PREDICTING EPIDEMIOLOGICAL TRANSITIONS IN DISEASE DYNAMICS

PREDICTING EPIDEMIOLOGICAL TRANSITIONS IN INFECTIOUS DISEASE DYNAMICS: SMALLPOX IN HISTORIC LONDON (1664-1930)

By OLGA KRYLOVA, B.Sc., M.Sc., B.Sc., M.A.St., M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment

of the Requirements for the Degree Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2011) McMaster University
(Mathematics) Hamilton, Ontario

TITLE: Predicting epidemiological transitions
in infectious disease dynamics:
Smallpox in historic London (1664-1930)

AUTHOR: Olga Krylova
M.Sc. (McMaster University, Canada)
M.A.St. (Cambridge University, UK)
B.Sc. (Trent University, Canada)
M.Sc. (Odessa State Academy, Ukraine)
B.Sc. (Odessa State Academy, Ukraine)

SUPERVISOR: Professor David J.D. Earn

NUMBER OF PAGES: xiii, 186

Abstract

Mathematical modelling has become a powerful tool used to predict the spread of infectious diseases in populations. Successful analysis and modelling of historical infectious disease data can explain changes in the pattern of past epidemics and lead to a better understanding of epidemiological processes. The lessons learned can be used to predict future epidemics and help to improve public health strategies for control and eradication.

This thesis is focused on the analysis and modelling of smallpox dynamics based on the weekly smallpox mortality records in London, England, 1664-1930. Statistical analysis of these records is presented. A timeline of significant historical events related to changes in variolation and vaccination uptake levels and demographics was established. These events were correlated with transitions observed in smallpox dynamics. Seasonality of the smallpox time series was investigated and the contact rate between susceptible and infectious individuals was found to be seasonally forced. Seasonal variations in smallpox transmission and changes in their seasonality over long time scale were estimated. The method of transition analysis, which is used to predict qualitative changes in epidemiological patterns, was used to explain the transitions observed in the smallpox time series. We found that the standard SIR model exhibits dynamics similar to the more realistic Gamma distributed SEIR model if the mean serial interval is chosen the same, so we used the standard SIR model for our analysis. We conclude that transitions observed in the temporal pattern of smallpox dynamics can be explained by the changes in birth, immigration and intervention uptake levels.

This thesis is dedicated to my father, who nourished my curiosity in science and encouraged my pursuit of a Ph.D. degree.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. David Earn. It has been an honour to be David's Ph.D. student and a delight to work on the fascinating project of studying smallpox dynamics in historic London that he introduced me to. His enthusiasm and love for the research always inspired me. I cannot thank him enough for all the encouragement and advice he gave me. I have learned a lot from him during this time and I am convinced that this knowledge will help me in the future. I also would like to thank my supervisory committee members Dr. Jonathan Dushoff and Dr. Gail Wolkowicz for their time, interest and valuable comments. Jonathan was extremely helpful with providing some valuable insight.

Many thanks goes to Raluca Eftimie who dedicated her time to read every section of my thesis and gave me very helpful advice and comments. Thanks to Karsten Hempel and Shawn Winnigton-Ball for proofreading my work. I would also like to thank Susan Marsh-Rollo for helping me with a picture for the natural history of smallpox, Kelly Hancock for a lot of help with data issues as well as Daihai He and all other members of the Mathematical Biology Group for their help and support. Thanks to my many friends in the math department who made these years of graduate school quite enjoyable.

I am very grateful to my family for all their love and encouragement. My dad, who believes that I can achieve anything as long as I put my mind into it, work hard and be patient and whose biggest dream was to see me succeeding in obtaining this degree. My mom for being there for me whenever I needed her. And of course I would like to express my greatest thanks to my biggest supporter, the person who helped me the most, who has been very patient with me, who spent many late nights discussing my ideas, who supported me no matter how hard it was, who believed

in me, always lifted my spirit in the darkest moments and encouraged me through all these years, my dearest boy-friend, Yuri. I appreciate your help more than you know. Thank you for being such a wonderful person and a great friend.

This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Thank you all, this Ph.D. would not have happened without you!

Contents

List of Figures

List of Tables

Declaration

I hereby declare that to the best of my knowledge the work presented in this thesis is original and my own work and effort under the supervision of Dr. D.J.D. Earn. Where other sources of information have been used, they have been acknowledged.

The material in this thesis is submitted for the degree of Ph.D. at McMaster University only and has not been submitted to any other universities.

Chapter 1

Introduction

Infectious diseases are among the leading causes of death in the world, contributing an overwhelming 15 million deaths each year to global human mortality (WHO, 2004). Even when a person survives an infection, decreased productivity, risk of morbidity and diminished quality of life pose a great burden on society. Despite high-tech innovations in the modern era, great improvements in sanitation and health care, tremendous research efforts made in the virology and epidemiology of infectious diseases, and great advances in treatment and prevention of infectious diseases, we still cannot prevent most infections. So far only one disease – smallpox – out of thousands of infectious diseases affecting humans, has been eradicated. Newly emerging diseases (SARS, West Nile virus, Ebola) and previously suppressed diseases (cholera, malaria, tuberculosis, E.coli, *etc.*) are currently on the rise (WHO, 1996). At the same time the possibility of a biological attack that could involve some of the deadliest viruses (smallpox, anthrax, plague) poses additional threats. Therefore, now more than ever, research dedicated to a greater understanding of infectious diseases is extremely valuable.

Mathematical models have become an effective tool that is extensively used to explain invasion and spread of infectious diseases through populations. Various deterministic and stochastic models have been developed to capture the main

characteristics of disease dynamics and have contributed enormously to the selection and implementation of optimal control measures (Anderson and May 1991; Hethcote 2000).

The birth of epidemiological modelling is often attributed to the work of Daniel Bernoulli, who developed a differential equation model to evaluate the impact of variolation on public health, and argued that it would increase a person's life expectancy by about three years (Bernoulli 1766; Blower and Bernoulli 2004; Dietz and Heesterbeek 2000, 2002). Since the introduction of this first model the area of study of infectious diseases through mathematical modelling has advanced immensely (Hethcote 2000): from the work of Hamer on measles epidemics and his introduction of the "law of mass action" to reflect the contact between individuals (Hamer 1906a,b,c); to the work of Sir Ronald Ross on malaria models (Ross 1911); to the work of Kermack and McKendrick, who formulated the notion of *epidemic threshold* - a critical density of susceptibles in a population for an epidemic to occur (Kermack and McKendrick 1927); to the stochastic models of Soper (Soper 1929), Barlett (Bartlett 1957a, 1960) and Bailey (Bailey 1964, 1975); to the extensive study of various models by Anderon and May (Anderson and May 1991); to the sophisticated present-day models that incorporate age structure, social and spatial structure of modern society, as well as various scenarios of public health interventions (Bozzette et al. (2003a); Kaplan et al. (2003); Burke et al. (2006); Riley and Ferguson (2006), Longini et al. (2007); Glasser et al. (2008) just to name a few).

The most standard models used to describe the course of an epidemic are the *Susceptible–Infectious–Recovered* and *Susceptible–Exposed–Infectious–Recovered* models, or in short the SIR and SEIR models (Anderson and May 1991; Kermack and McKendrick 1927). These models are widely used and extremely popular among researchers because of their simplicity for mathematical analysis, easy implementation and effectiveness. Incorporation of seasonal forcing into the models (i.e. that the contact rate between individuals varies seasonally) allows for

the analysis of the complex dynamics of the recurrent epidemics (London and Yorke 1973; Olsen and Schaffer 1990).

The temporal structure of recurrent epidemics varies significantly across times and places, and among diseases (London and Yorke (1973); Anderson and May (1991); Earn et al. (1998); Earn et al. (2000a); Grenfell et al. (2001); Bauch and Earn (2003b)). For example, the New York City (NYC) measles time series (1928–1972), which has been investigated using numerous mathematical models (London and Yorke (1973); Yorke and London (1973); Olsen and Schaffer (1990); Bolker and Grenfell (1993); Earn et al. (2000a); Bauch and Earn (2003a,b), and also analyzed in **Chapter 2**), shows transitions from a 2–3-year cycle to a strictly biennial cycle and then to a 3–4 year cycle. One of the goals of mathematical modelling is to predict and explain such changes in the historical incidence patterns of infectious diseases. The ability to predict the timing of the transitions from one type of periodic structure to another can lead to a better understanding of the nature of these transitions and provide a valuable insight about the factors that cause the spread of epidemics in populations. As a consequence, the knowledge obtained can be helpful in predicting current epidemics and in improving public health strategies for control and eradication.

The main goal of the current research is to analyze and model smallpox dynamics in historic London, England over almost three centuries. The method of *transition analysis* introduced by Earn and colleagues (Earn et al. (2000a); Bauch and Earn (2003a,b)) is used to predict and explain changes observed in the temporal pattern of smallpox time series. Transition analysis was designed to predict qualitative changes in epidemic dynamics induced by demographic and behavioural changes in the host population. Previously, this analysis was performed with the seasonally forced SIR and SEIR models. One implicit assumption built into the SIR-type models is that the lengths of time individuals spend in each disease stage (latent and infectious) are exponentially distributed. While this assumption sim-

plifies the mathematics it does not correspond to reality, where the distributions of the infectious and latent stages are narrow and do not resemble exponentials. We propose to use more realistic Gamma distributed SIR and SEIR models. In **Chapter 2** we revisit the method of transition analysis and investigate how the shape of the disease stage distributions affects predictions of this method. As an illustrative example, we apply the transition analysis to the well-known weekly NYC measles incidence time series (Yorke and London 1973). In **Chapter 5** we use transition analysis to explain changes in smallpox dynamics in London, 1664-1930.

The study of epidemic patterns would not be possible without access to historical incidence or disease mortality records. Available outbreak data, including some recent individual-based records, create a myriad of puzzles to be solved by mathematical modellers, e.g., why some infectious disease exhibit rare outbreaks with extinction and re-emergence (e.g. 17th century plague in London) and some show recurrent epidemics (e.g. measles in NYC), why epidemic patterns change from place to place and from time to time, *etc.* (Earn 2009). Moreover, the data are absolutely necessary for model validation and estimation of parameter values. Analysis of published records that cover relatively short periods of time (up to 100 years) have allowed researches to make many interesting discoveries (Anderson and May 1991; Bartlett 1957a; Earn et al. 2000a; Fine and Clarkson 1982). However longer data sets, which could potentially lead to a deeper understanding of epidemiological processes, have not been accessible. Recently, the weekly London Bills of Mortality spanning over three centuries were digitized at McMaster University by the Earn's research group. This rich data set provides a unique opportunity to study historical records of many infectious diseases. Note that, while data from the London *annual* bills were previously available, they did not provide the same level of detail. We are particularly interested in the study of smallpox epidemics, which are likely to be among the most accurate records in the bills due to the unique and easily identifiable presentation of the disease.

Smallpox, almost forgotten now, was considered to be one of the most terrifying infectious diseases because of its high mortality and morbidity rate (Fenner et al. 1988). During its circulation, smallpox caused devastation, killing thousands of people each year and often leaving its survivors scarred or disabled for life. Some historians have even argued that this disease was responsible for events that changed the course of human history (Hopkins 1983; Macaulay 1866; McNeill 1998; Razzell 1977). For example, the outcome of the war between the native population of Mexico and Cortez's army during the Spanish invasion of the 16th century was decided after the introduction of smallpox to the previously unexposed populations of the Aztec and Inca civilizations. The virus killed nearly half of the native population in less than six months.

Chapters 3, 4 and **5** are devoted to an in-depth analysis of the London smallpox mortality time series. Detailed statistical analysis of the data is presented in **Chapter 3**. First we investigate the history of smallpox to get a better understanding of how the disease spreads among populations. In particular, we discuss the origin and early history of smallpox, the invention of various preventative measures and the story of its eventual eradication. We also describe the natural history of smallpox infection, which is necessary to understand for further modeling work. Since our data set was specifically collected from London we outline the history of smallpox in England. This allows us to establish a timeline of important historical events related to implementation of control measures, population movement, wars, *etc.*, which may have influenced smallpox dynamics. Using standard statistical tools we analyze and describe the temporal pattern of smallpox epidemics. We are also able to correlate the established timeline with the observed changes in smallpox dynamics.

The analysis of the seasonal structure of smallpox time series performed in **Chapter 3** suggested the presence of seasonality in the observed data. Since seasonal variation in the infectious disease transmission is a key driver of epidemic

dynamics (London and Yorke 1973), in **Chapter 4** we estimate the amplitude and the seasonal pattern of the smallpox transmission rate from the smallpox time series. In order to do so we develop a simple and computationally efficient method that allows us to estimate seasonality of the transmission rate for long data sets, such as the London smallpox time series.

Chapter 5 of this thesis attempts to determine if demographic and behavioral changes can explain the transitions observed in the temporal pattern of smallpox epidemics in London. By applying the method of transition analysis to the smallpox mortality data we investigate if the changes in the rate of susceptible recruitment in the population could have triggered observed transitions in smallpox dynamics.

A summary of our major results and future research directions are presented in **Chapter 6**.

Chapters 2, 3, 4 and **5** are preliminary versions of independent papers that are intended to be submitted for publication. They are therefore formatted as separate papers.

