate the dataset.

figure C.17. Figure C.18 contains the plots of the prior and posterior distributions of the other quantities of interest.

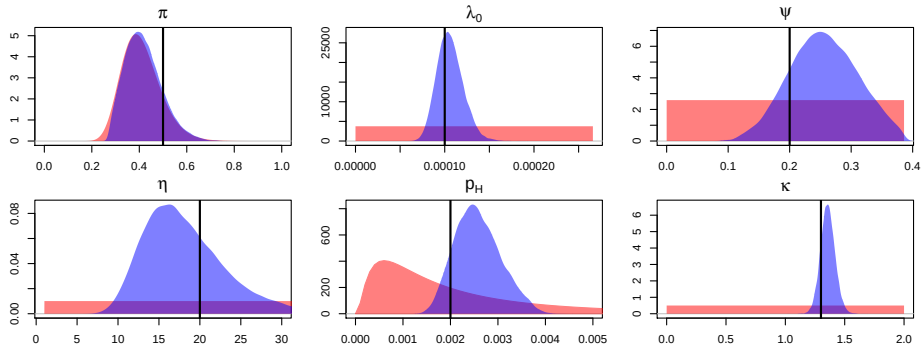


Figure C.17: Informative scenario - Prior (red) and posterior (blue) distributions of the parameters. The parameters are: π, λ_0, ψ, η, p_H and κ. The black vertical lines denote the values used to simulate the dataset.

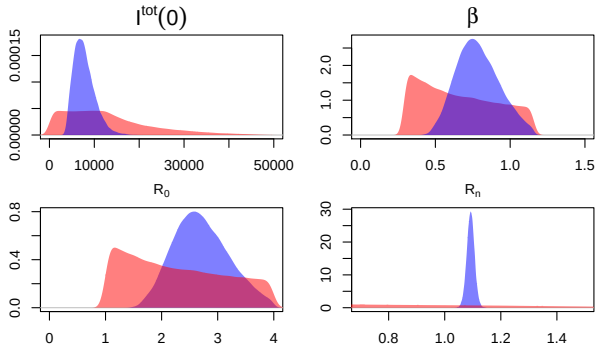


Figure C.18: Informative scenario - Prior (red) and posterior (blue) distributions of: the total number of initial infectious I_0^{tot}, the basic transmission rate β, the basic reproduction number R_0, and the effective reproduction number R_n.

Table C.8 reports median and 95% CrIs of all the parameters.

	Median (95% CrI)
π	0.409 (0.293 - 0.591)
λ_0	1.05 (0.802 - 1.37)
ψ	0.254 (0.148 - 0.362)
η	17.462 (10.574 - 31.958)
p_H	2.539 (1.744 - 3.596)
κ	1.357 (1.24 - 1.482)
I_0^{tot}	7412.1 (4260.8 - 13170)
β	0.774 (0.535 - 1.082)
R_0	2.671 (1.845 - 3.732)
R_n	1.092 (1.065 - 1.12)

Table C.8: Informative scenario - Posterior medians and 95% CrIs form the retrospective analysis of the Hospital admissions.

C.3.2 Results on the datasets updated every five weeks

The results on the predictive performance of the model is described here. Informative prior distributions are assumed on the parameters π and p_H.
Posterior predictive distributions are reported in Figure C.19. On the one hand, the precision of the predictions has increased a lot, making the results more useful; on the other hand, predictions could be precise and wrong, as shown in the second panel, where the future number of hospital admissions is underestimated.

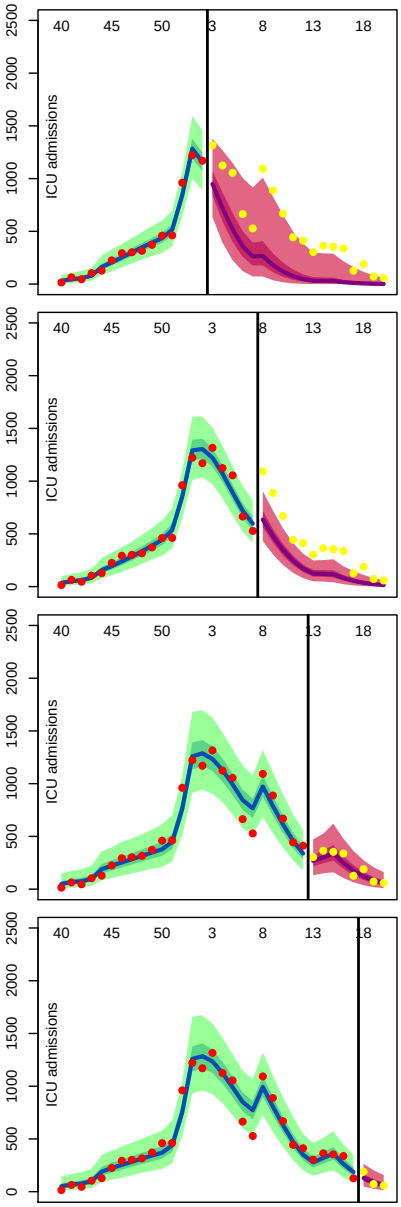


Figure C.19: Informative scenario - The black line displays the analysis time; the blue line and green shaded area represent median, quartile (dark green) and 95% CrIs (light green) of the posterior predictive distribution for the training weeks. The pink area displays posterior quartiles (deep pink) and 95% CrIs (light pink) for the predicted future observations, and the purple line displays the median; the red dots are the training data and the yellow dots are the objects of observation.

Appendix D

Appendix to Chapter 4

D.1 Monte Carlo methods

This appendix contains the foundations of many simulation methods on which sequential Monte Carlo (SMC) algorithms are based.

Vanilla Monte Carlo

Let X be a random variable (r.v.) with density function $f(x)$, called the *target density*. The goal is to compute the expected value $A(\phi)$ of a function $\phi(x)$ with respect to the distribution of X, i.e.:

$$A(\phi) = \mathbb{E}_f\left[\phi(x)\right] = \int_X \phi(x) f(x) \mathrm{d}x \tag{D.1}$$

Under the assumption that it is possible to simulate directly from $f(x)$, the integral in (D.1) can be approximated by:

$$A(\phi) = \mathbb{E}_f\left[\phi(x)\right] \approx \frac{1}{N} \sum_{n=1}^{N} \phi(x^{(n)})$$

where $\left\{x^{(n)}\right\}_{n=1}^{N}$ is a sample of size N from $f(x)$. This sample can also be used to obtain a Monte Carlo (MC) approximation of the target distribution:

$$\widehat{f}^{\mathrm{MC}}(x) = \frac{1}{N} \sum_{n=1}^{N} \delta_{x^{(n)}}(x) \tag{D.2}$$

where $\delta_a(x)$ is a Dirac point-mass in a

$$\delta_a(x) = \begin{cases} 1 \text{ if } x = a \\ 0 \text{ otherwise} \end{cases}$$

so that

$$\widehat{A}^{\mathrm{MC}}(\phi) = \int_X \phi(x) \sum_{n=1}^{N} \delta_{x^{(n)}}(x) \mathrm{d}x = \frac{1}{N} \sum_{n=1}^{N} \phi(x^{(n)})$$

Unfortunately, in many of the cases considered below it would be impossible to sample directly from the target distribution. A common alternative scenario occurs when $f(x)$ can be evaluated point-wise (up to proportionality), then importance sampling might be a feasible solution.

Importance sampling

Importance sampling (Marshall, 1954) explores the target distribution $f(x)$ by introducing a new distribution $q(x)$, called the *importance* or *proposal distribution* which is easy to sample from and whose support must contain the support of the target density.

Integral D.1 can hence be re-written as:

$$\begin{aligned} A(\phi) = \mathbb{E}_f\left[\phi(x)\right] &= \int_X \phi(x) f(x) \\ &= \int_X \phi(x) \frac{f(x)}{q(x)} q(x) \mathrm{d}x \\ &= \int_X \phi(x) \omega(x) q(x) \mathrm{d}x \\ &= \mathbb{E}_q\left[\phi(x)\omega(x)\right] \end{aligned}$$

where $\omega(x) = \frac{f(x)}{q(x)}$ is called the *weight function*. Since a sample $\left\{x^{(n)}\right\}_{n=1}^{N}$ can be drawn from the importance distribution $q(x)$, the integral can be approximated by vanilla MC integration:

$$\widehat{A}^{\mathrm{IS}}(\phi) = \frac{1}{N} \sum_{n=1}^{N} \phi(x^{(n)}) \omega(x^{(n)})$$

where $\omega(x^{(n)})$ is the *observed weight*, denoted $w^{(n)}$.

Self-normalised importance sampling deals with the situation when the target distribution is only known up to a normalising constant: $f(x) = \frac{\widetilde{f}(x)}{Z_f}$. In this case the non-normalised weight function can be defined as:

$$\widetilde{\omega}(x) = \frac{\widetilde{f}(x)}{q(x)}$$

and computed at the sampled values $\widetilde{\omega}(x^{(n)}) = \widetilde{w}^{(n)}$, which can then be normalised:

$$w^{(n)} = \frac{\widetilde{w}^{(n)}}{\sum_{i=1}^{N} \widetilde{w}^{(n)}}$$

and plugged in Equation D.1. Similarly to Equation D.2, an approximation of the target distribution can be obtained by the weighted samples as follows:

$$\widehat{f}^{\mathrm{IS}}(x) = \sum_{n=1}^{N} w^{(n)} \delta_{x^{(n)}}(x) \tag{D.3}$$

The pseudo-code to obtain N samples of values $x^{(n)}$, often called *particles*, and N weights $w^{(n)}$ that approximate the target distribution $f(x)$ is reported in Algorithm 9.

Result: $\left\{x^{(n)}, w^{(n)}\right\}_{n=1}^{N}$
Input: N, $\widetilde{f}(x)$, $q(x)$
for $n = 1, \dots N$ **do**
  sample
  $x^{(n)} \sim q(x)$

  compute the importance weights
  $\widetilde{w}^{(n)} = \frac{\widetilde{f}(x^{(n)})}{q(x^{(n)})}$
end
normalize the importance weights
$w^{(n)} = \frac{\widetilde{w}^{(n)}}{\sum_{i=1}^{N} \widetilde{w}^{(n)}} \qquad \forall n = 1, \dots, N$

Algorithm 9: Self-normalised Importance Sampling

The accuracy of importance-sampler-based approximations to Equation D.1 increases with N and, more importantly, with appropriate choices of the distribution $q(x)$: the greater the similarity between target and importance distribution, the better the approximation is. Finding suitable importance distributions, and approximating the target distribution via importance sampling, becomes more difficult with increasing dimensionality.

D.2 Simulating a dynamic system

The Severity process assumed in 4.4 is just a set of time non-homogeneous Poisson processes, where the event rates are often dependent on the number of prevalent individuals in a compartment. This is highly studied in chemical kinetics. In this section two algorithms are listed; these can be used to simulate data arising from a dynamic process that can be seen as a time non-homogeneous Poisson process with rates dependent on the prevalence of the state.