Chapter 2

Effects of the infectious period distribution on predicted transitions in childhood disease dynamics

Abstract

The population dynamics of infectious diseases occasionally undergo rapid qualitative changes, such as transitions from annual to biennial cycles or to irregular dynamics (Anderson and May 1991; Olsen and Schaffer 1990). Previous work based on the standard seasonally forced SEIR (*susceptible, exposed, infectious, removed*) model has found that transitions in the dynamics of many childhood diseases result from bifurcations induced by slow changes in birth and vaccination rates (Bauch and Earn 2003b; Earn 2009; Earn et al. 2000a). However, the standard formulation of the SIR and SEIR models assumes that the stage durations (infectious and latent periods) are exponentially distributed; while mathematically convenient, the exponential distribution is far from the biological reality of narrow distributions centred around the mean. Much recent work has indicated that realistically distributed stage durations strongly affect the dynamical structure of the seasonally forced SIR model (Black and McKane 2010a; Conlan et al. 2010; Keeling and Grenfell 2002; Lloyd

2001a,b; Nguyen and Rohani 2008; Wearing et al. 2005). We investigate whether inferences drawn from analyses of epidemiological transitions based on the SIR and SEIR models are robust to the shapes of the stage duration distributions. We find that with a fixed mean infectious period in the SIR model, the dynamical structure (and predicted transitions) vary substantially as a function of the shape of the infectious period distribution. In contrast, with fixed mean latent and infectious periods in the SEIR model, the shapes of the stage duration distributions have a less dramatic effect on model dynamical structure and predicted transitions. Finally, we find that all these results can be understood most easily by considering the distribution of the disease generation interval, as opposed to the distributions of individual disease stages. For a given mean generation interval, the dynamics of the SIR and SEIR are nearly equivalent and are very insensitive to the shapes of the disease stage distributions.

Keywords: SIR model, Gamma distribution, Erlang distribution, ordinary differential equations, delay differential equations, bifurcation theory, serial interval

2.1 Introduction

Mathematical modelling has proven to be an extremely powerful tool for understanding epidemiological patterns and predicting how demographic changes and control measures influence infectious disease dynamics (Anderson and May 1991; Earn 2009; Hethcote 2000). The most commonly used framework for modelling transmission dynamics involves dividing the population into compartments based on disease status and using ordinary differential equations (ODEs) to specify flows between the compartments. For diseases that confer permament immunity, the simplest case is the SIR model (Anderson and May 1991; Kermack and McKendrick 1927), in which the compartments represent *Susceptible*, *Infectious* and *Removed* individuals, while the SEIR model also includes an *Exposed* compartment, contain-

ing individuals who are in a latent stage (infected but not yet infectious). These simple models implicitly assume that the time an individual spends in each disease stage (e.g., latent or infectious) is drawn from exponential distributions (Brauer 2008; Hethcote 2000), which are unlike real distributions of disease stage durations.

The dynamical effects of exponential versus more realistic distributions of stage durations have been explored extensively in the literature (Lloyd (2001a,b); Keeling and Grenfell (2002); Wearing et al. (2005); Nguyen and Rohani (2008); Black and McKane (2010a); Conlan et al. (2010)), which has revealed that changing the shapes of these distributions while keeping their means fixed can have a large impact on predicted dynamics. Consequently, it is important to re-evaluate any inferences drawn about real data from models that assume exponentially distributed stage durations. In this paper we will investigate how the shapes of latent and infectious period distributions affect our predictions concerning epidemiological transitions (e.g., from annual to biennial epidemic cycles) and compare our results with conclusions previously made based on bifurcation theory applied to exponentially distributed models (Bauch and Earn 2003a,b; Earn et al. 2000a).

The shapes of real distributions of disease stage durations

Many authors have estimated infectious period distributions by fitting standard probability distributions (e.g., Normal (Bailey 1956a,b; Gough 1977), Log-normal (Nishiura 2007a; Nishiura and Eichner 2007), Gamma (Eichner and Dietz 2003; Wearing et al. 2005) or Fixed (Bailey 1956a,b)) to empirical data. For transmission modelling, a Gamma distribution with an integer shape parameter—also known as an *Erlang distribution*—is strongly preferred on theoretical grounds: the Erlang distribution is equivalent to a sequence of independent and identically distributed exponential distributions (Anderson and Watson 1980; Bailey 1964; Lloyd 2001a; Ma and Earn 2006), so compartmental transmission models with Erlang-distributed

stage durations can be expressed as ODEs (as opposed to the integro-differential equations required to express compartmental models with arbitrarily distributed stage durations).

The Erlang distribution with shape parameter n and scale parameter $n\gamma$, Erlang(n, $n\gamma$), has probability density

$$f\left(x; n, n\gamma\right) = \frac{(n\gamma)^n}{(n-1)!} x^{(n-1)} e^{-n\gamma x}, \qquad x > 0,\ n \in \mathbb{N}. \tag{2.1}$$

The mean is $1/\gamma$ and the variance is $1/n\gamma^2$.

The Erlang distribution is more restricted in shape than the general Gamma distribution, but it is sufficiently flexible to provide a good approximation of realistic stage duration distributions. **Figure 2.1** shows the probability density function of the Erlang distribution with mean $\frac{1}{\gamma} = 13$ days (dashed vertical line) and various shape parameters ($n = 1,\ 2,\ 3,\ 5,\ 8,\ 20,\ 100$).

We write SI^{n}R and SE^{m}I^{n}R to refer to the Erlang distributed SIR and SEIR models, where m and n refer to the shape parameters of the latent and infectious period distributions, respectively. Thus SI^{1}R ($n = 1$) and SE^{1}I^{1}R ($m = 1,\ n = 1$) denote the standard SIR and SEIR models with exponentially distributed latent and infectious periods. Estimated values of n and m can be inferred from appropriate clinical data and vary widely for different infectious diseases, e.g., $m = 2$, $n = 3$ for SARS and $m = 20$, $n = 20$ for measles (Wearing et al. 2005).

Dynamics of epidemic models with Erlang-distributed stage durations

In the past 20 years, the SI^{n}R and SE^{m}I^{n}R models—and other more general models—have received a great deal of attention. Equilibrium stability analyses have been conducted on "unforced" models that assume constant contact rates (Feng and Thieme 2000a,b; Hethcote and Tudor 1980; Lloyd 2001a,b; Zhang et al. 2008),

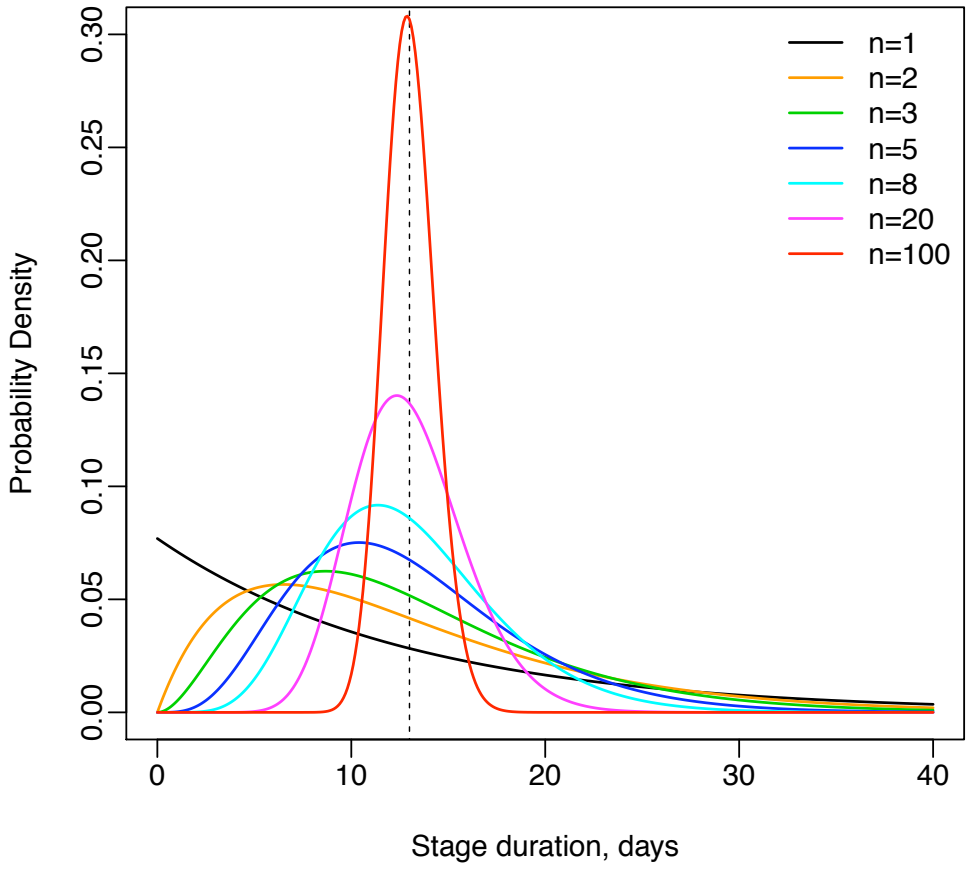


Figure 2.1: Probability density functions for Erlang distributions with mean of 13 days (vertical dashed line) and shape parameter n (**equation** (2.1)). Note that $n = 1$ yields the exponential distribution and $n \to \infty$ yields the Dirac delta distribution.

and bifurcation analyses have been conducted on "forced" models in which contact rates vary seasonally (Black and McKane 2010a; Conlan et al. 2010; Grossman 1980; Keeling and Grenfell 2002; Lloyd 2001a,b; Nguyen and Rohani 2008; Wearing et al. 2005). Lloyd (2001b) found that the biennial pattern observed in the SI^1R model is reproduced by the SI^nR model but with much weaker seasonality.

Nguyen and Rohani (2008) found that complex dynamics of whooping cough could be understood based on the multiple co-existing attractors of an SE^1I^5R model, whereas the simple SE^1I^1R model with the same parameter values always predicts an asymptotically annual cycle. Wearing et al. (2005) argued that the traditional assumptions of exponentially distributed latent and infectious periods may lead to underestimation of the basic reproduction number, $\mathcal{R}_0$, and hence to underestimation of the levels of control required to curtail an epidemic.

The primary theme of recent work on SI^nR and SE^mI^nR models has been that the shapes of stage duration distributions can significantly affect the qualitative dynamics of infectious diseases. Given this, it is important to re-examine previous work that has attempted to explain observed disease dynamics based on SI^1R or SE^1I^1R models, and determine whether the conclusions of these previous studies remain valid when the analyses are repeated using models with more realistically distributed stage durations. Our particular focus in this paper is on epidemiological transition analysis, by which we mean predicting qualitative changes in epidemic dynamics induced by demographic and behavioural changes in the host population (Bauch and Earn 2003b; Earn 2009; Earn et al. 2000a). As an illustrative example, we analyze measles incidence in New York City for the period 1928–1972, which was first investigated by London and Yorke (London and Yorke 1973; Yorke and London 1973) and has been the subject of numerous studies over the last 40 years (Bauch and Earn 2003b; Bolker and Grenfell 1993; Earn et al. 2000a; Olsen and Schaffer 1990).

We begin by describing the SI^nR and SE^mI^nR models in **Section 2.2** and the method of transition analysis in **Section 2.3**. In **Section 2.4** we apply transition analysis, based on SI^nR and SE^mI^nR models, to measles dynamics in New York City from 1928 to 1972. We consider the role of the distribution of the disease generation interval (as opposed to the latent and infectious periods) in **Section 2.5** and summarize our results in **Section 2.6**.

2.2 Models

2.2.1 SI^1R and SE^1I^1R

Assuming the population is large and homogeneously mixed, the (unforced) SI^1R model can be cast as a simple system of nonlinear ordinary differential equations (Anderson and May 1991; Kermack and McKendrick 1927):

$$\frac{dS}{dt} = \Phi - \beta SI - \mu S\,, \tag{2.2a}$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I\,, \tag{2.2b}$$

$$\frac{dR}{dt} = \gamma I - \mu R\,. \tag{2.2c}$$

Here, S, I and R are the numbers of susceptible, infectious, and recovered (immune) individuals in the population. The μ, β and γ are the rates of *per capita* death, transmission and recovery, respectively. μ quantifies death from "natural causes" (disease-induced mortality is assumed to be negligible). Φ denotes the number of births per unit time, which is often time-dependent in practice (He and Earn 2007). If $\Phi = \mu N$, where $N = S + I + R$ is the total population size, then births balance deaths and the population size remains constant. β is the rate at which contacts between susceptible and infectious individuals cause new infections (per susceptible per infected), so βSI is the number of new infections that occur per unit time (the *incidence rate*). Note that **equations** (2.2a) and (2.2b) do not depend on R. Therefore they completely specify the system dynamics and **equation** (2.2c) can be ignored.

For our purposes, the birth term (Φ) is particularly important because secular changes in this term can induce dynamical transitions (Bauch and Earn 2003a,b; Earn 2009; Earn et al. 2000a). We estimate Φ based on demographic data and do not assume that it scales with population size (e.g., we do *not* assume $\Phi = \mu N$). Nevertheless, it is convenient to express Φ in units that are similar to those of the

per capita death rate μ. We therefore write $\Phi = \nu N_0$, where N_0 is the population size at a particular "anchor time" t_0 (see also **Section 2.3.2**). ν represents births *per capita* at time t_0, but not at other times. We rewrite **equation** (2.2) as

$$\frac{dS}{dt} = \nu N_0 - \beta SI - \mu S\,, \tag{2.3a}$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I\,, \tag{2.3b}$$

$$\frac{dR}{dt} = \gamma I - \mu R\,. \tag{2.3c}$$

A fundamental characteristic of an infectious disease is its *basic reproduction number,* $\mathcal{R}_0$, which is defined as the mean number of susceptible individuals infected by one infectious individual in a completely susceptible population (Anderson and May 1991). This number determines whether the infection dies out ($\mathcal{R}_0 < 1$) or spreads ($\mathcal{R}_0 > 1$) in a population. For the $\mathrm{SI^1R}$ model,

$$\mathcal{R}_0^{\mathrm{SI^1R}} = \frac{\nu N_0}{\mu}\frac{\beta}{\gamma+\mu}\,. \tag{2.4}$$

The first factor here ($\nu N_0/\mu$) does not normally appear in formulae for $\mathcal{R}_0$ because it is typically assumed that births balance deaths, and the population size is typically absorbed into the transmission rate β. We assume that ν changes slowly enough that it can be regarded as constant for the purposes of defining $\mathcal{R}_0$ at a given time.

The $\mathrm{SI^1R}$ model can easily be extended to the $\mathrm{SE^1I^1R}$ model, which includes a *latent stage*, by replacing **equation** (2.3b) with the two equations:

$$\frac{dE}{dt} = \beta SI - \sigma E - \mu E\,, \tag{2.5a}$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \mu I\,. \tag{2.5b}$$

The latent period is defined to be the time from initial infection to becoming infectious. In **equations** (2.5a) and (2.5b) E represents the number of *exposed* individ-

uals (individuals in the latent stage). For the SE^1I^1R model, the basic reproduction number is

$$\mathcal{R}_0^{SE^1I^1R} = \frac{\nu N_0}{\mu} \frac{\beta\sigma}{(\sigma+\mu)(\gamma+\mu)}. \tag{2.6}$$

In the standard SI^1R and SE^1I^1R formulation (**equations** (2.3) and (2.5)), the lengths of the latent and infectious stages are exponentially distributed. To see this, suppose that during the infectious stage the only process occurring was recovery from infection. Then **equation** (2.3b) would reduce to $dI/dt = -\gamma I$, which implies that the distribution of time spent in the infectious class (the infectious period) is exponential with mean $1/\gamma$ (if I_0 individuals are infectious at time 0 then $I_0 e^{-\gamma t}$ are still infectious at time t). Similarly the latent period is exponentially distributed with mean $1/\sigma$.

2.2.2 SI^nR and SE^mI^nR

Arbitrarily distributed stage durations can be included into SIR and SEIR models via integro-differential equations (Feng and Thieme 2000a; Hethcote and Tudor 1980; Ma and Earn 2006). Unfortunately, the resulting dynamical systems are mathematically and computationally difficult to study. To avoid the complications involved with integro-differential equations, most research on epidemic models with non-exponentially distributed stage durations has restricted attention to a convenient class of realistic (but not arbitrary) distributions, namely Gamma distributions with integer shape parameter (also known as *Erlang distributions*) (Anderson and Watson 1980; Bailey 1964; Lloyd 2001a; Ma and Earn 2006). The idea is to exploit the fact that the sum of a sequence of independent exponentially distributed random variables is Gamma distributed (Therrien and Tummala 2011). If we break up the infectious stage into a sequence of n substages, each exponentially distributed with mean $1/(n\gamma)$, then the full infectious period distribution will be the Erlang distribution with shape parameter n and scale parameter $n\gamma$, Erlang(n, $n\gamma$).

The resulting SI^{n}R model can then be represented by a simple system of ODEs:

$$\frac{dS}{dt} = \nu N_0 - \beta SI - \mu S\,, \tag{2.7a}$$
$$\frac{dI_1}{dt} = \beta SI - (n\gamma + \mu)I_1\,, \tag{2.7b}$$
$$\frac{dI_2}{dt} = n\gamma I_1 - (n\gamma + \mu)I_2\,, \tag{2.7c}$$
$$\vdots$$
$$\frac{dI_n}{dt} = n\gamma I_{n-1} - (n\gamma + \mu)I_n\,, \tag{2.7d}$$

and the basic reproduction number is (Feng et al. 2007)

$$\mathcal{R}_0^{\mathrm{SI}^n\mathrm{R}} = \frac{\nu N_0}{\mu}\frac{\beta}{n\gamma + \mu}\sum_{j=0}^{n-1}\left(\frac{n\gamma}{n\gamma + \mu}\right)^j. \tag{2.8}$$

Note that the number of individuals in the infectious stage (I) is the sum of all individuals currently in each infectious substage,

$$I = \sum_{j=1}^{n} I_j\,. \tag{2.9}$$

Division into n subclasses is purely a mathematical device and has no biological meaning.

The two extreme cases of the SI^{n}R model occur for $n = 1$, in which case the model reduces to the standard SI^{1}R model (**equation** (2.3)), and the limit as $n \to \infty$, which yields a fixed infectious period of $\tau = 1/\gamma$ (**equation** (2.1), see also **Figure 2.1**), (i.e., the infectious period has a Dirac delta distribution $\delta(t - \tau)$: all individuals who become infectious at time t recover at exactly time $t+\tau$). In this limit, the system becomes a delay differential equation, which can be seen directly as follows. Since the incidence rate at time t is $\beta(t)S(t)I(t)$ and the probability that an individual alive at time t survives to time $t + \tau$ is $e^{-\mu\tau}$ (Hethcote and Tudor 1980), the recovery rate at time t is $\beta(t-\tau)e^{-\mu\tau}S(t-\tau)I(t-\tau)$. Thus, in the limit

$n \to \infty$, the SI^{n}R model approaches the system

$$\frac{dS}{dt} = \nu N_0 - \beta(t)S(t)I(t) - \mu S(t)\,, \tag{2.10a}$$

$$\frac{dI}{dt} = \beta(t)S(t)I(t) - \beta(t-\tau)e^{-\mu\tau}S(t-\tau)I(t-\tau) - \mu I(t)\,, \tag{2.10b}$$

$$\frac{dR}{dt} = \beta(t-\tau)e^{-\mu\tau}S(t-\tau)I(t-\tau) - \mu R(t)\,. \tag{2.10c}$$

We obtain the SE^{m}I^{n}R model (with mean latent period $\tau_E = 1/\sigma$ and mean infectious period $\tau_I = 1/\gamma$) by subdividing the exposed class into m subclasses (Keeling and Grenfell 2002; Nguyen and Rohani 2008),

$$\frac{dS}{dt} = \nu N_0 - \beta SI - \mu S\,, \tag{2.11a}$$

$$\frac{dE_1}{dt} = \beta SI - (m\sigma + \mu)E_1\,, \tag{2.11b}$$

$$\frac{dE_2}{dt} = m\sigma E_1 - (m\sigma + \mu)E_2\,, \tag{2.11c}$$

$$\vdots$$

$$\frac{dE_m}{dt} = m\sigma E_{m-1} - (m\sigma + \mu)E_m\,, \tag{2.11d}$$

$$\frac{dI_1}{dt} = m\sigma E_m - (n\gamma + \mu)I_1\,, \tag{2.11e}$$

$$\frac{dI_2}{dt} = n\gamma I_1 - (n\gamma + \mu)I_2\,, \tag{2.11f}$$

$$\vdots$$

$$\frac{dI_n}{dt} = n\gamma I_{n-1} - (n\gamma + \mu)I_n\,. \tag{2.11g}$$

In the limit that $m \to \infty$ and $n \to \infty$ we obtain the delay differential equation (Keeling and Grenfell 2002),

$$\frac{dS}{dt} = \nu N_0 - \beta(t)S(t)I(t) - \mu S(t)\,, \tag{2.12a}$$

$$\frac{dE}{dt} = \beta(t)S(t)I(t) - \beta(t-\tau_E)e^{-\mu\tau_E}S(t-\tau_E)I(t-\tau_E) - \mu E(t)\,, \tag{2.12b}$$

$$\frac{dI}{dt} = \beta(t-\tau_E)e^{-\mu\tau_E}S(t-\tau_E)I(t-\tau_E) - \tag{2.12c}$$

$$- \beta(t-\tau_I)e^{-\mu\tau_I}S(t-\tau_I)I(t-\tau_I) - \mu I(t)\,, \tag{2.12d}$$

$$\frac{dR}{dt} = \beta(t-\tau_I)e^{-\mu\tau_I}S(t-\tau_I)I(t-\tau_I) - \mu R(t)\,. \tag{2.12e}$$

For the $\mathrm{SE}^m\mathrm{I}^n\mathrm{R}$ model $\mathcal{R}_0$ can be computed by the following formula (Feng et al. 2007):

$$\mathcal{R}_0^{\mathrm{SE}^m\mathrm{I}^n\mathrm{R}} = \frac{\nu N_0}{\mu}\left(\frac{m\sigma}{m\sigma+\mu}\right)^m \frac{\beta}{n\gamma+\mu}\sum_{j=0}^{n-1}\left(\frac{n\gamma}{n\gamma+\mu}\right)^j . \tag{2.13}$$

Note that the basic reproduction number is the product of the mean transmission rate and the mean duration of infectiousness. Altering the distribution of the infectious period alters the probability that an infectious host will die before transmitting. This death-delay interaction changes the mean duration of infectiousness, which results in the differences in the formulae for $\mathcal{R}_0$ above. For diseases of short duration such as measles and smallpox, the mean host lifetime is much longer than the duration of infectiousness. Consequently, $\mu \ll \sigma$ and $\mu \ll \gamma$, so $\mathcal{R}_0$ can always be written:

$$\mathcal{R}_0 \approx \frac{\nu N_0}{\mu}\frac{\beta}{\gamma}\,. \tag{2.14}$$

2.2.3 Seasonal forcing

Exponential distribution of disease stage durations is one unrealistic assumption used in standard SIR-type models. Another is treating the transmission rate β as a constant. More realistic seasonally forced models are implemented by allowing

the transmission rate to vary periodically with a period of one year. The two most commonly used seasonal patterns are *sinusoidal forcing* (Aron and Schwartz 1984; Grossman 1980; Keeling and Grenfell 2002; Olsen and Schaffer 1990),

$$\beta(t) = \beta_0(1 + \alpha \cos(2\pi t)), \tag{2.15}$$

with mean β_0 and amplitude α ($0 \leq \alpha \leq 1$) and *term-time forcing* (Bauch and Earn 2003b; Earn et al. 2000a; Finkenstadt and Grenfell 2000; Keeling and Grenfell 2002; Nguyen and Rohani 2008; Schenzle 1984)

$$\beta(t) = \begin{cases} \beta_{\mathrm{H}} & \text{school days,} \\ \beta_{\mathrm{L}} & \text{non-school days,} \end{cases} \tag{2.16}$$

where $\beta_{\mathrm{H}} > \beta_{\mathrm{L}}$ (the transmission rate is high when school is in session and low otherwise). Earn et al. (2000a) found that the qualitative dynamics of the term-time forced $\mathrm{SE^1I^1R}$ model are essentially equivalent to the dynamics of the sinusoidally forced $\mathrm{SE^1I^1R}$ model but with lower seasonal amplitude, α. The same is true for the $\mathrm{SI^1R}$ model. Since our focus is on qualitative dynamics, we use sinusoidal forcing for simplicity.

In the next section (**Section 2.3**) we will discuss in detail the *transition analysis* used to predict changes in the qualitative pattern of infectious disease time series. In **Section 2.4**, we use the Erlang distributed $\mathrm{SI}^n\mathrm{R}$ and $\mathrm{SE}^m\mathrm{I}^n\mathrm{R}$ models described above to understand how the shape of the stage duration distributions influences predictions of transition analysis.

2.3 Predicting epidemiological transitions

Many infectious disease time series display occasional, rapid changes in qualitative dynamics, such as transitions from annual to biennial cycles or to irregular dynamics (Anderson and May 1991; Olsen and Schaffer 1990). Previous work has shown that these transitions appear to be driven by demographic and behavioural changes since they induce bifurcations in the SE^1I^1R model (Bauch and Earn 2003b; Earn 2009; Earn et al. 2000a). We would like to know whether the qualitative inferences made previously based on the SE^1I^1R model remain valid when the analysis is repeated with more realistic SE^mI^nR models.

Earn et al. (2000a) used the SE^1I^1R model to show that knowing the changes in birth and vaccination rates—or, more generally, changes in the rate at which susceptible individuals are recruited into the population—it is possible to predict the occurrence of bifurcations that change the period of epidemic cycles. We briefly revisit that argument here in the more general context of the SE^mI^nR model.

2.3.1 Theoretical motivation for transition analysis

In **equation** (2.11a), the term (ν) was formulated as the birth rate but can be thought of more generally as the susceptible recruitment rate. Suppose that this rate changes to ν', which might occur because the birth rate has changed or because we have begun to vaccinate a proportion p of the population (in which case $\nu' = \nu(1-p)$). To understand the dynamical effect of this change from ν to ν', consider the following simple change of variables:

$$S' = \frac{\nu}{\nu'}S\,, \quad E'_j = \frac{\nu}{\nu'}E_j\,, \quad I'_k = \frac{\nu}{\nu'}I_k\,, \quad \text{for } 1 \le j \le m,\ 1 \le k \le n. \quad (2.17)$$

If we insert these expressions in **equation** (2.11) and solve for the equations

for the primed variables we obtain, for example,

$$\frac{dS'}{dt} = \nu N_0 - \beta\frac{\nu'}{\nu}S'I' - \mu S' \,. \tag{2.18}$$

That is, the equations for the primed variables are identical to the original equations (with the original susceptible recruitment term ν), but with the transmission rate changed from β to $\beta\nu'/\nu$. Thus the dynamical effect of a change in susceptible recruitment by a given factor is identical to the dynamical effect of changing the transmission rate by exactly that factor,

$$\nu \to \frac{\nu'}{\nu} \quad \Longrightarrow \quad \beta \to \beta\frac{\nu'}{\nu} \text{ and } \mathcal{R}_0 \to \mathcal{R}_0\frac{\nu'}{\nu} \,. \tag{2.19}$$