Standard Gillespie Algorithm

Consider a system X where v chemical reactions might happen. This system is assumed to behave as a Markov jump process. The hazard of each reaction $h_i, \forall i = 1, \dots v$, depends on a constant rate c_i and the current state of the process x; i.e. $h_i = h(c_i, x)$. To simulate this process Gillespie proposed the following algorithm (Gillespie, 1976).

1. Initialise the system at $t = 0$ with constants rates $c_1, c_2, \dots, c_v$ and initial numbers of molecules for each species, $x_1, x_2, \dots, x_u$;
2. For each $i = 1, 2, \dots, v$, calculate $h_i(x, c_i)$ based on the current state x;
3. Calculate $h_0(x, c) = \sum_{i=1}^{v} h_i(x, c_i)$, the combined reaction hazard;
4. Simulate time to next event, t', as an $Exp(h_0(x, c))$ random quantity;
5. Put $t := t + t'$;
6. Simulate the reaction index, j, as a discrete random quantity with probabilities $\frac{h_i(x,c_i)}{h_0(x,c)}, i = 1, 2, \dots, v$;

7. Update x according to reaction j;
8. Output x and t;
9. If $t < T_{max}$, return to step 2.

This algorithm gives exact simulation of the Markov Jump process.

τ-leap Gillespie Algorithm

The standard Gillespie algorithm is very slow when the number of reactions that happen is high. In 2001 he proposed an approximated version of the algorithm that relies on the *leap condition.*

Definition 5. *Leap Condition: Require τ to be small enough that the change in the state during $[t, t+\tau]$ will be so slight that no propensity function (i.e. hazard) will suffer an appreciable (i.e., macroscopically non-infinitesimal) change in its value.*

If the Leap condition holds the simulation of the standard Gillespie algorithm can be approximated by the following (Gillespie, 2001):

1. Initialise the system at $t = 0$ with rate constants $c_1, c_2, \ldots, c_v$ and initial numbers of molecules for each species, $x_1, x_2, \ldots, x_u$;
2. For each $i = 1, 2, \ldots, v$, calculate $h_i(x, c_i)$ based on the current state, x;
3. Simulate $R_i(t; t+\tau)$, the number of reactions of type i that happen in the interval $(t; t+\tau)$ from: $R_i(t; t+\tau) \sim$ Poisson $(\tau h_i(x, c_i))$;
4. Update $t = t + \tau$ and x at $t + \tau$ accordingly;
5. If $t + \tau < T_{max}$, return to step 2.

This simulation scheme is significantly more efficient than the classical Gillespie algorithm, above all when the rate of any reaction is high.

D.3 Code for the Bootstrap Particle Filter in Section 4.5

```
BPF <- function(data,            # list of elements, each containing
                                 # a time series of data
                delta.dsets,     # vector of same length as data, each elements
                                 # says the timescale in days of
                                 # the data used (eg 7=weekly data)
                Xts.dsets,       # vector of same length as data, each elements
                                 # contains a character vector stating the components
                                 # of X on which the data are centered
                Theta,           # list of paramters, thhe first 15 elemnts are
                                 # the named Xproc paramteres and the last two
                                 # are the Yobs paramter: a list of verctor with
```

```
                                # the detection probabilities and a vector
                                # containing the overdispersion of each dataset
                NP,             # number of particles
                dt,             # time step at which to run the Xstate process
                X0,             # state in 0: a list with as first element a
                                # vector of length 3 containing the number of
                                # people in each SEI state at 0 and all the other
                                # elements contain the prevalence of the severity
                                # states at 0
                N               # total population size
                ){
  ...

    for (d in 1:length(data)){
      if(round(tday[t], 5) %in% DayData[[d]]){
        # select the ts for mu
        tfm   <-  which(tday>DayData[[d]][tday[t]/delta.dsets[d]]&
                          tday<=DayData[[d]][tday[t]/delta.dsets[d]+1])
        if (length(tfm)>1){
          xtdt  <- apply(Xpar[,tfm,
                              Xts.dsets[[d]]]
                         ,MARGIN = c(1,2), sum)
          muts  <- apply(xtdt, MARGIN = 1, sum)
        }else if (length(tfm)==1){
          muts  <- apply(as.matrix(Xpar[, tfm, Xts.dsets[[d]]]), MARGIN = 1, sum)
        }
        rts   <- ifelse((muts*Theta$det.dsets[[d]][tday[t]/delta.dsets[d]])/
                        (Theta$eta.dsets[d]-1)>0,
                        (muts*Theta$det.dsets[[d]][tday[t]/delta.dsets[d]])/
                        (Theta$eta.dsets[d]-1),
                        1)
        wts   <- dnbinom(data[[d]][tday[t]/delta.dsets[d]],
                        mu=muts, size=rts, log=T)
        ...
      }
    }
  return(list("lL"=LL,
              "Rtraj"=Xpar))
}
```

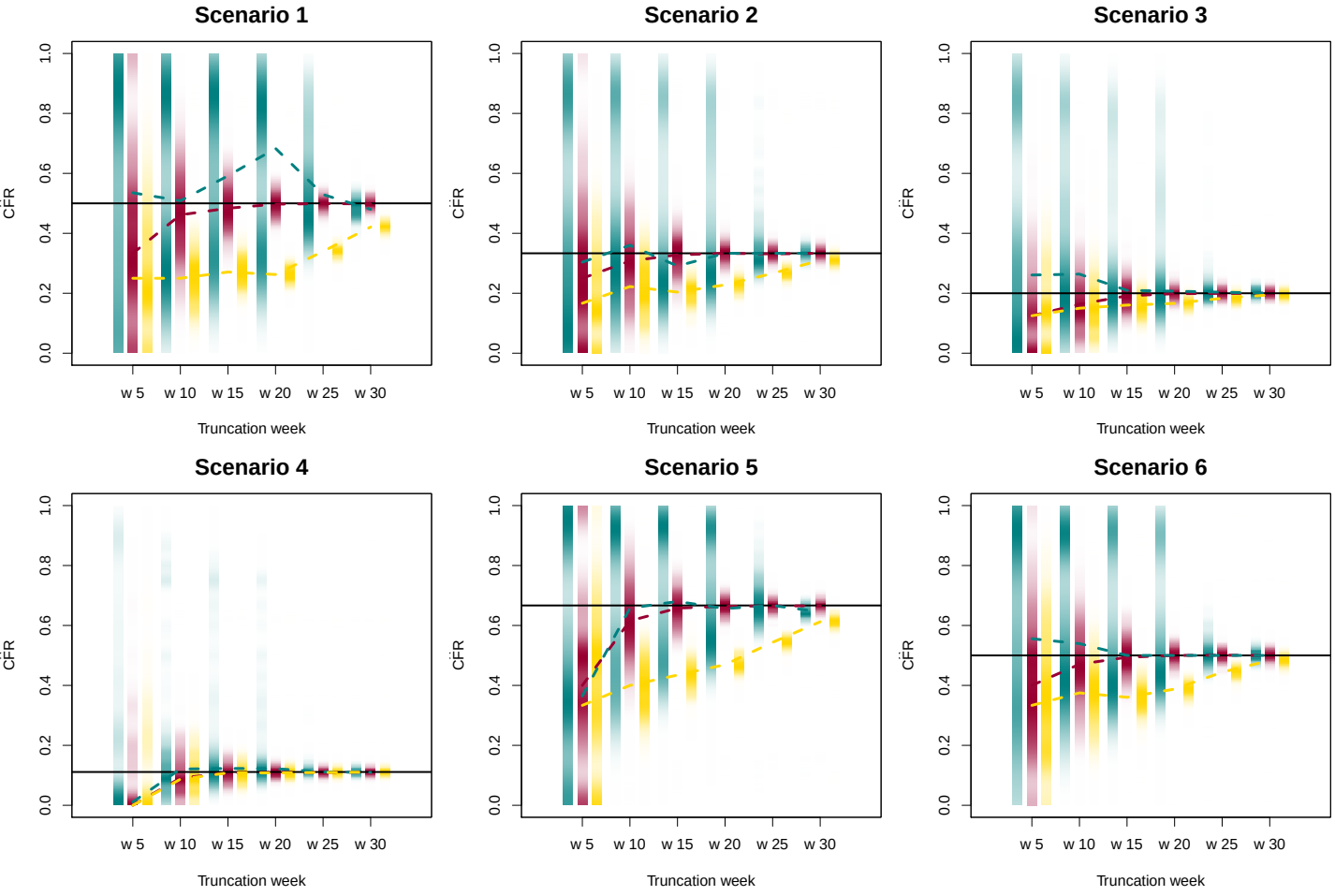
D.4 Simulation study

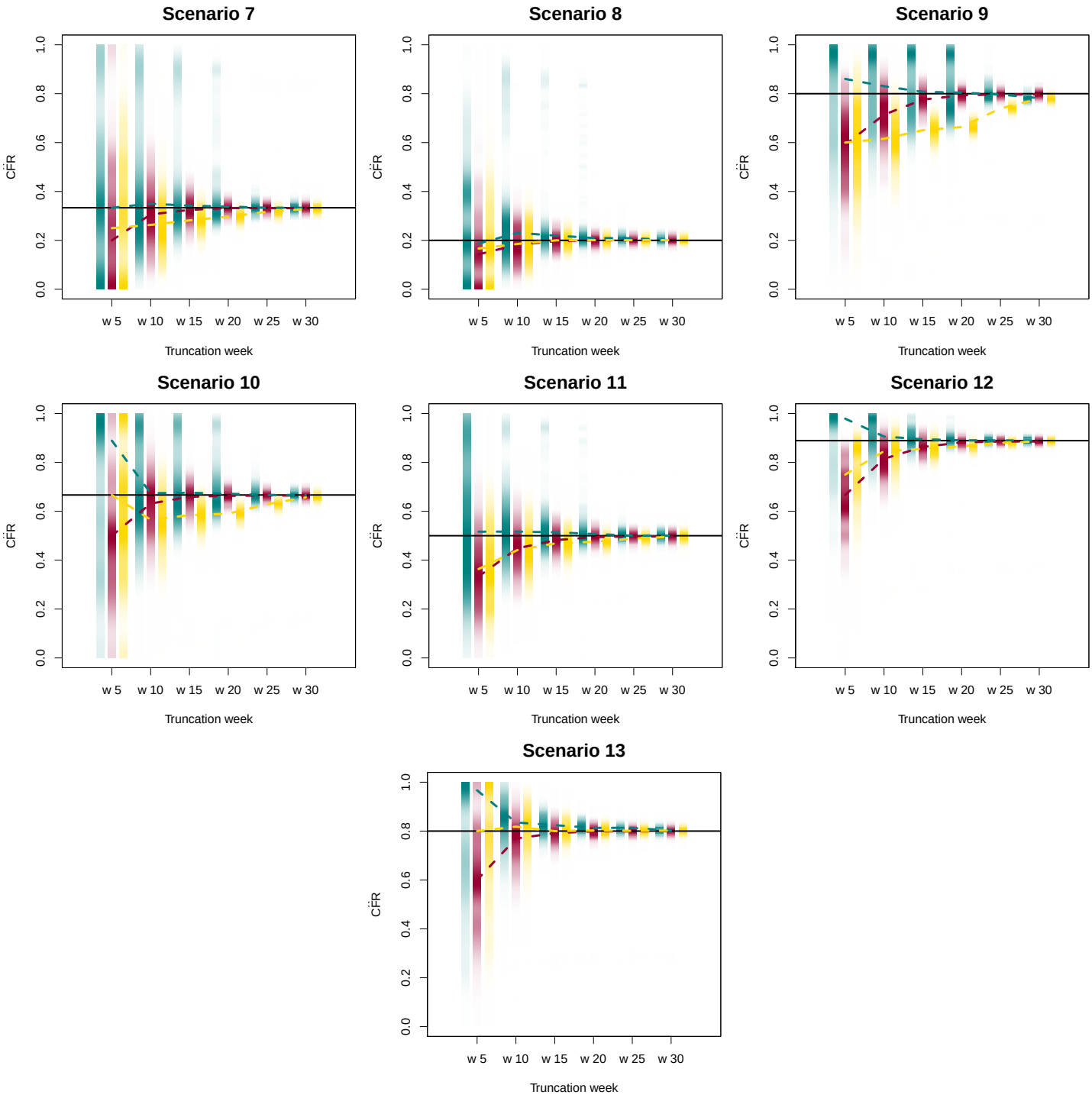
The scenarios considered are listed in Table D.1