Consequently, we can use a bifurcation diagram with the transmission rate β, or equivalently the basic reproduction number $\mathcal{R}_0$ (since $\mathcal{R}_0$ is proportional to β), as the control parameter to predict transitions in dynamical behaviour induced by changes in susceptible recruitment rate. **Figure 2.2** shows such a bifurcation diagram based on the sinusoidally forced SI^{1}R model (**equation** (2.3)) with parameters chosen to correspond to measles (and with an estimated value of $\mathcal{R}_0 = 17$ at some given time, say t_0, marked with a dotted vertical line). If the susceptible recruitment rate was ν_0 at time t_0 and ν_1 at time t_1 then we would predict that at time t_1 the system would behave as if the basic reproduction number had changed by the factor ν_1/ν_0, i.e. the *effective* reproduction number at time t, is

$$\mathcal{R}_{0,\text{eff}} = \mathcal{R}_0\frac{\nu_1}{\nu_0} \,. \tag{2.20}$$

There is an important subtlety upon which our ability to predict transitions depends critically. In the equation for dS/dt (**equations** 2.3a, 2.7a, 2.11a, 2.10a, 2.12a) the susceptible recruitment rate appears as a constant (ν does not depend explicitly on time t or population size N) and we use mass-action incidence (βSI)

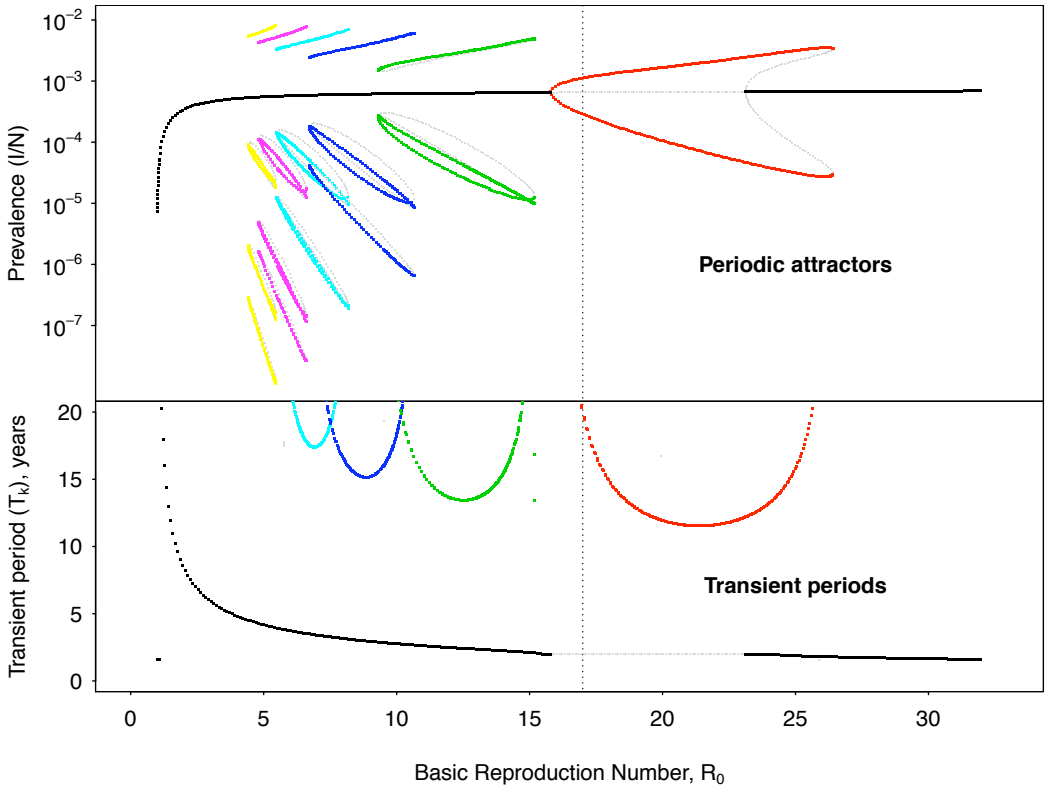


Figure 2.2: Asymptotic and perturbation analysis of the sinusoidally forced SI^{1}R model (2.3,2.15) parameterized for measles ($\gamma^{-1} = 13$ days, $\nu = 0.02$ yr^{-1}, $\alpha = 0.08$). The **top panel** (asymptotic analysis) shows the bifurcation diagram for the model with control parameter $\mathcal{R}_0$. The ordinate shows the proportional prevalence of infection at the start of each year, so annual cycles (black) are indicated by a single point at each $\mathcal{R}_0$, biennial cycles (red) by two points, triennial cycles (green) by three, and so on. Coloured curves correspond to stable cycles while grey curves indicate unstable cycles. A dotted vertical line is drawn at $\mathcal{R}_0 = 17$, indicating the estimate of the basic reproduction number at the "anchor time" t_0. Two types of bifurcations occur in this diagram: period doublings (also called pitchforks or flips) and tangent bifurcations (also called folds or saddle-node bifurcations). The **bottom panel** shows the natural period of damped oscillations (the transient period) onto each attractor, as described in step 2 of §2.3.2. The transient period curves are coloured according to the corresponding attractor in the top panel. The grey line indicates a region where the annual cycle is unstable.

rather than standard incidence ($\beta SI/N$). If the susceptible recruitment term were taken to be νN rather than νN_0, and we were to use standard incidence then the

variable change in **equation** (2.17) would have no effect (the differential equations are invariant to the scaling transformation given by **equation** (2.17)) and we would never predict dynamical transitions resulting from changes in the susceptible recruitment rate. One can debate on theoretical grounds whether one model formulation or another is most plausible biologically (Heesterbeek and Metz 1996); we favour our formulation because it leads to correct predictions concerning dynamical transitions (Bauch and Earn 2003b; Earn et al. 2000a). We are interested in the effects of changes in ν over time, but the changes of interest occur slowly compared with the epidemic time scale, which is why we can treat ν as constant in the dS/dt equation.

2.3.2 The method of transition analysis

Given a time series of reported disease incidence or mortality (for a disease for which we have estimates of the mean latent and infectious periods), a full *transition analysis* proceeds as follows (Bauch and Earn 2003b; Earn et al. 2000a). First, in order to clarify what needs to be explained, plot the disease time series together with its estimated frequency structure at each time point (e.g., Fourier power spectra for subsets of the full time series or, preferably, a wavelet spectrum for the full time series (Bauch 2008; Grenfell et al. 2001)). Second, for some "anchor time" t_0 in the time series, obtain an estimate of the basic reproduction number $\mathcal{R}_0$, preferably using data other than the focal time series (e.g., annual age-specific data (Anderson and May 1991)). Third, estimate the susceptible recruitment rate ν at each point of the disease time series and infer the effective reproductive number $\mathcal{R}_{0,\text{eff}}$ at all times by inserting the estimated ν values into **equation** (2.20) (where $\nu_0 = \nu(t_0)$ and $\nu_1 = \nu(t)$ for an arbitrary time t). Fourth, identify time intervals during which ν is roughly constant (hence during which the dynamical features of the disease time series can be expected to be approximately stationary). Finally, based on the estimated value of $\mathcal{R}_{0,\text{eff}}$ in each of the "dynamically stationary time intervals",

predict transitions in qualitative dynamical behaviour (e.g., changes in the structure of the wavelet spectrum, especially the positions of peaks), as follows.

1. ***Asymptotic analysis*** (to identify the periods of *attractors* of the model, which are reached asymptotically) (Earn et al. (2000a); Bauch and Earn (2003a,b); Nguyen and Rohani (2008)): Construct a bifurcation diagram with $\mathcal{R}_0$ as the control parameter, over a range of $\mathcal{R}_0$ that includes the value estimated for time t_0 and the full range of $\mathcal{R}_{0,\text{eff}}$ determined via **equation** (2.20) (e.g., top panel of **Figure 2.2**). From this diagram, we can easily infer the periods of cyclical attractors of the system. We call these *resonant periods* because they are exact subharmonics (i.e., integer multiples) of the period of seasonal forcing (one year). (See **Appendix 2.6** for a step-by-step guide to creating diagrams like **Figure 2.2** using XPPAUT (Ermentrout 2002).)

2. ***Perturbation analysis*** (to estimate the periods of the *transients* associated with each attractor): Over the same range of $\mathcal{R}_0$ as in the asymptotic analysis, plot the periods of the transients associated with—i.e., the periods of damped oscillations onto—each cyclical attractor (e.g., bottom panel of **Figure 2.2**). We call these *non-resonant periods* because they can take any real value and are not entrained by seasonal forcing. Non-resonant periods may be detected in observed epidemic time series because transients can be sustained by demographic stochasticity (Bartlett 1957b; Bauch and Earn 2003b). Non-resonant periods can be calculated by linearizing about the fixed points and cycles of the model's one-year-stroboscopic map (Bauch and Earn 2003a,b). If the period of a given attractor is k and the dominant eigenvalue of the associated k-cycle of the stroboscopic map is λ_k (which is complex for typical disease parameters) then the associated transient period is

$$T_k = \frac{2\pi k}{|\text{Arg}(\lambda_k)|}\,. \tag{2.21}$$

3. ***Stochastic analysis*** (to estimate the relative importance of transient versus asymptotic dynamics): The wavelet spectrum has peaks at the most important periods in the time series (which we attempt to predict with steps 1 and 2 above) but also shows the magnitude of the peaks, which cannot be estimated by asymptotic and perturbation analysis of a deterministic model. The relative magnitudes of spectral peaks of observed time series can be estimated from spectra of simulations of stochastic realizations of the model, with the expectation that smaller population sizes (which are subject to greater demographic stochasticity) will stimulate more transient dynamics, leading to larger spectral peaks at non-resonant periods (Bauch and Earn 2003b; Earn 2009). Because the stochastic analysis addresses the details rather than the main features of dynamical transitions, we do not conduct it in this paper (though we make occasional reference to stochastic effects). We note, however, that understanding these details is an area of very active research, and powerful analytical approaches for estimating power spectra for recurrent epidemic processes have been developed recently (Alonso et al. 2007; Black and McKane 2010a,b; Lima 2009). Ultimately, a complete transition theory would need to account for all the dynamical characteristics of stochastic epidemic models, which include alternation between asymptotic and transient behaviour (Bauch and Earn 2003b), switching between different attractors (Earn et al. 2000a; Schwartz and Smith 1983), phase-locked cycles at one fixed period (He and Earn 2007) and interactions with repellors (Rand and Wilson 1991).

In **Section 2.4** we use the SI^nR and SE^mI^nR models to conduct transition analysis of the well known New York City measles time series (Yorke and London 1973). Our main question is: Do we predict different transitions if we base our theoretical analysis on the SI^nR rather than the SI^1R model, or the SE^mI^nR rather

than SE^1I^1R model?

Another question that we will address is: Can we approximate the dynamics of the SE^mI^nR model using the SI^nR model? This question is motivated by the fact that the dynamics of the SE^1I^1R model can be approximated using the SI^1R model. It is well-known that the equilibrium and stability properties (e.g., the period of damped oscillations onto the equilibrium) of the unforced SI^1R and SE^1I^1R models correspond if the mean infectious period in the SI^1R model is associated with the sum of the mean latent and mean infectious periods in the SE^1I^1R model (p. 668, Anderson and May (1991)). The measles bifurcation diagram shown in **Figure 2.2** for the sinusoidally forced SI^1R model is virtually identical to the term-time forced SE^1I^1R measles bifurcation diagram produced previously by Earn et al. (2000a). Therefore, we analyze the SI^nR model with mean infectious period $1/\gamma = 13$ days (the sum of the real mean latent period of 8 days and the real mean infectious period of 5 days for measles).

2.4 Transition analysis using SI^nR and SE^mI^nR models

In this section, we use the well-known measles incidence time series for New York City (1928–1972) as an illustrative example with which to compare the results of transition analysis using SI^nR and SE^mI^nR models with stage duration distributions varying from exponential to fixed. The New York City measles data were originally digitized and studied by London and Yorke (London and Yorke 1973; Yorke and London 1973). Previous transition analysis of these data (Bauch and Earn 2003b; Earn et al. 2000a) has been restricted to the pre-vaccine period (up to 1963). Here, we are able to extend our analysis to 1972 using vaccination data for 1963–1972 (see **Appendix 2.6**).

2.4.1 Description of the data

Reported incidence and inferred frequency structure

The upper panel of **Figure 2.3** shows monthly reported cases of measles in New York City (together with estimated susceptible recruitment rate) and the lower panel shows the frequency structure of the data over time as a wavelet spectrum. Two spectral peaks are evident for the full duration of the time series, one at a period of one year and a second at a period that changes over time (2–3 years from 1928 to about 1946, exactly 2 years from about 1946 to 1965, and 2–4 years from about 1965 to the end of 1972).

Estimated susceptible recruitment

Based on age-incidence and age-seroprevalence data for England and Wales (1950–1968), the basic reproduction number for measles has been estimated to be $\mathcal{R}_0 \simeq 17$ in a pre-vaccination era (Anderson and May 1991, Figures 3.9 and 3.10, and Table 4.1, p. 70). Since in New York City the birth rate was approximately the same as in England and Wales (also in pre-vaccination era), we use this value as an estimate for $\mathcal{R}_0$ in New York City in 1960, which we take to be our *"anchor time"* t_0.