	pD	pR	β_1	β_2
Scenario 1	0.1	0.1	0.2	0.5
Scenario 2	0.1	0.2	0.3	0.333
Scenario 3	0.1	0.4	0.5	0.2
Scenario 4	0.1	0.8	0.9	0.111
Scenario 5	0.2	0.1	0.3	0.667
Scenario 6	0.2	0.2	0.4	0.5
Scenario 7	0.2	0.4	0.6	0.333
Scenario 8	0.2	0.8	1	0.2
Scenario 9	0.4	0.1	0.5	0.8
Scenario 10	0.4	0.2	0.6	0.667
Scenario 11	0.4	0.4	0.8	0.5
Scenario 12	0.8	0.1	0.9	0.889
Scenario 13	0.8	0.2	1	0.8

Table D.1: Values assumed to simulate the scenarios.

The results for all the simulated scenarios are displayed below.




Scenario 7
Scenario 8
Scenario 9
Scenario 10
Scenario 11
Scenario 12
Scenario 13
CFR
0.0
0.2
0.4
0.6
0.8
1.0
w 5
w 10
w 15
w 20
w 25
w 30
Truncation week

Appendix E

Appendix to Chapter 5

E.1 Independent versus dependent likelihood: toy example

A simple model to describe how the observation of the mandatory and the sentinel UK Severe Influenza Surveillance System (USISS) schemes can be generated is exposed here. This is a simplistic model: the timely-aspect of the data is not considered, assuming that only the total number of cases detected by the sentinel and by the mandatory scheme is to be modelled.

Denote: by N the total cases; by H the hospitalizations; by H^{S} the hospitalizations in a sentinel trust; by I^{S} the Intensive Care Unit (ICU) admissions in a sentinel trust; by I^{N} the ICU admissions in non-sentinel trusts; by y^{M} the cases reported to the mandatory scheme; and by y^{S} the cases reported in the sentinel scheme. These quantities can be expressed as a chain of Binomial random variable (r.v.) with parameters: $p_{H|E}$, the probability of hospitalization given infection; p_{Sent}, the probability of being in a sentinel trust given hospitalization; $p_{I|H}$, the probability of ICU admission given hospitalization; d_{S}, the probability of being detected in the sentinel scheme given hospitalization; and d_{M} the probability of being detected in the mandatory scheme given ICU admission. The r.v.s describing the dynamics are:

$$\begin{aligned}
(a) \qquad & H \sim \mathrm{Binom}(N, p_{H|E}) \\
(b) \qquad & H^{\mathrm{S}} \sim \mathrm{Binom}(H, p_{Sent}) \\
(cI) \qquad & I^{\mathrm{S}} \sim \mathrm{Binom}(H^{\mathrm{S}}, p_{I|H}) \\
(cII) \qquad & I^{\mathrm{N}} \sim \mathrm{Binom}(H - H^{\mathrm{S}}, p_{I|H}) \\
(d) \qquad & y^{\mathrm{S}} \sim \mathrm{Binom}(H^{\mathrm{S}}, d_{\mathrm{S}}) \\
(e) \qquad & y^{\mathrm{M}} \sim \mathrm{Binom}(I^{\mathrm{S}} + I^{\mathrm{N}}, d_{\mathrm{M}})
\end{aligned} \tag{E.1}$$

The two data y^{M} and y^{S} overlap. Let $y^{\mathrm{M}\bar{\mathrm{S}}}$ be the number of people registered in the mandatory but not in the sentinel scheme, $y^{\bar{\mathrm{M}}\mathrm{S}}$ the number of people registered in the sentinel but not in the mandatory scheme and y^{MS} the overlap (i.e. the people registered in both).

A graphical representation of the disjoint sets that sum up to the number of cases N is reported in Figure E.1. Where the red compartment represent the people detected only by the mandatory scheme $y^{\mathrm{M}\bar{\mathrm{S}}}$; the yellow compartment represent the people detected only by the sentinel scheme $y^{\bar{\mathrm{M}}\mathrm{S}}$; the blue compartment represent the people detected by both the mandatory and the sentinel scheme y^{MS}; and the grey compartment represent the people undetected. The

probabilities of these disjoint sets are:

$$\begin{aligned}
p_1 =& p_{H|E} \cdot p_{Sent} \cdot p_{I|H} \cdot d_{\text{M}} \cdot d_{\text{S}} \\
p_2 =& p_{H|E} \cdot p_{Sent} \cdot d_{\text{S}} \cdot \left[(1 - p_{I|H}) + p_{I|H} \cdot (1 - d_{\text{M}})\right] \\
p_3 =& p_{H|E} \cdot d_{\text{M}} \cdot p_{I|H} \cdot \left[(1 - p_{Sent}) + p_{Sent} \cdot (1 - d_{\text{S}})\right] \\
p_4 =& 1 - p_1 - p_2 - p_3 = \left[1 - p_{H|E}\right] \\
&+ p_{H|E} \cdot (1 - p_{Sent}) \cdot \left[(1 - p_{I|H}) + p_{I|H} \cdot (1 - d_{\text{M}})\right] \\
&+ p_{H|E} \cdot (p_{Sent}) \cdot (1 - d_{\text{S}}) \cdot \left[(1 - p_{I|H}) + p_{I|H} \cdot (1 - d_{\text{M}})\right]
\end{aligned} \tag{E.2}$$

And therefore the disjoint sets are distributed according to a Multinomial r.v.:

$$\left(y^{\text{MS}}, y^{\overline{\text{M}}\text{S}}, y^{\text{M}\overline{S}}, U\right) \sim \text{Multinom}(N, p_1, p_2, p_3, p_4) \tag{E.3}$$

Figure E.2 relates these disjoint quantities to the observed data y^{M} and y^{S}.

The two data y^{M} and y^{S} are, marginally, a Binomial sample of the N cases, and therefore, under the hypothesis of independence, their likelihood would be the product of the two Binomial densities:

$$\begin{aligned}
y^{\text{S}} &\sim \text{Binom}\left(N; p^{\text{S}} = (p_1 + p_2)\right) \\
y^{\text{M}} &\sim \text{Binom}\left(N; p^{\text{M}} = (p_1 + p_3)\right)
\end{aligned} \tag{E.4}$$

which would lead to a likelihood of the following form:

$$p(y^{\text{S}}, y^{\text{M}} | N, p_1, p_2, p_3) = \binom{N}{y^{\text{S}}} \left[p^{\text{S}}\right]^{y^{\text{S}}} \left[1 - p^{\text{S}}\right]^{(N - y^{\text{S}})} \cdot \binom{N}{y^{\text{M}}} \left[p^{\text{M}}\right]^{y^{\text{M}}} \left[1 - p^{\text{M}}\right]^{(N - y^{\text{M}})} \tag{E.5}$$

If instead, the data are allowed to be dependent, and the joint Multinomial distribution of Equation E.3 is adopted, the likelihood takes the form of a convolution over the unknown overlap between the two datasets as follows:

$$\begin{aligned}
p(y^{\text{S}}, y^{\text{M}} | N, p_1, p_2, p_3) =& \sum_{m=0}^{min(y^{\text{S}}, y^{\text{M}})} p \left(\begin{array}{c|cc} y^{\text{S}} - m & & \frac{p_2}{1-p_1} \\ y^{\text{M}} - m & N - m; & \frac{p_3}{1-p_1} \\ N - y^{\text{S}} - y^{\text{M}} + m & & \frac{(1-p_3-p_2-p_1)}{1-p_1} \end{array} \right) \\
&\times p(m | N, p_1) \\
=& \sum_{m=0}^{min(y^{\text{S}}, Y^{\text{M}})} \frac{(N-m)!}{(y^{\text{S}} - m)!(y^{\text{M}} - m)!(N - y^{\text{S}} - y^{\text{M}} + m)!} \times \\
&\left(\frac{p_2}{1 - p_1}\right)^{(y^{\text{S}} - m)} \left(\frac{p_3}{1 - p_1}\right)^{(y^{\text{M}} - m)} \left(\frac{(1 - p_3 - p_2 - p_1)}{1 - p_1}\right)^{(N - y^{\text{S}} - y^{\text{M}} + m)} \\
&\times \binom{N}{m} (p_1)^m (1 - p_1)^{(N-m)}
\end{aligned} \tag{E.6}$$

The two likelihoods reported in Equations E.5 and E.6 are not equal.

A very simple case of two simulated data is proposed below.

y^{M} and y^{M} have been generated with the parameter-set: $p_1 = 0.3$, $p_2 = 0.2$, $p_3 = 0.1$. The log-likelihood curve for each parameter, conditional on the others being equal to the generating parameters and given the data y^{M} and y^{M}, is drawn below in Figure E.3.

The curves are nor the same, and therefore the inference drawn according to the independent Binomial likelihood is misspecified since it does not account for the overlap between datasets.

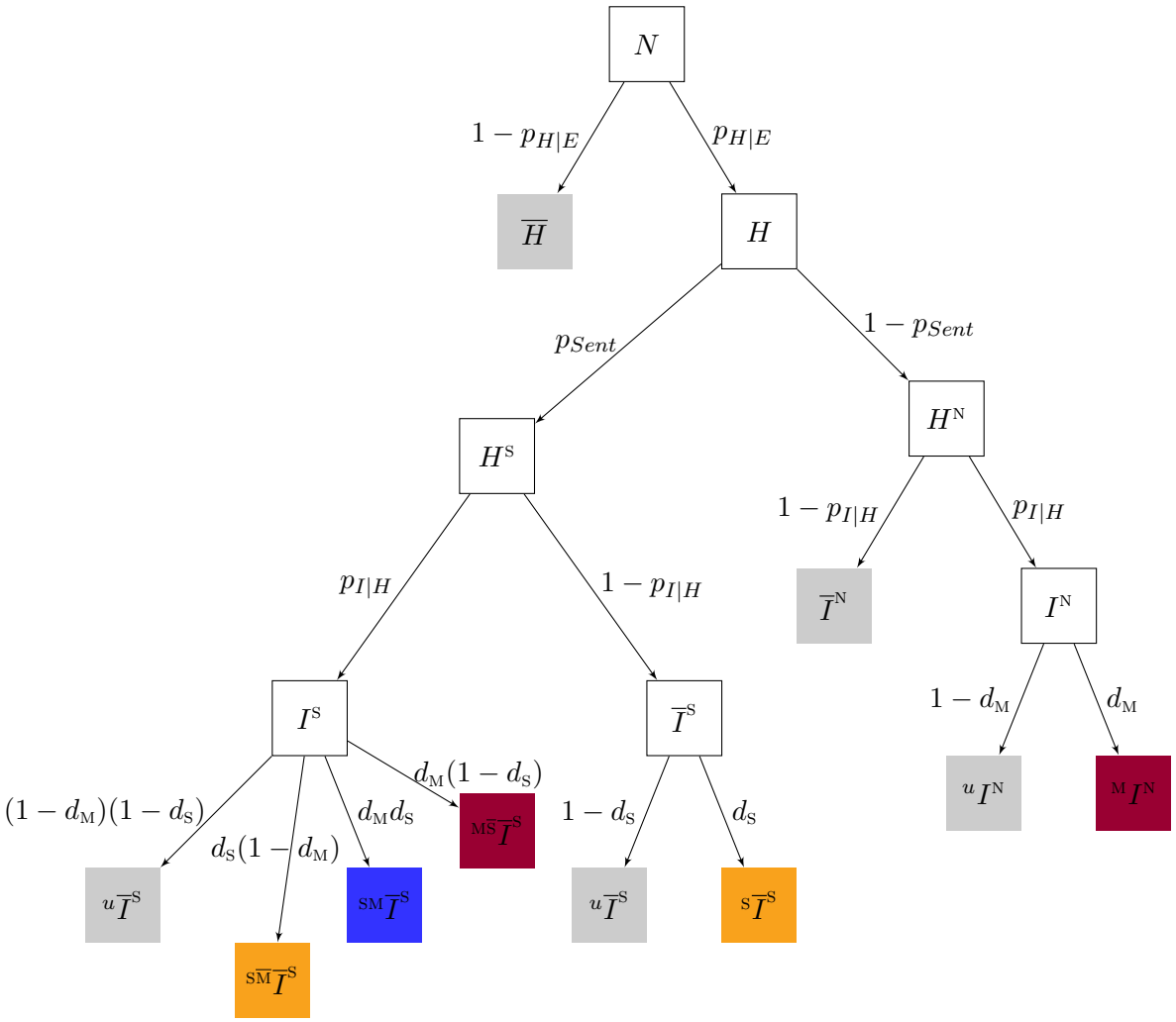


Figure E.1: Multinomial model. Where N, H, I are as defined above and $\overline{H}, \overline{I}$ are the infected-non-hospitalised and the hospitalised-non-ICU-admitted people respectively. The superscript following the compartment letter $(\star)^{\mathrm{S}}, (\star)^{\mathrm{N}}$ denotes the people admitted to Sentinel, or Non-sentinel trusts respectively. The superscript preceding the compartment letter denotes the scheme that detected those people: ${}^{\mathrm{SM}}(\star)$, detected by the mandatory and the sentinel scheme, ${}^{\mathrm{S\overline{M}}}(\star)$, detected by the sentinel but not by the mandatory scheme, etc.

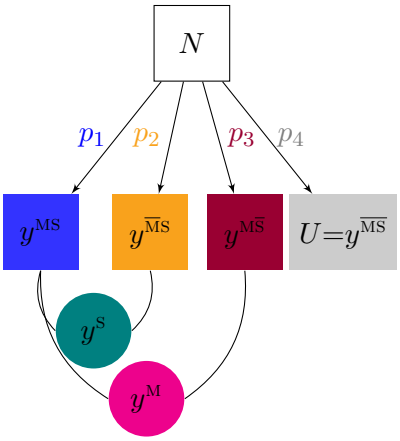


Figure E.2: Multinomial representation of the disjoint quantities composing the data.

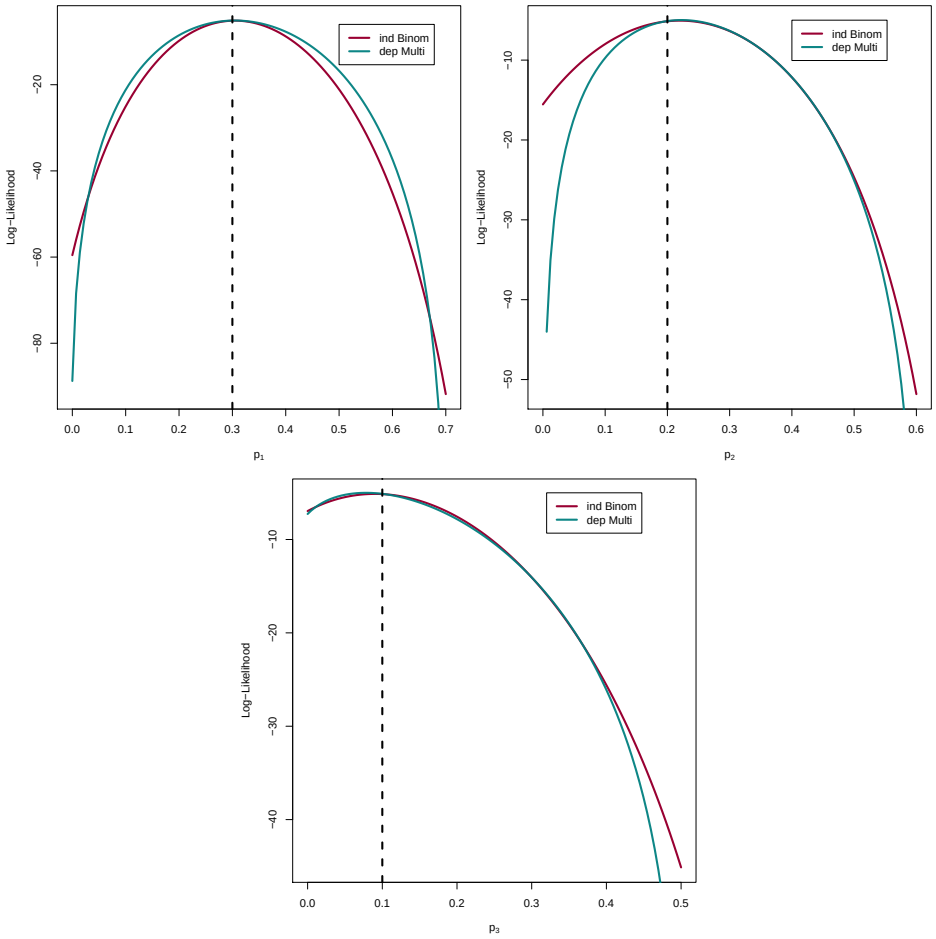


Figure E.3: Log-likelihood curve for the three parameters according to Equation E.5 (red) and E.6.

E.2 Specifics of the simulation study in Section 5.4

The simulation study is run on 500 datasets per scenario. On each dataset both an analysis using a product of two independent Poisson densities and an analysis accounting for dependence is run. In the latter a Monte Carlo within Metropolis (MCWM) algorithm is used: in the acceptance ratio of each step of the Markov chain Monte Carlo (MCMC) a set of N=2000 particles is used to approximate the likelihood.

The wrapping algorithm used to produce the sample is identical for dependent and independent case, except for the formulation of the likelihood. All the parameters of the system have been transformed to lie in $(-\infty; +\infty)$. A block-update Metropolis Hastings (MH) algorithm with multivariate-Normal proposal is used to explore the parameter space. The variance-covariance matrix of the proposal is estimated in previous Gibbs-sampling steps where independent univariate Normals is used as proposal distributions.

The Gibbs-sampling steps consists of 3 independent chains that run an adaptive phase of 10,000 iterations and a further sampling phase of 55,000 iterations, of which the first 5,000 are discarded as burn-in period.

The block-update phase consists of 3 chains that ran an adaptive phase of 20,000 iterations and a sampling phase of 110,000 iterations, of which the first 10,000 are discarded as burn-in period. A thinning factor of 100 is used so that the final approximation of the posterior

distribution consists of 3,000 samples.

E.3 Further results of the simulation study in Section 5.4

E.3.1 Increasing ${}^{0}\theta^{\mathrm{H}}$

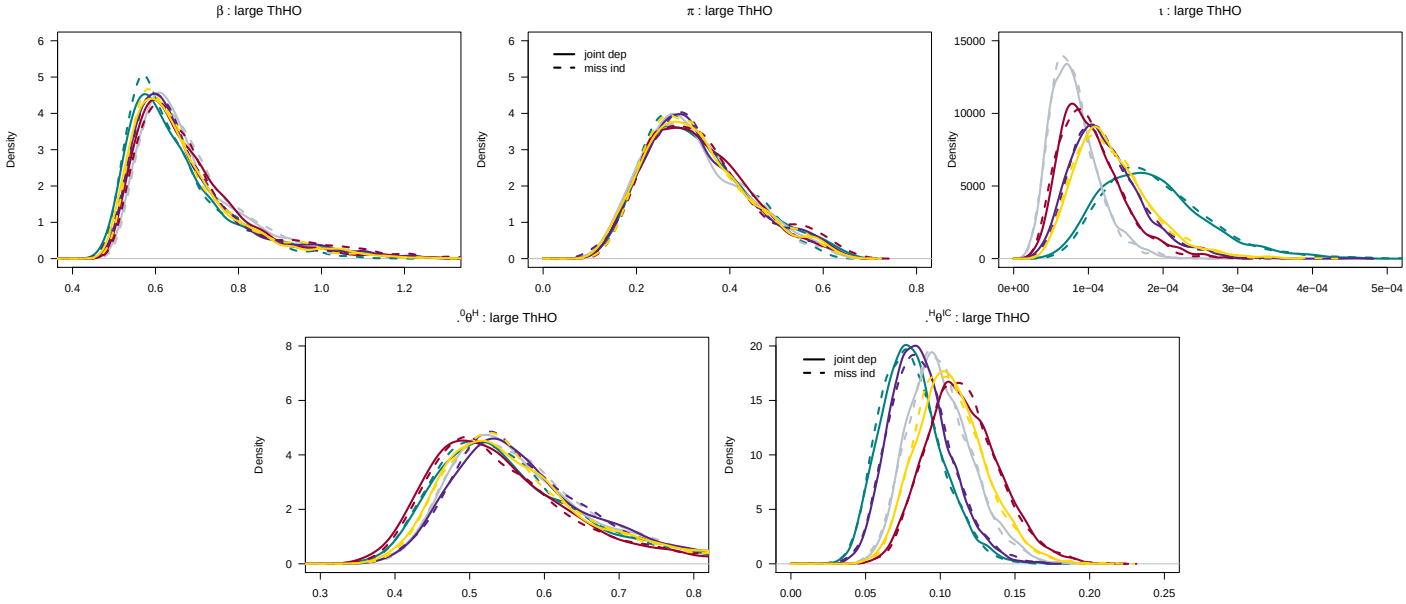


Figure E.4: Posterior distribution of the transmission and severity parameters from 5 datasets. The colour of the posterior density identifies the dataset analysed while dashed lines refer to results from the misspecified independent model and filled lines to results from the model approximating the joint dependent likelihood.