Measles vaccine was introduced in the United States in 1963 (CDC), so susceptible recruitment until 1963 can be taken to be associated entirely with births. However, newborns do not enter the well-mixed susceptible pool immediately, for two reasons: (i) maternally acquired immunity can take up to a year to wane (Anderson and May 1991, p. 50), (ii) before entering pre-school, children typically have much lower contact rates with other susceptibles. Hence the impact of changes in birth rate on transmission dynamics is delayed, approximately by the time between birth and entering the well-mixed susceptible pool. We took this delay, τ_S, to be 2 years, but our conclusions are not sensitive to this parameter (e.g., taking it to be 0 or 5 years makes little difference (light red/blue curves in **Figure 2.3**)). Notice

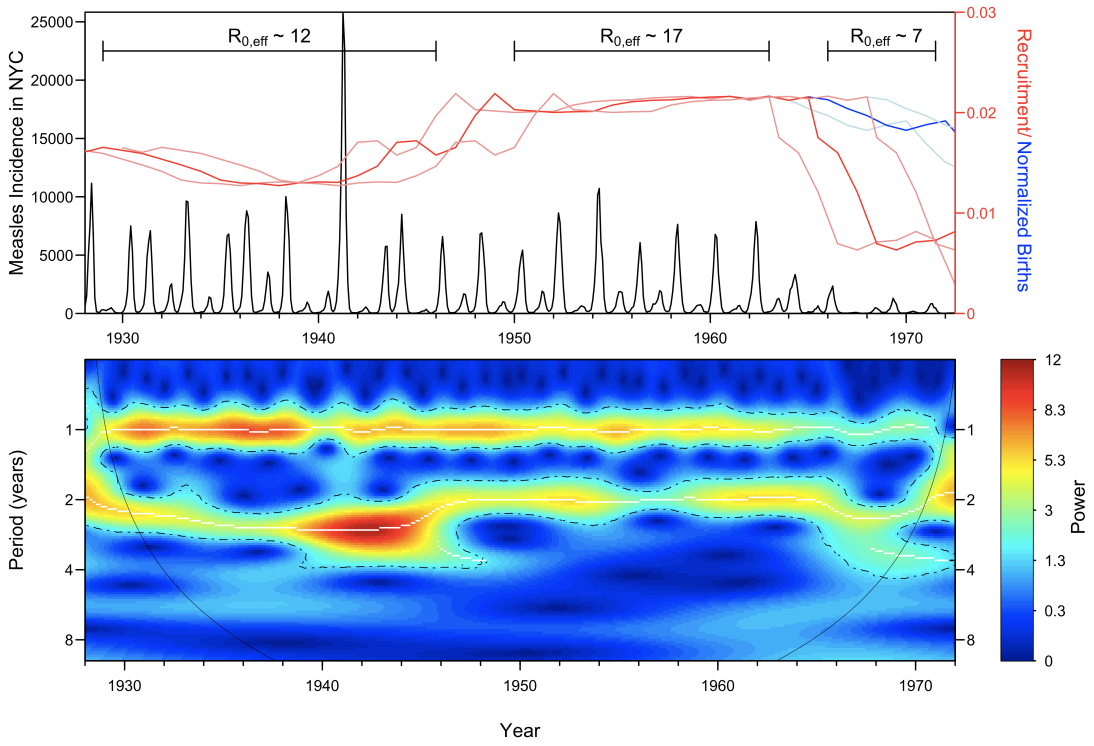


Figure 2.3: Measles in New York City, 1928–1972. The **upper panel** shows monthly reported measles incidence (black) and annual susceptible recruitment relative to the population size in 1960 (red). Before the introduction of vaccination in 1963, annual susceptible recruitment coincided with annual births (blue). We shift the recruitment curve forward by 2 years to account for the delay between birth and entering the well-mixed population. Light red and light blue curves show the susceptible recruitment rate without delay and with a delay of 5 years. The line segments at the top of the upper panel highlight time intervals with distinct effective $\mathcal{R}_0$, calculated with **equation** (2.20). The **lower panel** shows the wavelet power spectrum of the measles incidence time series (log-transformed and normalized to unit variance). The white curves show the local maxima of wavelet power (squared modulus of wavelet coefficients (Cazelles et al. 2008, p. 291)) at each time. The dot-dashed curves indicate 95% confidence region, estimated from 1000 bootstrapped time series generated by the method of (Cazelles et al. 2008, pp. 292–293). Below the "cone of influence" (Torrence and Compo (1998);(Cazelles et al. 2008, p. 291)), the calculation of wavelet power is less accurate because it includes edges of the time series that have been zero-padded to make the length of the series a power of 2. The wavelet spectrum was computed using `MATLAB` code kindly provided by Bernard Cazelles (Cazelles et al. 2007, 2008; Torrence and Compo 1998).

that τ_S should be < 5 since the mean age at infection is 5 years (Anderson and May 1991, Fig. 8.1, p. 156). Thus, we take the susceptible recruitment rate in 1960 to be the ratio of the number of births in 1958 ($B(t_0 - \tau_S) = 167,660$) to the estimated population of New York City in 1960 ($N_0 = 7,781,984$), i.e., $\nu(t_0) \simeq 0.02$. At other times t,

$$\nu(t) = \frac{B(t-\tau_S)}{N_0}\big(1 - p(t-\tau_S)\big)\,, \tag{2.22}$$

where $p(t)$ is the proportion of new recruits at time t who were vaccinated before entering the well-mixed susceptible pool. Note in **equation** (2.22) that it is N_0, not $N(t)$ that appears: recruitment is normalized relative to the population size at the "anchor time" t_0 (Earn et al. 2000a). After 1963, the susceptible recruitment rate is substantially reduced by the introduction of vaccination (**Figure 2.3**).

The birth and measles vaccination data that we insert in **equation** (2.22) are discussed in **Appendix 2.6**. The resulting annual susceptible recruitment rate is shown in the top panel of **Figure 2.3**. There are three distinct periods during which the recruitment rate was roughly constant: 1929–1946 with $\nu \approx 0.015$, 1950–1963 with $\nu \approx 0.02$, and 1966–1971 with $\nu \approx 0.008$. Therefore, from **equation** (2.20), we estimate the effective reproduction number to be $\mathcal{R}_{0,\text{eff}} \approx 12$ for 1928–1946, $\mathcal{R}_{0,\text{eff}} \approx 17$ for 1950–1963 and $\mathcal{R}_{0,\text{eff}} \approx 7$ for 1966–1971.

2.4.2 Asymptotic and Perturbation analysis

Previous transition analyses of the New York City measles incidence time series were based on the $\text{SE}^1\text{I}^1\text{R}$ model with the mean latent and infectious periods $\tau_E = 8$ days and $\tau_I = 5$ days respectively (Bauch and Earn 2003b; Earn et al. 2000a). Given data from which the full latent and infectious period distributions can be estimated (rather than just their means), it would be sensible to fit Erlang distributions to the actual stage duration distributions and begin the transition analysis from the corresponding $\text{SE}^m\text{I}^n\text{R}$ model. For example, Wearing et al. (2005) used measles

case data from Gloucestershire, UK, for the period 1947–51 (Hope-Simpson 1952) to estimate $\tau_E = 8$ days with the shape parameter $m \approx 20$ and $\tau_I = 5$ days with the shape parameter $n \approx 20$. Even in situations in which only the means of the stage duration distributions can be estimated, an SE^mI^nR model (with $m > 1$ and $n > 1$) is likely to be a more accurate representation of reality than an SE^1I^1R model. So, for example, Keeling and Grenfell (2002) considered an SE^mI^nR model with $m = 8$ and $n = 5$, i.e., one day on average in each latent and infectious substage, as a reasonable improvement of the SE^1I^1R model.

Our primary question, however, is how the predictions of transition analysis vary as a function of stage duration distribution and whether the previous transition analyses based on the SE^1I^1R model have led us to correct or incorrect inferences. We therefore consider the full range of Erlang distributions for the latent and infectious periods and study the SE^mI^nR model with $1 \leq m \leq \infty$ and $1 \leq n \leq \infty$. Note that we chose the *mean* latent and infectious periods to be fixed ($1/\sigma = 8$ days; $1/\gamma = 5$ days). Because our general goal is to evaluate the robustness of dynamical inferences to model structure, we begin by analysing the simpler SI^nR model with $1 \leq n \leq \infty$.

Predictions of the SI^nR model

Asymptotic analysis

Figure 2.4 shows a sequence of SI^nR bifurcation diagrams for various values of the shape parameter ($n = 1, 3, 10, \infty$) together with the corresponding distributions of the infectious period (each with a mean of 13 days). Stable branches are colour-coded according to the period of the attractor. Unstable branches are shown in grey. The case $n = 1$ (top panel) is identical to the top panel of **Figure 2.2**. As n increases from 1 to ∞, each of the branches undergoes further bifurcations. Chaotic attractors (superimposed in light grey) are evident for $n = 10$ and dominate for a substantial range of $\mathcal{R}_0$ for $n = \infty$.

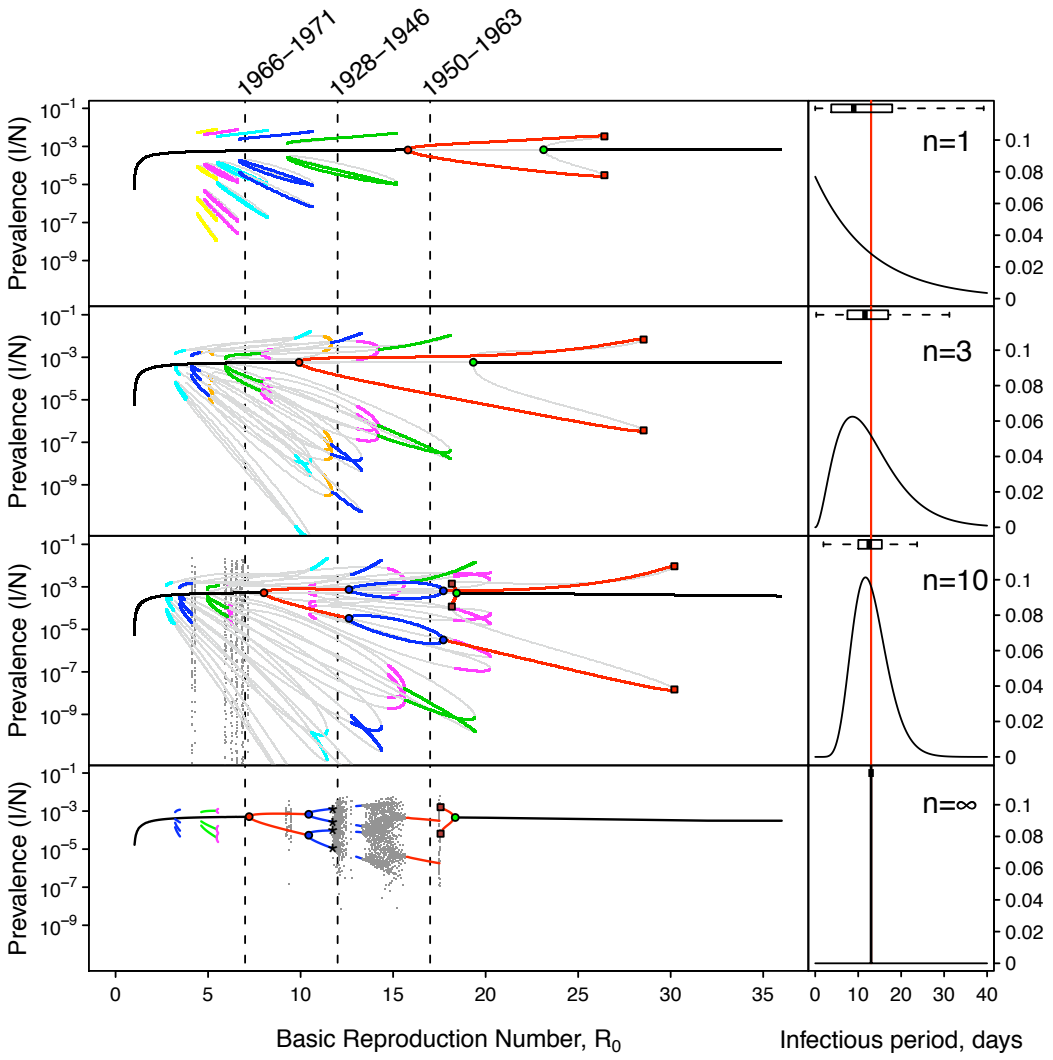


Figure 2.4: SI^{n}R measles bifurcation diagrams as a function of $\mathcal{R}_0$ for several values of the shape parameter of the infectious period distribution ($n = 1,\ 3,\ 10,\ \infty$), with other parameters fixed (mean infectious period $1/\gamma = 13$ days, birth rate $\nu = 0.02$/year, seasonal forcing amplitude $\alpha = 0.08$). Colours represent attractors with different periods. Light grey curves show unstable branches. Circles represent period doubling (flip) bifurcations while squares denote tangent (saddle-node) bifurcations of the main branch. Dashed vertical lines highlight $\mathcal{R}_{0,\text{eff}} = 7$, 12 and 17, which correspond to the estimated effective reproduction number for measles in New York City for the year ranges indicated, as in **Figure 2.3**. Each right panel shows the corresponding probability distribution of the infectious period and a box plot showing the 5%, 25%, 50%, 75% and 95% quantiles and the distribution; a vertical red line shows the mean infectious period (13 days). **(continued on the next page)**

Figure 2.4: For finite n, the bifurcation diagrams were computed using standard continuation software (XPPAUT, Ermentrout (2002)), whereas the fixed-delay limit ($n = \infty$) was computed by "brute force", i.e., by numerical integration of the delay differential **equation** (2.10) until convergence on an attractor; hence unstable branches are not shown in the limit $n = \infty$. Brute force bifurcation diagrams were also computed for finite n to reveal regions of chaotic behaviour which are also shown in light grey.

The vertical dashed dark grey line at $\mathcal{R}_0 = 17$ in **Figure 2.4** corresponds to the estimated basic reproduction number for the year $t_0 = 1960$. The effective reproduction number is also estimated to be 17 throughout the 13 year period t =1950–1963, since the birth rate did not change appreciably during this time and measles vaccine was not yet invented. The other two vertical dashed grey lines at $\mathcal{R}_0 = 7$ and $\mathcal{R}_0 = 12$ correspond, respectively, to the estimated effective reproduction number during the periods t =1928–1946 and t =1966–1971, as computed from **equations** (2.20) and (2.22).