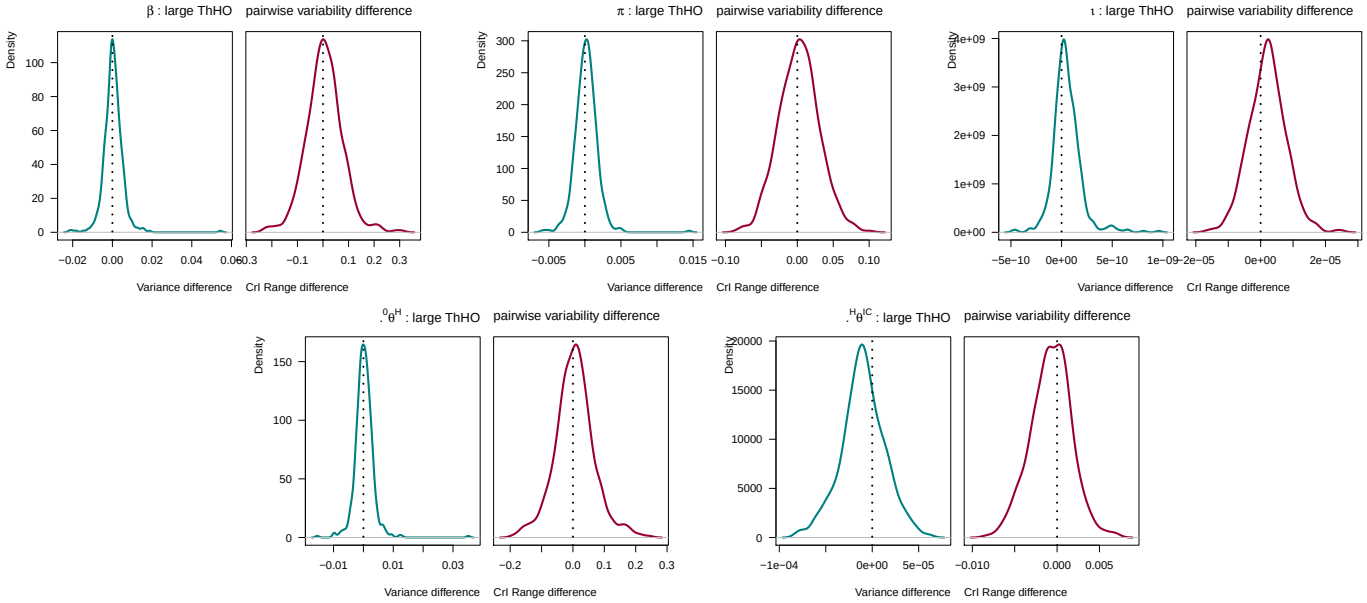


Figure E.5: Distribution of the pairwise difference in variance (green plots) and in 95% CrI length (red plots) of the posterior distribution of the transmission and severity parameters from all the datasets.

E.3.2 Increasing ${}^{\mathrm{H}}\theta^{\mathrm{IC}}$

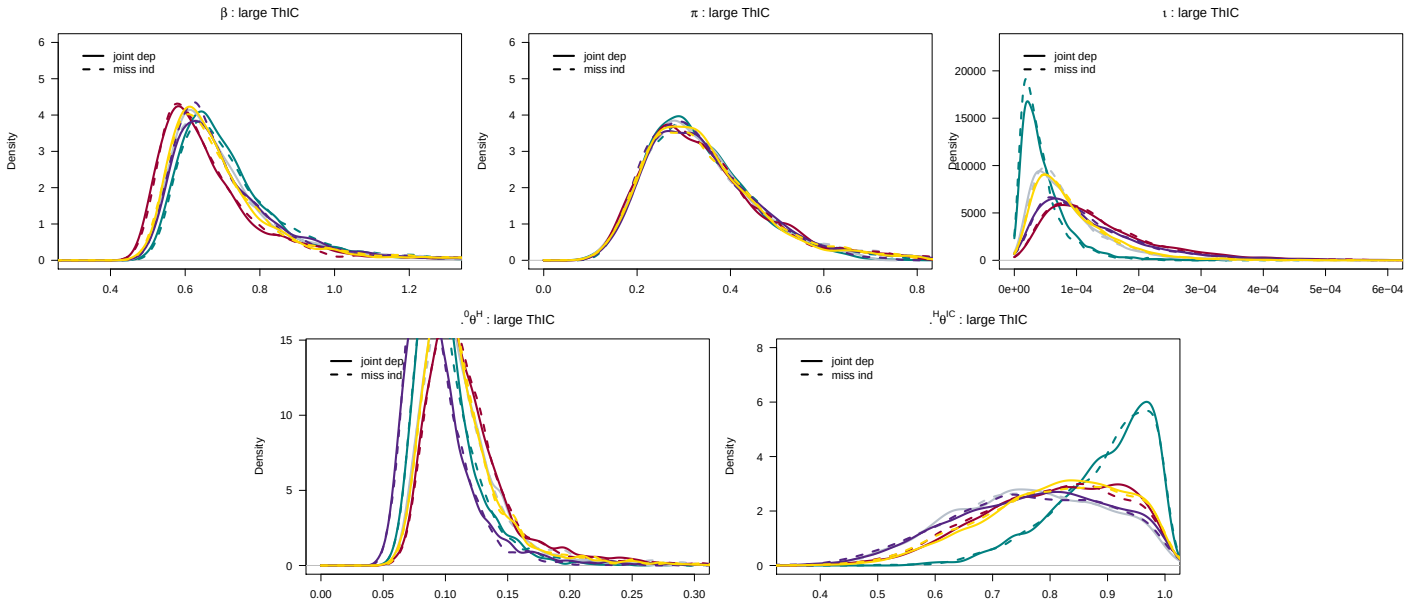


Figure E.6: Posterior distribution of the transmission and severity parameters from 5 datasets. The colour of the posterior density identifies the dataset analysed while dashed lines refer to results from the misspecified independent model and filled lines to results from the model approximating the joint dependent likelihood.

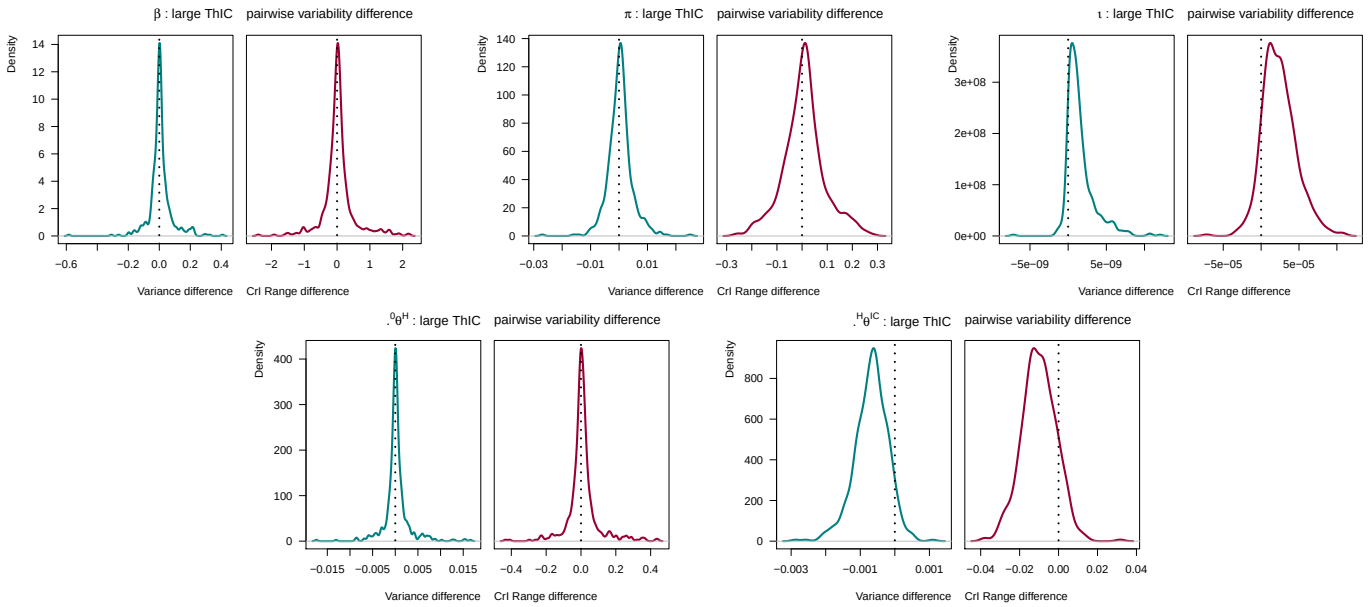


Figure E.7: Distribution of the pairwise difference in variance (green plots) and in 95% CrI length (red plots) of the posterior distribution of the transmission and severity parameters from all the datasets.