The bifurcation tree of the standard SI^{1}R model ($n = 1$) shows a biennial cycle for $\mathcal{R}_0 = 17$, coexistence of annual and triennial cycles for $\mathcal{R}_0 = 12$, and coexistence of annual and 4- and 5-year cycles for $\mathcal{R}_0 = 7$. Hence, the model correctly predicts the biennial pattern observed from 1950 to 1963 in New York City, but appears at first sight to predict incorrectly that there are multiple coexisting non-annual cycles at other times. However, in the ranges of $\mathcal{R}_0$ for which multiple attractors coexist, and in particular for $\mathcal{R}_0 = 12$ and $\mathcal{R}_0 = 7$, stochastic simulations spend almost all of their time in the basin of the annual attractor (Bauch and Earn 2003b). Thus, the resonant period of one year observed in New York City from 1928 to 1946 and from 1966 to 1971 is also consistent with the SI^{1}R model.

Because of the series of bifurcations that occur rapidly as n is increased, the SI^{n}R model for any $n > 1$ exhibits more complex dynamics than the SI^{1}R model and is harder to reconcile with the observed transitions in New York City measles. More often than the SI^{1}R model, the SI^{n}R model with $n > 1$ has coexisting long-

period stable cycles that are not observed in practice. As with the SI^{1}R model, stochastic simulations can be expected to remain primarily in the vicinity of the "primary" attractor, but unlike the SI^{1}R model, the primary attractor of the SI^{n}R with $n > 1$ often predicts the wrong resonant period for New York City measles. For example, for $n = 10$, the dominant attractor for $\mathcal{R}_0 = 17$ has a period of four years (not two years) and the dominant attractor for $\mathcal{R}_0 = 12$ has period two (not one). In the presence of noise, the four year cycle may be difficult to distinguish from a two-year cycle, but the predicted two-year cycle for $\mathcal{R}_0 = 12$ is nothing like the measles data it ought to explain.

Perturbation analysis

Just as perturbing an orbit away from a stable equilibrium can induce transient, damped oscillations onto the equilibrium, perturbing an orbit away from a periodic attractor can induce transient, damped oscillations onto the stable cycle. Although more cumbersome to calculate for a non-equilibrium attractor (Bauch and Earn 2003b), transient orbits in the vicinity of a periodic attractor have a well-defined characteristic period of oscillation. **Figure 2.5** summarizes the transient dynamics of the SI^{n}R models for $n = 1$, 3 and 10. For each periodic attractor, the non-resonant period, i.e. the period of damped oscillations onto the attractor, is plotted on the y-axis as a function of $\mathcal{R}_0$. The curves are colour-coded according to the colours of the corresponding attractors in **Figure 2.4**. Dotted grey lines are used in ranges of $\mathcal{R}_0$ where the corresponding periodic orbits are unstable; in these regions, the model displays phase-locked transient dynamics at the indicated period (i.e., the transient period is fixed and is the same as the period of the stable attractor), which is a prerequisite for a period-doubling bifurcation (He and Earn 2007).

In the case of the SI^{1}R model, the non-resonant periods associated with all the non-annual attractors are too long to be observable in the New York City measles time series. The non-resonant period associated with the annual attractor

does agree well with the wavelet spectrum shown in **Figure 2.3**. For the SI^{n}R models with $n > 1$, the non-resonant periods associated with multi-year attractors are shorter and often should be observable in principle. For example, for $\mathcal{R}_{0,\text{eff}} = 12$ the SI^{10}R model ($n = 10$) predicts a transient period of 4.5 years. However, it is not observed in the incidence power spectra (**Figure 2.3**). The lack of any indication of non-resonant periods associated with non-annual attractors in the wavelet spectrum for measles in New York City appears to cast further doubt on the usefulness of the SI^{n}R model for measles.

Summary of SI^{n}R transition analysis

Overall, from the point of view of measles transition analysis, the SI^{1}R model is just as successful as the SE^{1}I^{1}R model studied previously (Bauch and Earn 2003b; Earn et al. 2000a). However, the SI^{n}R model with $n > 1$ is far less successful; as n increases the dynamical structure of the model becomes more and more complex and the predicted resonant and non-resonant periods stray further and further from the observed spectral peaks in the New York City measles time series.

The upper panel of **Figure 2.6** summarizes our asymptotic analyses of the full sequence of SI^{n}R measles models ($n = 1$ to ∞) with a two-parameter $(\mathcal{R}_0, n)$ bifurcation diagram for the *main branch* of the bifurcation tree in **Figure 2.4**. The boundaries of the colour-coded regions in **Figure 2.6** correspond to the major bifurcation points highlighted with circles (for flips) and squares (for saddle-nodes) in **Figure 2.4**. As $n \to \infty$ (i.e., as the infectious period distribution approaches a delta function), the main branch of the bifurcation tree undergoes a period doubling cascade in the grey region ($\mathcal{R}_0 \sim 12$–15). The lower panel of **Figure 2.6** also describes the $(\mathcal{R}_0, n)$ plane, but shows contours of constant non-resonant periods associated with the annual cycle on the main branch (this is the most likely non-resonant period to be observable since it is the shortest; see **Figure 2.5**). The hatched region is characterized by phase-locked transient dynamics at a period of two years.

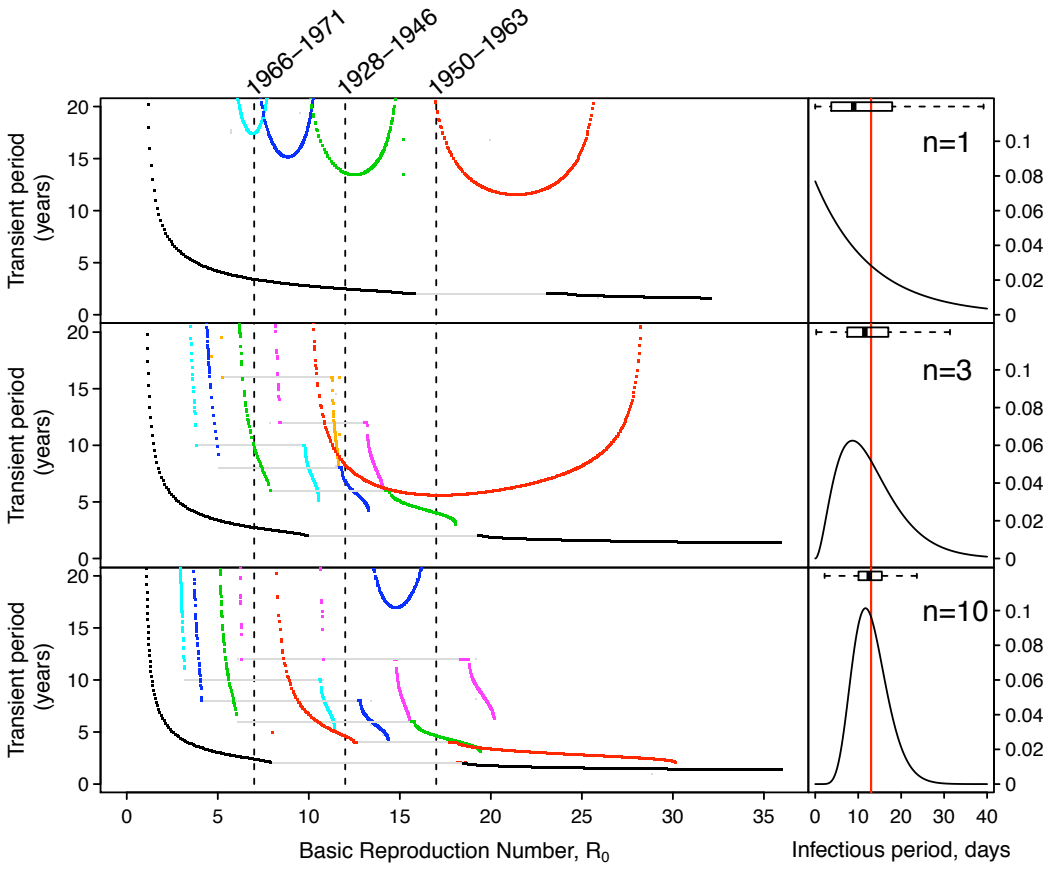


Figure 2.5: Transient dynamics of the measles SI^{n}R model for $n = 1$, 3 and 5 as a function of $\mathcal{R}_0$. This figure complements **Figure 2.4**. Each panel shows the transient periods associated with the periodic attractors shown in the same colour in **Figure 2.4** (e.g., the black curve shows the transient period associated with the annual attractor). The light grey lines show ranges of $\mathcal{R}_0$ where the associated periodic cycles exist but are unstable. Dashed vertical lines correspond to the values of $\mathcal{R}_{0,\text{eff}} = 7$, 12 and 17. As in **Figure 2.4**, the right panels show the associated infectious period distribution.

Note that because n is a discrete parameter it cannot be used as a continuation parameter in XPPAUT, hence we had to resort to separate continuation analyses for each n. The sequence of main-branch bifurcation diagrams that we constructed for the SI^{n}R measles model (using 24 values of n from 1 to ∞) is shown in **Appendix 2.6**.

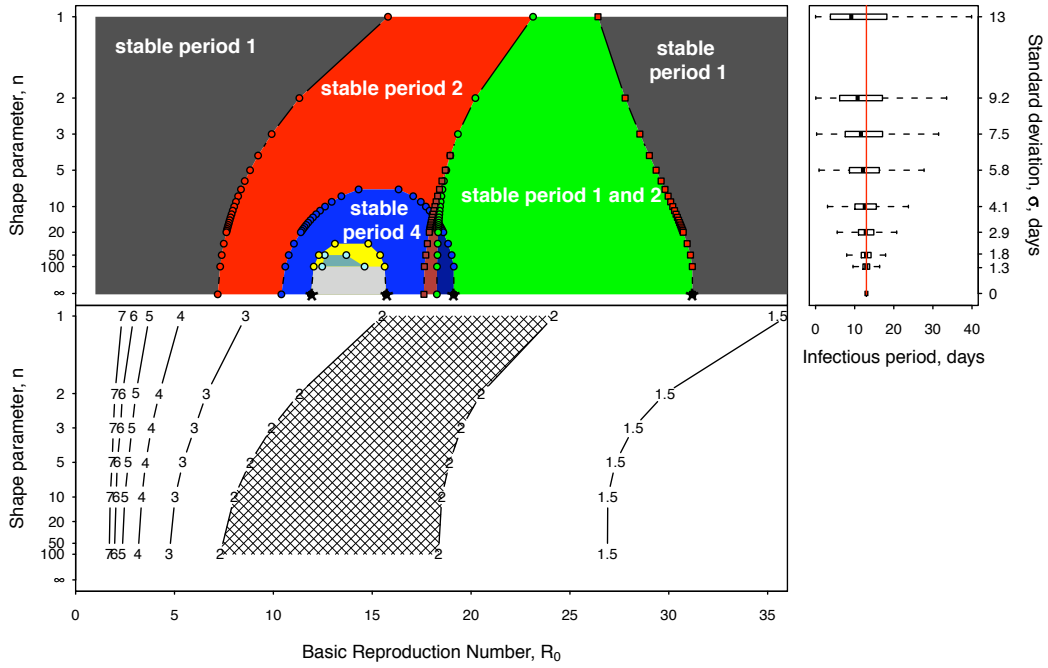


Figure 2.6: Dynamical structure of the SI^{n}R model with a mean infectious period $1/\gamma = 13$ days. *Upper panel:* Two-parameter ($\mathcal{R}_0$ vs n) bifurcation diagram corresponding to the main branch of the one-parameter bifurcation diagrams shown in **Figure 2.4**. Circles represent period doubling (flip) bifurcations while squares denote tangent (fold) bifurcations as in **Figure 2.4**. Regions of different colours indicate different asymptotic dynamics on the main branch: a single annual attractor (dark grey), a single biennial attractor (red), a single four-year attractor (blue), a single eight-year attractor (yellow), a single sixteen-year attractor (light blue), co-existence of annual and biennial attractors (green), co-existence of two distinct biennial attractors or co-existence of biennial and four-year attractors (brown), and coexistence of annual and four-year attractors (dark blue). In the light grey region, there are cascades of further period doublings that appear to end in chaos as $n \to \infty$. The stars indicate bifurcation points that were estimated by extrapolation to $n = \infty$ rather than by direct calculations based on the fixed delay model, **equation** (2.10). *Lower panel:* Contours of constant transient period (associated with the annual cycle) in the $(\mathcal{R}_0, n)$ plane (*cf.* black curves in the bottom panel of **Figure 2.2** and in each panel of **Figure 2.5**). In the hatched region, the transient period is phase-locked at precisely two years (He and Earn 2007), whereas the transient period changes smoothly between the other contours.

Predictions of the SE^{m}I^{n}R model

We now apply precisely the same analyses to the more realistic SE^{m}I^{n}R models. **Figures 2.7–2.9** for the SE^{m}I^{n}R models correspond to **Figures 2.4–2.6** for the SI^{n}R models.

Since we are now modelling both the latent and infectious stages directly, we can use accepted estimates for their mean durations (mean latent period $1/\sigma = 8$ days, mean infectious period $1/\gamma = 5$ days (Wearing et al. 2005). In addition, we now have two shape parameters (m for the latent stage and n for the infectious stage). We examine several illustrative m, n values studied previously in the literature: $m = 1, n = 1$ (Anderson and May 1991; Earn et al. 2000a), $m = 8, n = 5$ (Keeling and Grenfell 2002) and $m = 20, n = 20$ (Wearing et al. 2005).

Figure 2.7 presents asymptotic analysis of the SE^{m}I^{n}R model. The bifurcation structure of the model changes as m and n are increased, but the changes are less substantial than **Figure 2.4** shows as n is increased in the SI^{n}R model. **Figure 2.8** presents the results of perturbation analysis of the SE^{m}I^{n}R model. Again, narrowing the stage duration distributions alters the transient periods, but less than **Figure 2.5** shows for the SI^{n}R model.