E.3.3 Increasing ζ^{H}

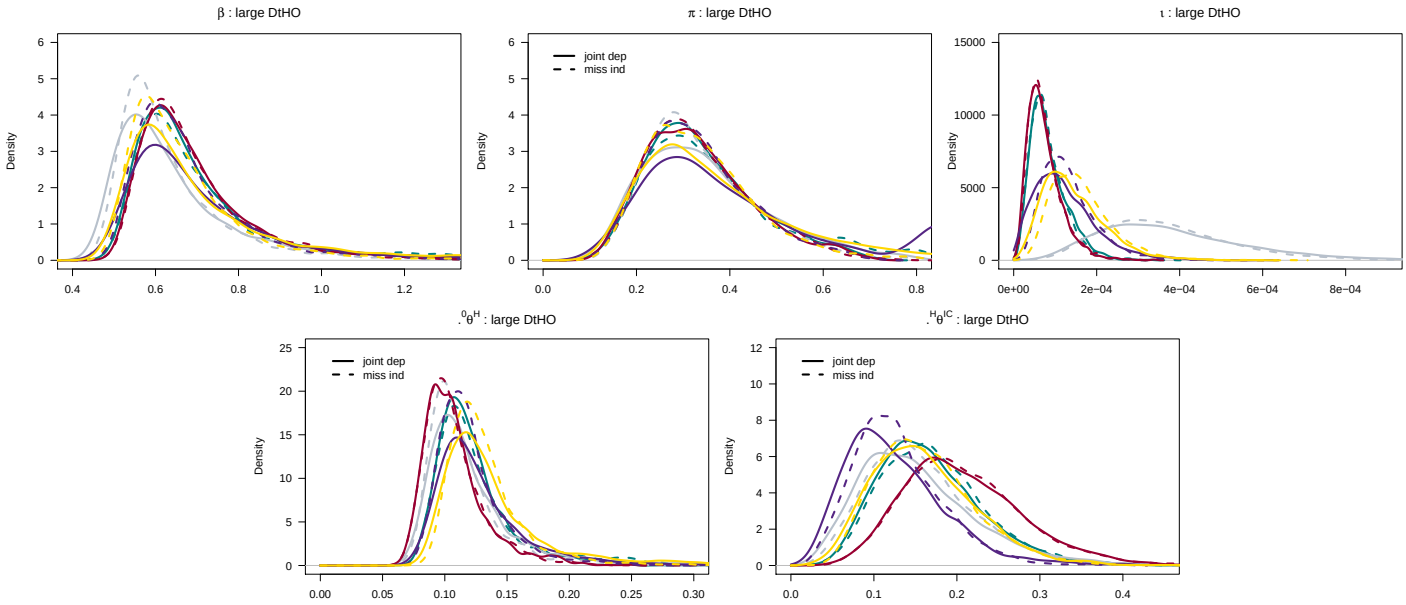


Figure E.8: Posterior distribution of the transmission and severity parameters from 5 datasets. The colour of the posterior density identifies the dataset analysed while dashed lines refer to results from the misspecified independent model and filled lines to results from the model approximating the joint dependent likelihood.

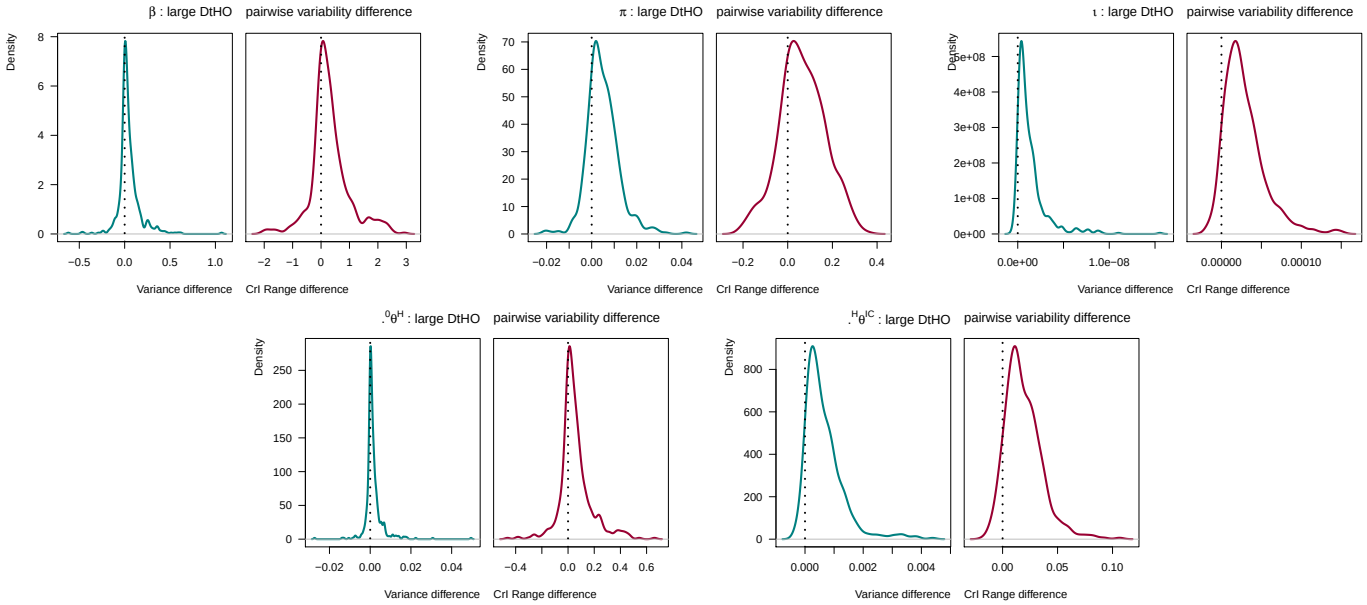


Figure E.9: Distribution of the pairwise difference in variance (green plots) and in 95% CrI length (red plots) of the posterior distribution of the transmission and severity parameters.

E.3.4 Increasing ζ^{IC}

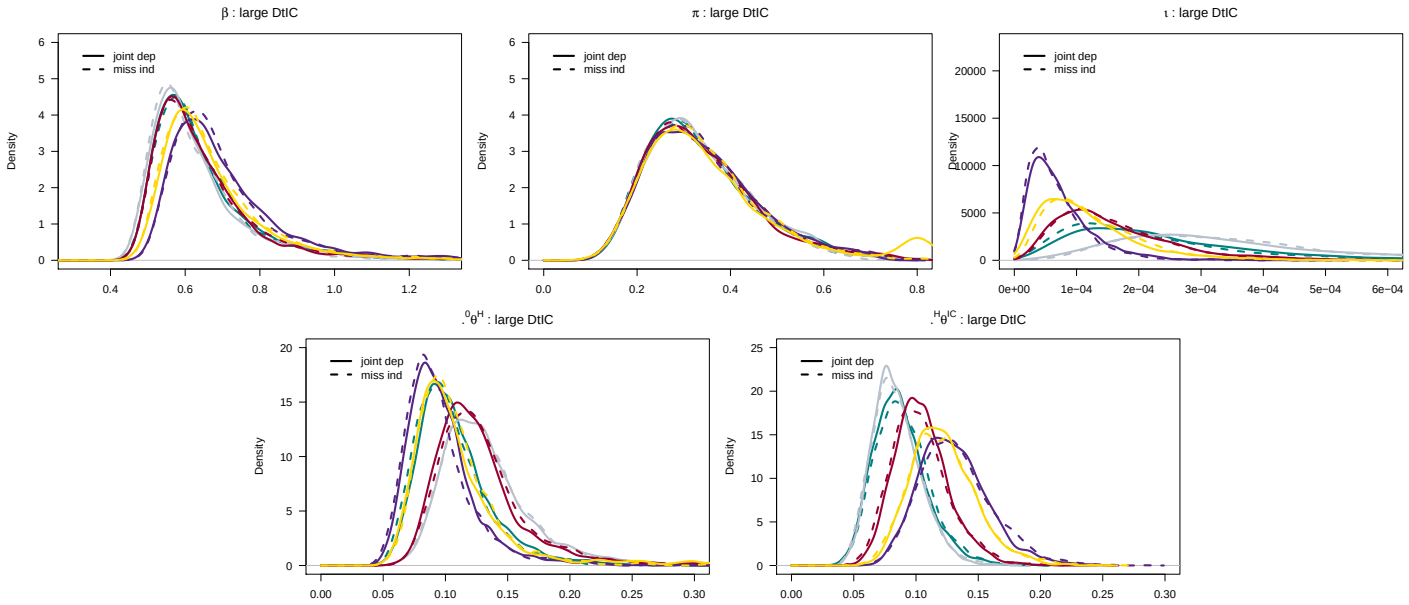


Figure E.10: Posterior distribution of the transmission and severity parameters from 5 datasets. The colour of the posterior density identifies the dataset analysed while dashed lines refer to results from the misspecified independent model and filled lines to results from the model approximating the joint dependent likelihood.

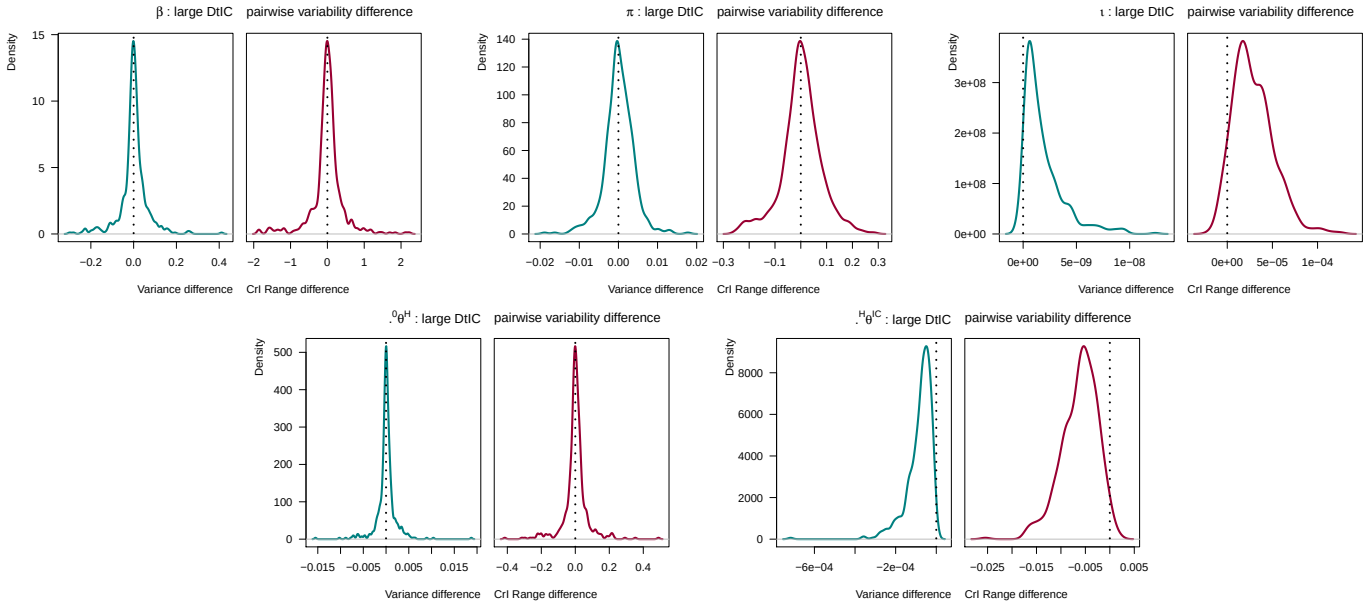


Figure E.11: Distribution of the pairwise difference in variance (green plots) and in 95% CrI length (red plots) of the posterior distribution of the transmission and severity parameters.

Appendix F

Appendix to Chapter 6

F.1 Derivation of the distribution of μ_u^{G}

To simplify notation, in what follows the incidence of influenza cases convolved over the delay from infection to General Practitioner (GP) consultation is denoted by:

$$\Upsilon_u = \sum_{c=0}^{C} \xi_{u-c}^{0} \, {}^{0}f_c^{\mathrm{F}}$$

Moreover, all the indexes referring to the day u are dropped: Υ stands for Υ_u; μ^{F} stands for μ_u^{F} etc.