The degree of dependence of SE^{m}I^{n}R dynamics on stage duration distributions is clearest from the two-parameter bifurcation diagrams and transient-period contour plots shown in **Figure 2.9**, which should be compared with **Figure 2.6** for the SI^{n}R model. Regardless of the shapes of the stage duration distributions, the predicted resonant and non-resonant periods are very similar. Regardless of m and n, for $\mathcal{R}_0 = 17$ we predict a resonant period of two years and an unobservably long non-resonant period (> 7 years), for $\mathcal{R}_0 = 12$ we predict a one-year resonant period and a 2–3 year non-resonant period, and for $\mathcal{R}_0 = 7$ we predict a one-year resonant and 3–4 year non-resonant period. Consequently, transition analysis based on any of these SE^{m}I^{n}R models is consistent with the New York City measles time series and wavelet spectrum (**Figure 2.3**) as well as for the other measles time series

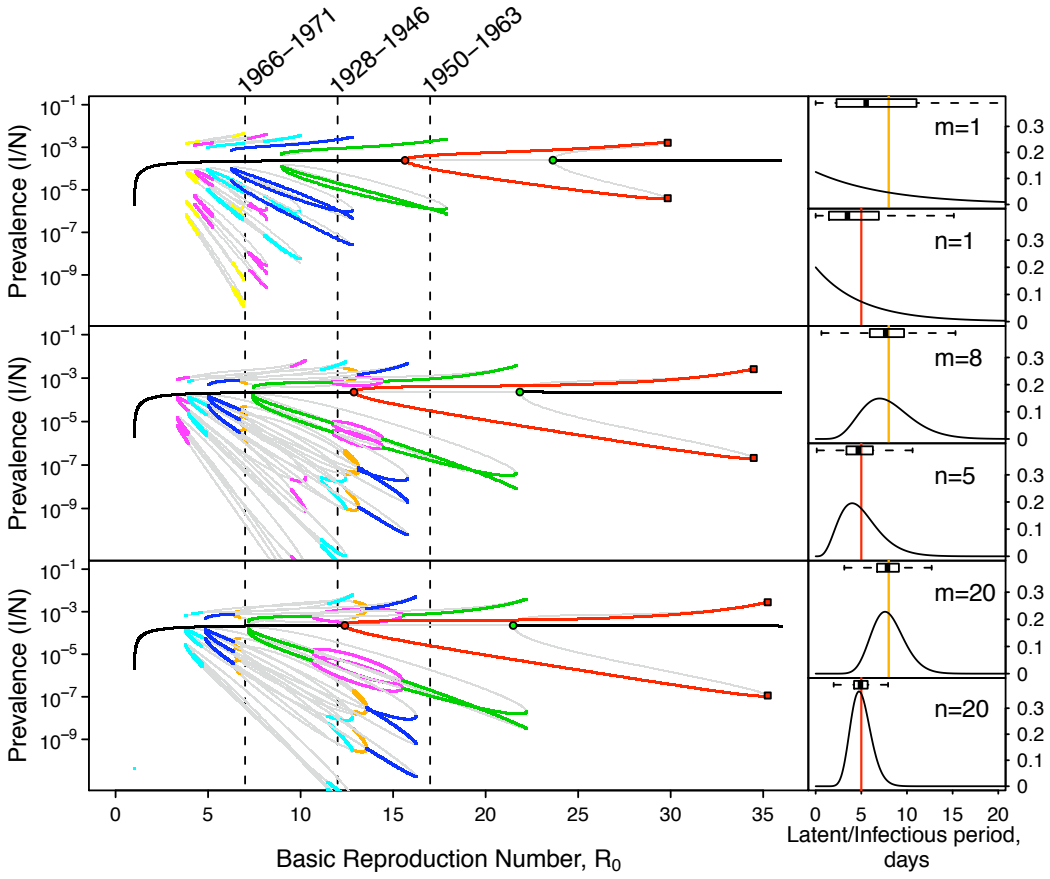


Figure 2.7: SE^{m}I^{n}R bifurcation diagrams as a function of $\mathcal{R}_0$ for several values of the shape parameters of the latent and infectious period distributions. The mean stage durations are chosen to correspond to measles (mean latent period $1/\sigma = 8$ days, mean infectious period $1/\gamma = 5$ days). The other fixed parameters are the birth rate ($\nu = 0.02$/year) and the amplitude of (sinusoidal) seasonal forcing ($\alpha = 0.08$). Colours represent attractors with different periods. Light grey curves indicate unstable branches. Circles represent period doubling (flip) bifurcations while squares denote tangent (fold) bifurcations on the main branch. Dashed vertical lines highlight $\mathcal{R}_{0,\text{eff}} = 7$, 12 and 17 for year ranges indicated in **Figure 2.3**. The right panels show probability densities and box plots of the latent and infectious periods, with means highlighted by vertical lines (orange for latent period, red for infectious period).

considered previously (Bauch and Earn 2003a,b; Earn et al. 2000a).

We are led to conclude that transition analysis is robust to the shapes of the distributions of the latent and infectious periods (provided we include both).

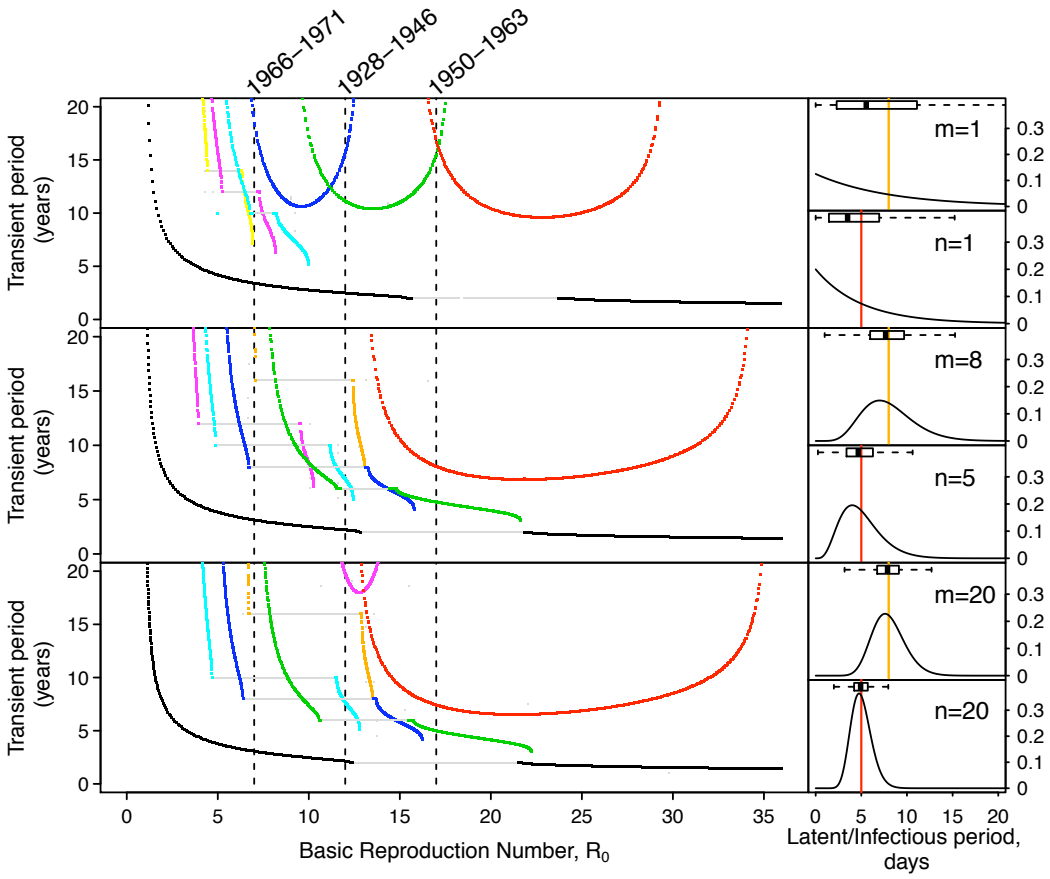


Figure 2.8: Transient periods of the measles SE^mI^nR model, for which attractors are shown in corresponding colours in **Figure 2.7**. Light grey lines indicate regions where the corresponding periodic cycles exist but are unstable. Annotation is as in **Figure 2.7**.

2.5 The role of the serial interval distribution in the dynamics of the SI^nR and SE^mI^nR models

It is surprising that narrowing the infectious period distribution in the SI^nR model (apparently making it more realistic) makes the model worse as a predictor of dynamical transitions (**Figure 2.6**). Since the effect of narrowing the shapes of the latent and infectious period distributions in the SE^mI^nR is much smaller (**Figure 2.9**), it is tempting to infer that the inclusion of a latent stage is essential for producing

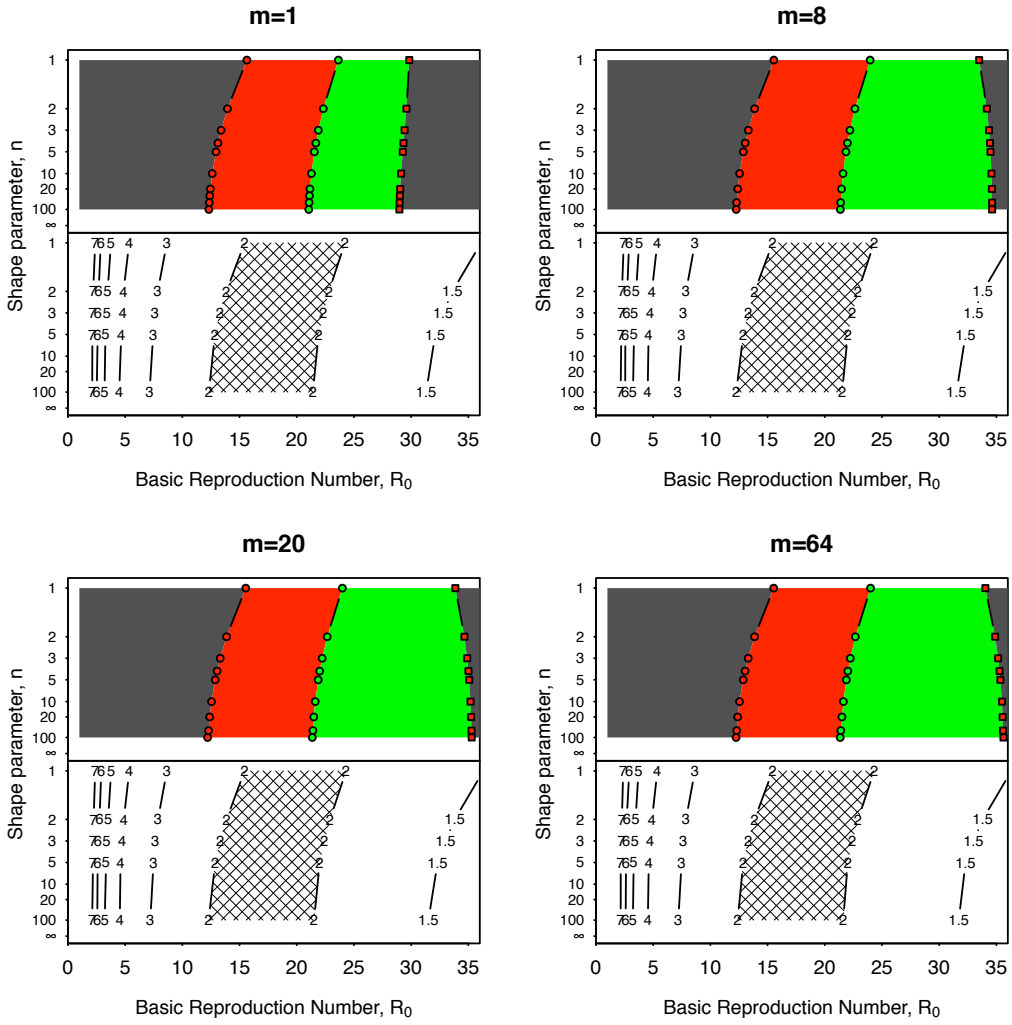


Figure 2.9: Two-parameter bifurcation diagrams and transient-period contour plots for the measles SE^{m}I^{n}R model (mean latent period $1/\sigma = 8$ days, mean infectious period $1/\gamma = 5$ days). Each panel corresponds to different values of the shape parameter (m) of the latent period distribution. Annotation is as in **Figure 2.6**.

a robust model of an infection that really does have a significant latent period. In fact, in this section, we identify the key factor that changes the structure of the SI^{n}R bifurcation diagram as n gets larger, and we argue ultimately that any SI^{n}R or SE^{m}I^{n}R model is as good as any other from the point of view of transition analysis (including the SI^{1}R or SE^{1}I^{1}R models) provided they are parameterized appropriately.

When using an SIR rather than SEIR model, we chose the mean infectious period to be 13 days, the sum of the actual mean latent (8 days) and mean infectious (5 days) periods. Our motivation was that it is well known that the dynamics of the unforced SI^1R model is almost identical to that of the unforced SE^1I^1R model if this association is made. In particular, the period of damped oscillations about the equilbrium is then identical in the SI^1R and SE^1I^1R models (Anderson and May 1991, p. 668).