The goal is to derive the distribution of $\mu^{\mathrm{G}} = \zeta^{\mathrm{G}}\alpha(\mu^{\mathrm{F}} + \mu^{\mathrm{B}}) = \zeta^{\mathrm{G}}\omega\mu^{\mathrm{F}} + \zeta^{\mathrm{G}}\omega\mu^{\mathrm{B}}$ that follows form the two following prior distributions:

$${}^{0}\theta^{\mathrm{F}} \sim \text{Gamma}(\iota^{\mathrm{F}}\varepsilon^{\mathrm{F}}, \varepsilon^{\mathrm{F}}) \tag{F.1}$$

$$\mu^{\mathrm{B}} \sim \text{Gamma}\left(b\frac{\varepsilon^{\mathrm{F}}}{\Upsilon}, \frac{\varepsilon^{\mathrm{F}}}{\Upsilon}\right) \tag{F.2}$$

From the properties of the Gamma random variable (r.v.) reported in Appendix A, $\mu^{\mathrm{F}} = \Upsilon \cdot {}^{0}\theta^{\mathrm{F}}$ still distributes as the following Gamma:

$$\mu^{\mathrm{F}} = \Upsilon \cdot {}^{0}\theta^{\mathrm{F}} \sim \text{Gamma}\left(\iota^{\mathrm{F}}\varepsilon^{\mathrm{F}}, \frac{\varepsilon^{\mathrm{F}}}{\Upsilon}\right) \tag{F.3}$$

Likewise the two quantities $\zeta^{\mathrm{G}}\omega\mu^{\mathrm{F}}$ and $\zeta^{\mathrm{G}}\omega\mu^{\mathrm{B}}$ distribute as follows

$$\begin{aligned} \zeta^{\mathrm{G}}\omega\mu^{\mathrm{F}} &\sim \text{Gamma}\left(\iota^{\mathrm{F}}\varepsilon^{\mathrm{F}}, \frac{\varepsilon^{\mathrm{F}}}{\Upsilon\zeta^{\mathrm{G}}\omega}\right) \\ \zeta^{\mathrm{G}}\omega\mu^{\mathrm{B}} &\sim \text{Gamma}\left(b\frac{\varepsilon^{\mathrm{F}}}{\Upsilon}, \frac{\varepsilon^{\mathrm{F}}}{\Upsilon\zeta^{\mathrm{G}}\omega}\right) \end{aligned} \tag{F.4}$$

Where $\zeta^{\mathrm{G}}\omega\mu^{\mathrm{F}}$ has mean:

$$M_1 = \frac{\iota^{\mathrm{F}}\varepsilon^{\mathrm{F}}}{\frac{\varepsilon^{\mathrm{F}}}{\Upsilon\zeta^{\mathrm{G}}\omega}} = \iota^{\mathrm{F}}\Upsilon\zeta^{\mathrm{G}}\omega \tag{F.5}$$

and variance

$$V_1 = \frac{\iota^{\mathrm{F}}\varepsilon^{\mathrm{F}}}{(\frac{\varepsilon^{\mathrm{F}}}{\Upsilon\zeta^{\mathrm{G}}\omega})^2} = \frac{\iota^{\mathrm{F}}}{\varepsilon^{\mathrm{F}}}\left(\Upsilon\zeta^{\mathrm{G}}\omega\right)^2 \tag{F.6}$$

and $\zeta^{\mathrm{G}}\omega\mu^{\mathrm{B}}$ has mean

$$M_2 = b\frac{\varepsilon^{\mathrm{F}}}{\Upsilon}\frac{\Upsilon\zeta^{\mathrm{G}}\omega}{\varepsilon^{\mathrm{F}}} = b\zeta^{\mathrm{G}}\omega \tag{F.7}$$

and variance

$$V_2 = b\frac{\varepsilon^{\mathrm{F}}}{\Upsilon}\left(\frac{\Upsilon\zeta^{\mathrm{G}}\omega}{\varepsilon^{\mathrm{F}}}\right)^2 = \frac{b\Upsilon}{\varepsilon^{\mathrm{F}}}\left(\zeta^{\mathrm{G}}\omega\right)^2 \tag{F.8}$$

Stewart et al. (2007) shows by simulation that if X and Y distribute according to a Gamma with respective means μ_x and μ_y and variances σ_x^2 and σ_y^2, then their sum can be approximated by a Gamma with mean $\mu_x + \mu_y$ and variance $\sigma_x^2 + \sigma_y^2$. Therefore, $\mu^{\mathrm{G}} = \zeta^{\mathrm{G}}\omega\mu^{\mathrm{F}} + \zeta^{\mathrm{G}}\omega\mu^{\mathrm{B}}$ can be approximated by a Gamma r.v. with mean and variance:

$$\begin{aligned} M_1 + M_2 &= \iota^{\mathrm{F}}\Upsilon\zeta^{\mathrm{G}}\omega + b\zeta^{\mathrm{G}}\omega \\ &= \zeta^{\mathrm{G}}\omega\left(\iota^{\mathrm{F}}\Upsilon + b\right) \\ V_1 + V_2 &= \frac{\iota^{\mathrm{F}}}{\varepsilon^{\mathrm{F}}}\left(\Upsilon\zeta^{\mathrm{G}}\omega\right)^2 + \frac{b\Upsilon}{\varepsilon^{\mathrm{F}}}\left(\zeta^{\mathrm{G}}\omega\right)^2 \\ &= \frac{\Upsilon}{\varepsilon^{\mathrm{F}}}\left(\zeta^{\mathrm{G}}\omega\right)^2\left(\iota^{\mathrm{F}}\Upsilon + b\right) \end{aligned} \tag{F.9}$$

From which the shape and rate parameter of the Gamma can be derived:

$$\begin{aligned} \alpha_{\mu^{\mathrm{G}}} = \frac{(M_1+M_2)^2}{V_1+V_2} &= \frac{\left(\zeta^{\mathrm{G}}\omega\right)^2\left(\iota^{\mathrm{F}}\Upsilon + b\right)^2}{\frac{\Upsilon}{\varepsilon^{\mathrm{F}}}\left(\zeta^{\mathrm{G}}\omega\right)^2\left(\iota^{\mathrm{F}}\Upsilon + b\right)} \\ &= \frac{\left(\iota^{\mathrm{F}}\Upsilon + b\right)\varepsilon^{\mathrm{F}}}{\Upsilon} \\ &= \iota^{\mathrm{F}} + \frac{b\varepsilon^{\mathrm{F}}}{\Upsilon} \end{aligned} \tag{F.10}$$

$$\begin{aligned} \beta_{\mu^{\mathrm{G}}} = \frac{(M_1+M_2)}{V_1+V_2} &= \frac{\left(\zeta^{\mathrm{G}}\omega\right)\left(\iota^{\mathrm{F}}\Upsilon + b\right)}{\frac{\Upsilon}{\varepsilon^{\mathrm{F}}}\left(\zeta^{\mathrm{G}}\omega\right)^2\left(\iota^{\mathrm{F}}\Upsilon + b\right)} \\ &= \frac{\varepsilon^{\mathrm{F}}}{\Upsilon(\zeta^{\mathrm{G}}\omega)} \end{aligned} \tag{F.11}$$

If the original notation is adopted again, the shape and rate parameters used in Equation 6.2.2 can be recognised.

F.2 Prior distribution for the background influenza-like illness (ILI)

Data on the weekly count of GP consultations for ILI from the first of January 2015 to the first of September 2017 are analysed.

An HHH model is fitted to the data. Observations are assumed to be Negative-Binomially distributed around a mean that comprises an epidemic autoregressive component and an endemic component, i.e.:

$$\begin{aligned} Y_t &\sim \mathrm{NegBinom}\left(\mu = \exp(\alpha Y_{t-1})\cdot b_t, \eta\right) \\ b_t &= \exp\left\{\nu_1 + \nu_2\cos\left(\frac{2\pi t}{52}\right) + \nu_3\sin\left(\frac{2\pi t}{52}\right)\right\} \end{aligned}$$

The fit of the model to the data is reported in Figure E.1.

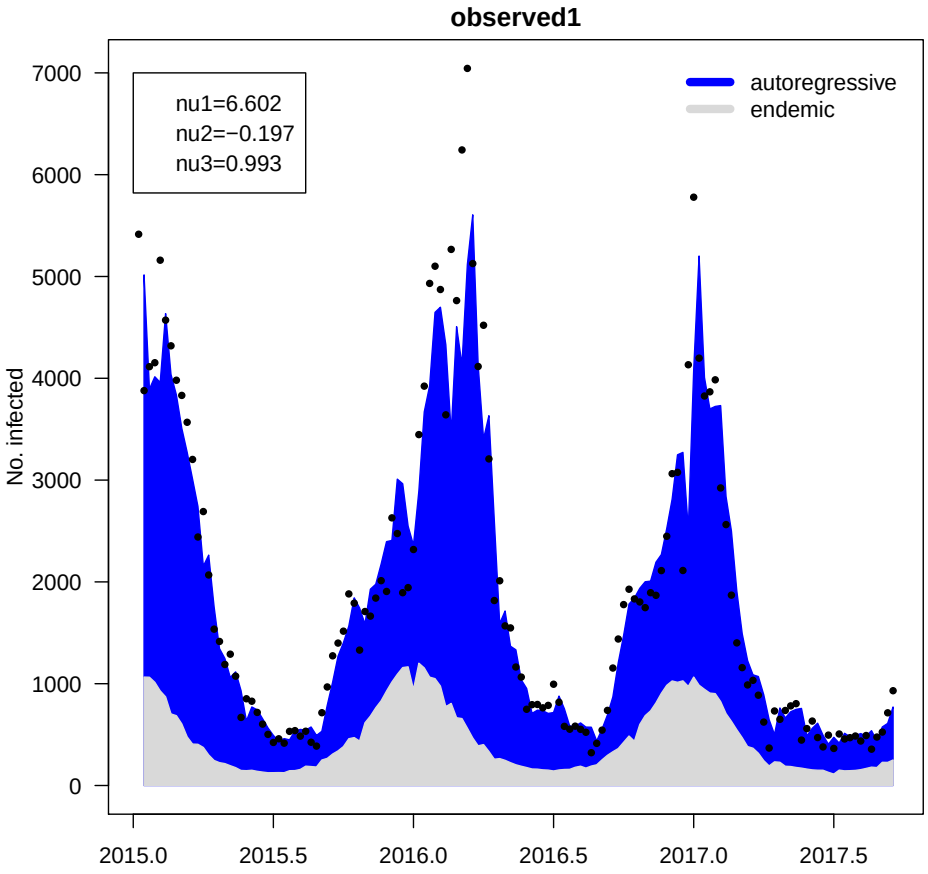


Figure F.1: Endemic/epidemic model of ILI-consultations counts.

F.3 Derivation of prior variance for ω_4

This derivation is taken from (Birrell et al., 2016).

A prior distribution is defined so that $\omega_1, \ldots \omega_7$ are Log-Normally distributed. Let

$$\begin{aligned} \alpha_1 &= \log(\omega_1) \\ \alpha_2 &= \log(\omega_2) \\ \alpha_3 &= \log(\omega_3) \\ \alpha_5 &= \log(\omega_5) \\ \alpha_6 &= \log(\omega_6) \\ \alpha_7 &= \log(\omega_7) \end{aligned} \tag{F.12}$$

and

$$\begin{aligned} \alpha_4 &= \log(\omega_4) \\ &= \log\left(1 \Big/ \prod_{i \in 1,2,3,5,6,7} \omega_i\right) \\ &= -\sum_{i \in 1,2,3,5,6,7} \alpha_i \end{aligned} \tag{F.13}$$

Let

$$\begin{aligned}
\alpha_1 &\sim \text{Norm}(0, \sigma^2) \\
\alpha_2 &\sim \text{Norm}(0, \sigma^2) \\
\alpha_3 &\sim \text{Norm}(0, \sigma^2) \\
\alpha_5 &\sim \text{Norm}(0, \sigma^2) \\
\alpha_6 &\sim \text{Norm}(0, \sigma^2) \\
\alpha_7 &\sim \text{Norm}(0, \sigma^2)
\end{aligned} \tag{F.14}$$

So the variance of the remaining parameter is

$$\begin{aligned}
\text{Var}(\alpha_4) &= \text{Var}\left(-\sum_{i \in 1,2,3,5,6,7} \alpha_i\right) \\
&= \sum_{i \in 1,2,3,5,6,7} \text{Var}(\alpha_i) + 2\sum_{i \neq j} \text{Cov}(\alpha_i, \alpha_j) \\
&\text{assume equal covariances } \text{Cov}(\alpha_i, \alpha_j) = c \\
&= 6\sigma^2 + 2\,(15c)
\end{aligned} \tag{F.15}$$

This parameter is imposed to have the same variance as all the others:

$$6\sigma^2 + 2\,(15c) = \sigma^2 \tag{F.16}$$

from which:

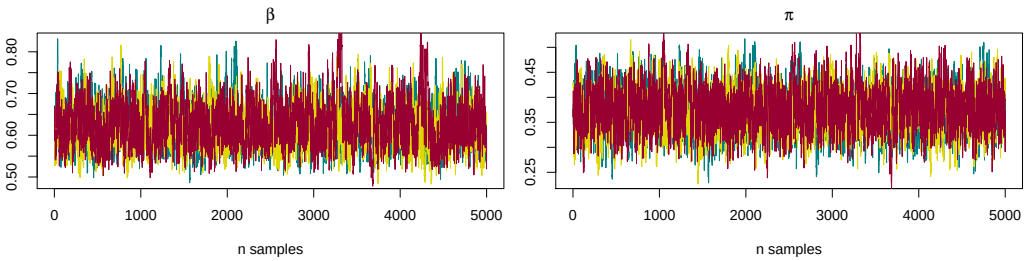
$$c = -\frac{1}{6}\sigma^2 \tag{F.17}$$

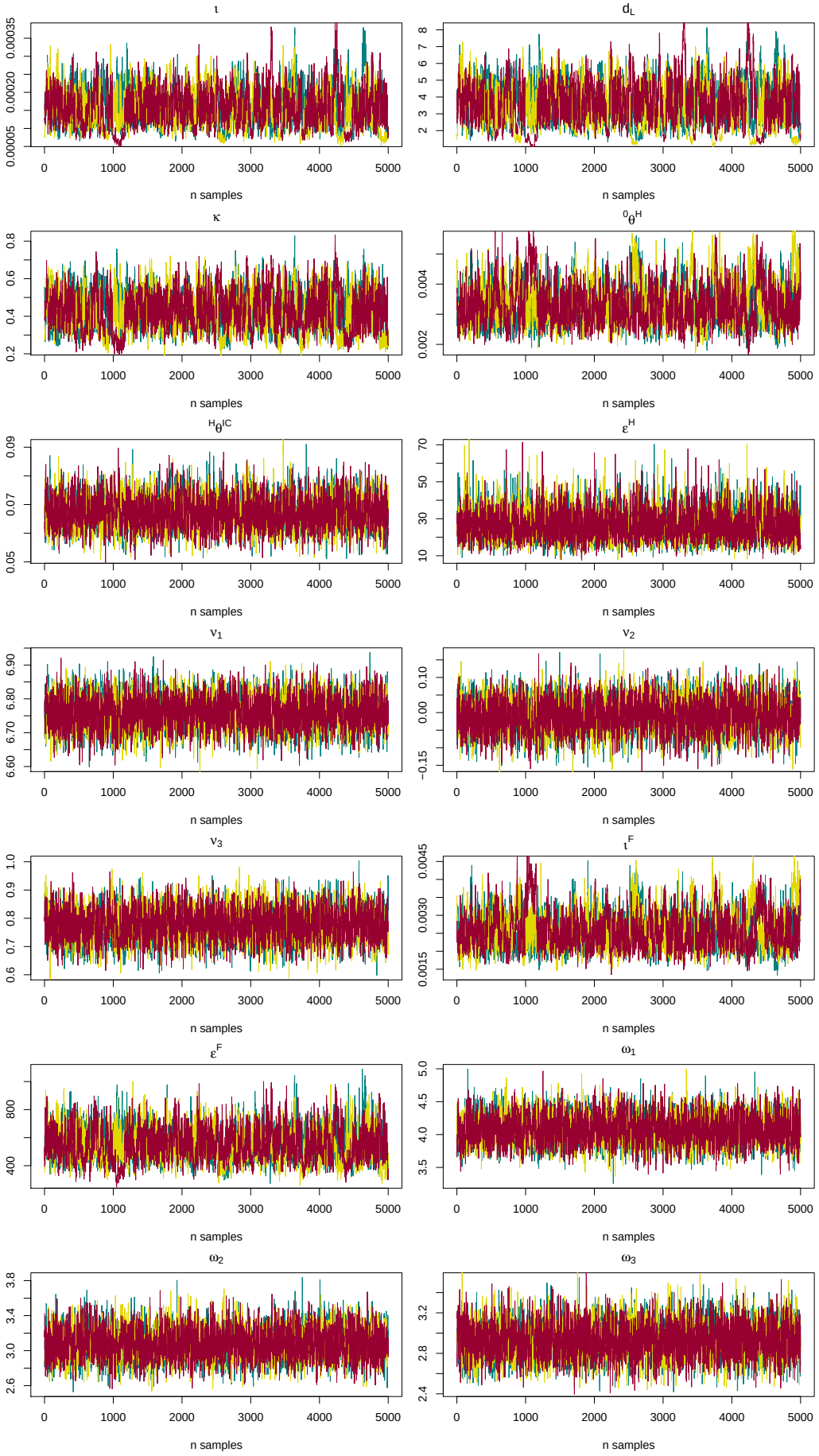
so that

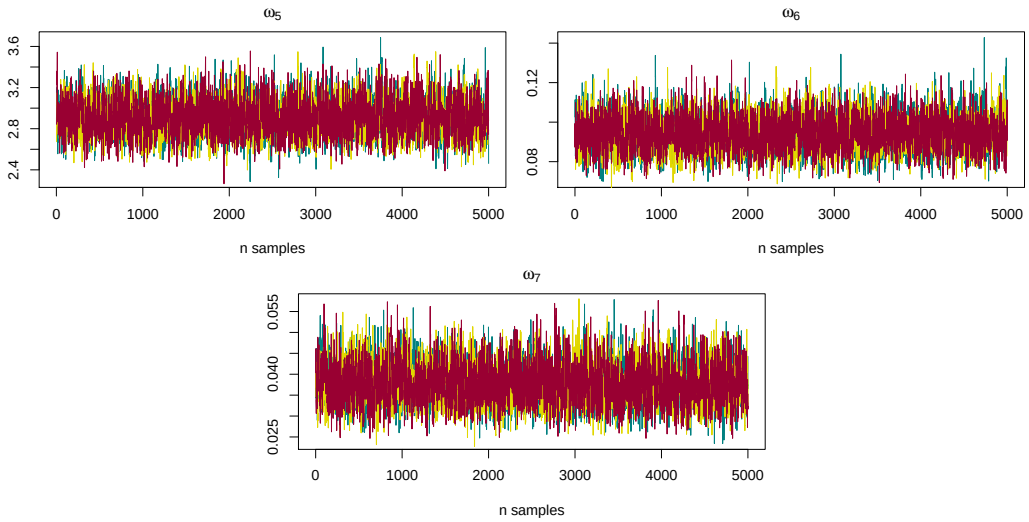
$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \end{pmatrix} \sim \text{Norm}\left(0; \sigma^2 \begin{pmatrix} 1 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 \\ -1/6 & 1 & -1/6 & -1/6 & -1/6 & -1/6 \\ -1/6 & -1/6 & 1 & -1/6 & -1/6 & -1/6 \\ -1/6 & -1/6 & -1/6 & 1 & -1/6 & -1/6 \\ -1/6 & -1/6 & -1/6 & -1/6 & 1 & -1/6 \\ -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & 1 \end{pmatrix}\right) \tag{F.18}$$

F.4 Convergence assessment

The Markov chain Monte Carlos (MCMCs) samples of all the parameters are displayed below.




ι
0.00005
0.00020
0.00035
0 1000 2000 3000 4000 5000
n samples
d_L
2 3 4 5 6 7 8
0 1000 2000 3000 4000 5000
n samples
κ
0.2 0.4 0.6 0.8
0 1000 2000 3000 4000 5000
n samples
$^0\theta^H$
0.002 0.004
0 1000 2000 3000 4000 5000
n samples
$^H\theta^{IC}$
0.05 0.07 0.09
0 1000 2000 3000 4000 5000
n samples
ε^H
10 30 50 70
0 1000 2000 3000 4000 5000
n samples
ν_1
6.60 6.70 6.80 6.90
0 1000 2000 3000 4000 5000
n samples
ν_2
−0.15 0.00 0.10
0 1000 2000 3000 4000 5000
n samples
ν_3
0.6 0.7 0.8 0.9 1.0
0 1000 2000 3000 4000 5000
n samples
ι^F
0.0015 0.0030 0.0045
0 1000 2000 3000 4000 5000
n samples
ε^F
400 800
0 1000 2000 3000 4000 5000
n samples
ω_1
3.5 4.0 4.5 5.0
0 1000 2000 3000 4000 5000
n samples
ω_2
2.6 3.0 3.4 3.8
0 1000 2000 3000 4000 5000
n samples
ω_3
2.4 2.8 3.2
0 1000 2000 3000 4000 5000
n samples


ω5
2.4
2.8
3.2
3.6
0
1000
2000
3000
4000
5000
n samples
ω6
0.08
0.12
0
1000
2000
3000
4000
5000
n samples
ω7
0.025
0.040
0.055
0
1000
2000
3000
4000
5000
n samples