It is instructive to note that the mean disease generation time or *serial interval*[1] in the SE^1I^1R model is equal to the sum of the mean latent and infectious periods. So the association we have made between the mean infectious period in the SI^1R model and the sum of the mean latent and infectious periods in the SE^1I^1R model amounts to making sure both models have the same mean serial interval. But for more general SE^mI^nR models, the mean serial interval is *not* equal to the sum of the mean latent and infectious periods. Indeed, the mean serial interval in an SE^mI^nR model is (Svensson 2007, Eq. 5.9)

$$T_{\text{serial}} = \frac{1}{\sigma} + \left(\frac{n+1}{2n}\right)\frac{1}{\gamma}. \tag{2.23}$$

From this formula we see that the mean serial interval does not depend on the shape of the latent period distribution (only its mean $1/\sigma$), but decreases as the infectious period distribution gets narrower (i.e., as n increases) if the mean infectious period is kept fixed. If the mean serial interval is the key factor affecting the dynamics of the SE^mI^nR model then we can now easily see why **Figure 2.6** shows so much more variation than **Figure 2.9**: the mean serial interval T_{serial} decreases from 13 to 6.5 days as n increases from 1 to ∞ in the SI^nR model ($1/\sigma = 0$, $1/\gamma = 13$ days), whereas T_{serial} decreases only from 13 days to 10.5 days as n increases from 1 to ∞

[1]The serial interval is also called the generation interval, the generation time, or the case-to-case interval. It is the time from initial infection of a primary case to initial infection of a secondary case Fine (2003).

in the SE^{m}I^{n}R model ($1/\sigma = 8$ days for any value of m, $1/\gamma = 5$ days).

Figure 2.10 shows another version of the two-parameter ($\mathcal{R}_0$ vs n) bifurcation diagram for the SI^{n}R model. Rather than fixing the mean infectious period as in **Figure 2.6**, for each n we set the mean serial interval to be the same as that in SE^{m}I^{n}R model with the same value of n. The result in **Figure 2.10** is now negligibly different from each of the panels of **Figure 2.9**.

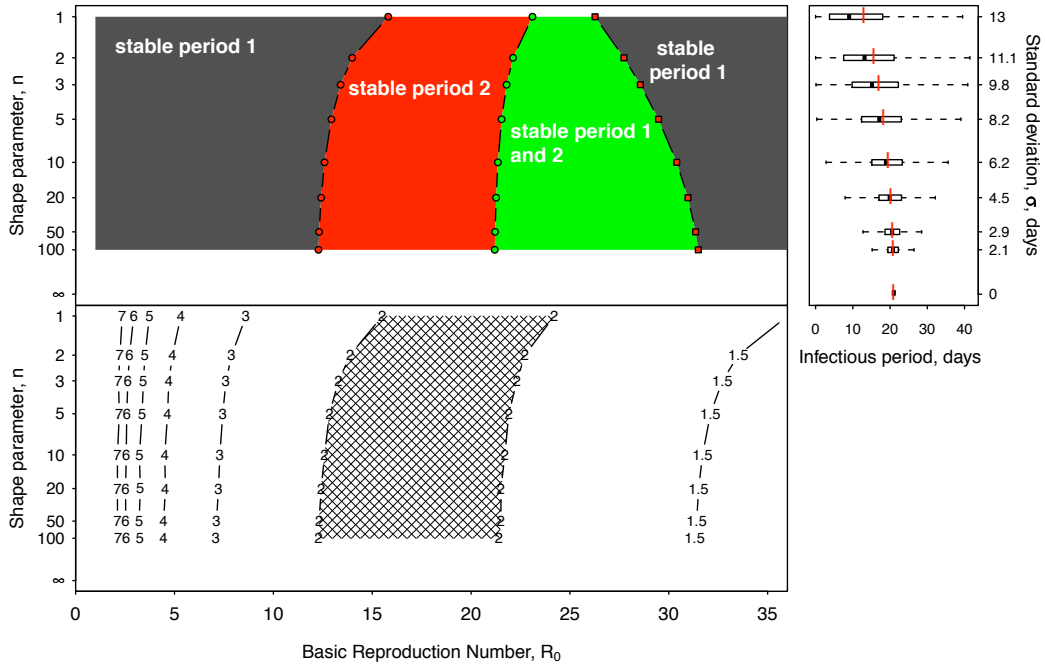


Figure 2.10: Two-parameter bifurcation diagram and transient-period contour plot for the measles SI^{n}R model with the mean serial interval chosen to be the same as in SE^{m}I^{n}R model. Annotation is as in **Figure 2.6**. Notice that it is very similar to the SE^{m}I^{n}R model diagrams in **Figure 2.9**.

Finally, in **Figure 2.11** we show yet another version of the $\mathcal{R}_0$ vs n bifurcation diagram for the SI^{n}R model, this time keeping the mean serial interval fixed at 13 days for all values of n. From this diagram (some details of which are discussed in **Appendix 2.6**), it is now clear that from the point of view of transition analysis—and to a large extent more generally for understanding the dynamics of

SE^mI^nR models—the key parameter that needs to be estimated is the mean serial interval, not the mean latent or mean infectious period themselves and certainly not the shapes of these distributions. For a given mean serial interval, it makes little difference which SI^nR or SE^mI^nR model we use, so we might as well work with the simplest, the SI^1R model.

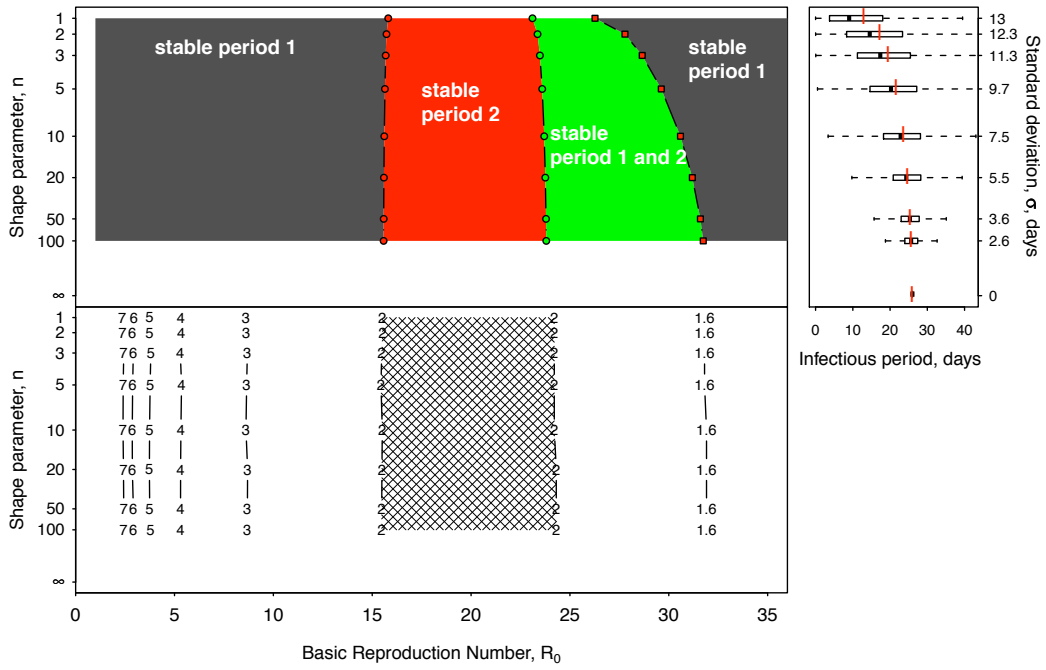


Figure 2.11: Two-parameter bifurcation diagram and transient-period contour plot for the measles SI^nR model with fixed mean serial interval, $T_{\text{serial}} = 13$ days. Annotation is as in **Figure 2.6**. Notice that the main period doubling bifurcation point from annual to biennial cycle occurs for approximately the same value of $\mathcal{R}_0$ regardless of the shape (n) of the infectious period distribution. The transient dynamics shown in the bottom panel are also the same for all values of n.

2.6 Discussion

We set out to determine whether the results of previous "transition analyses" of recurrent epidemic patterns of childhood diseases (Bauch and Earn 2003b; Earn 2009; Earn et al. 2000a) were robust to the assumed shapes of the latent and infectious period distributions (which were taken to be exponential in previous work). We undertook a systematic analysis of the sequence of SI^nR and SE^mI^nR models, parameterized for measles, and concluded that for a given mean serial interval, transition analyses based on any SI^nR or SE^mI^nR model will lead to the same predictions. Consequently, transition analyses of measles dynamics can be safely conducted using the very simplest SI^1R model. It is important to emphasize, however, that the mean serial interval must be estimated correctly for this to work; in particular, it is not true that the real mean serial interval is the sum of the mean latent and infectious periods.

The key graph that establishes that SI^nR dynamics are nearly invariant if the mean serial interval is fixed is **Figure 2.11** (where the mean serial interval is set to 13 days). In future work, the equivalent graph should be created for a sequence of mean serial intervals that covers the range of typical recurrent infectious diseases, in order to verify that transition analyses of other diseases can also be safely conducted with the simple SI^1R model.

Consistent with previous work (Lloyd 2001a,b; Nguyen and Rohani 2008), we found that if we fix the mean infectious period (rather than the mean serial interval) then narrowing the infectious period distribution (which reduces the mean serial interval) leads to more complex dynamics. Previous work has also investigated the stochastic dynamics of SI^nR and SE^mI^nR models and examined characteristics such as the critical community size for disease persistence (Conlan et al. 2010; Lloyd 2001a). In future work, inferences concerning the stochastic dynamics of these models should be re-examined in light of the now-evident importance of the mean serial interval for their deterministic dynamics.

Appendix A: Vaccination level calculations

Live measles virus vaccine was licensed in the US in 1963 (CDC). A national campaign to eliminate measles was launched by the Center for Disease Control (CDC) in October of 1966 (Witte and Axnick 1975). The campaign aimed to target infants at approximately one year of age and all remaining susceptible children when they enter kindergarten, the first grades of elementary school or other places of gathering (Conrad et al. 1971; Dull and Witte 1968). We thus assume that the individuals targeted for vaccination were children of one year (infants), four years (pre-kindergarten), five years (kindergarten) and six years (first year of elementary school) of age (USE). For example, the targeted population in 1963 (column 4 of **Table 2.1**) is the total number of children born in 1962 (one year old), 1959 (four years old), 1958 (five years old), and 1957 (six years old):

$$4,167,362 + 4,295,000 + 4,255,000 + 4,308,000 = 17,025,362 \qquad (2.24)$$

Since the data on the distribution of measles vaccine by cities was not available, we used the vaccination level calculated for the whole US (column 5 of **Table 2.1**) as an estimate of the vaccine coverage in New York City (**Figure 2.3**).

For the years after 1963 we had to account for vaccination being implemented already. Hence the size of the targeted population was still the total number of one, four, five and six year olds, but now reduced by the number of children already vaccinated. For example, the targeted population in the year 1964 (column 4 of **Table 2.1**) is the total number of children born in 1963 (one year old), 1960 (four years old), 1959 (five years old), 1958 (six years old) reduced by the number

[1]Source: Department of Health and Human Services, National Center for Health Statistics `http://www.infoplease.com/ipa/A0005067.html`

[2]Source: Witte and Axnick (1975)

Year	Births, US [1]	Vaccine doses [2]	Size of targeted population	Estimated vaccination level
1957	4,308,000			
1958	4,255,000			
1959	4,295,000			
1960	4,257,850			
1961	4,268,326			
1962	4,167,362			
1963	4,098,020	3,200,000	17,025,362	0.187955
1964	4,027,490	3,800,000	15,298,855	0.248385
1965	3,760,358	6,000,000	14,724,270	0.407490
1966	3,606,274	7,900,000	12,196,284	0.647738
1967	3,520,959	6,400,000	9,657,979	0.662665
1968	3,501,564	5,300,000	8,695,492	0.609511
1969	3,600,206	4,900,000	7,999,115	0.612568
1970	3,731,386	4,500,000	7,833,993	0.574420
1971	3,555,970	8,300,000	14,618,977	0.567755
1972	3,258,411	8,200,000	10,138,244	0.808819

Table 2.1: Vaccination level for measles in the United States.

of children already vaccinated in 1963, $\approx$19% of five and six year olds:

$$\begin{aligned} &4,098,020 + 4,257,850 + 4,295,000 + 4,255,000 - \\ &-(4,295,000 + 4,255,000) \cdot 0.187955 = 15,298,855 \end{aligned} \tag{2.25}$$

We then estimate the vaccination level as the ratio of the distributed vaccine doses to the size of targeted population.

In 1971 the combined measles, mumps and rubella (MMR) vaccine was licensed (Banatvala and Brown 2004). Since it was a new vaccine that protected against three infectious diseases, we assume that the targeted population in 1971 was again all children of one, four, five, and six years old even if they were previously vaccinated with measles vaccine. **Table 2.1** summarizes our calculations of the vaccination level for the US. Note that we did not consider vaccine efficacy (Linnemann 1973), migration of the population, or other factors that could have in-

fluenced the proportion of those vaccinated who were actually immunized. Despite the limitations of our methods, our estimates are in excellent agreement with the only two published annual vaccination rates we have found: 61.4% in 1969 and 57.2% in 1970, as reported by the United States Immunization Survey (Landrigan and Conrad 1971).

Appendix B: Main bifurcation point of the SE^mI^nR model vs SI^nR model with the same mean serial interval

From the point of view of understanding transitions in measles dynamics, the most important bifurcation is that the first period doubling (from annual to biennial cycles) that occurs as $\mathcal{R}_0$ is increased in the SI^nR and SE^mI^nR models (e.g., **Figure 2.9**). In this appendix, we examine the precise value of $\mathcal{R}_0$ at which this bifurcation occurs in a sequence of SI^nR and SE^mI^nR models, all with the same mean serial interval (**Figure 2.10**). In order to keep the mean serial interval constant as we narrowed the shape of the infectious period distribution (i.e., increasing n), we used **equation** (2.23) to determine by exactly how much we needed to lengthen the mean infectious period ($1/\gamma$); the mean latent period was set to 8 days in all the SE^mI^nR models. For example, for $n = 2$,

$$T_{\text{serial}}^{\text{SE}^m\text{I}^2\text{R}} = 8 + \frac{2+1}{2 \cdot 2} \cdot 5 = 11.75 \text{ days} \tag{2.26}$$

$$\frac{1}{\gamma_{\text{SI}^2\text{R}}} = 11.75 \cdot \frac{4}{3} \approx 15.67 \text{ days} \tag{2.27}$$

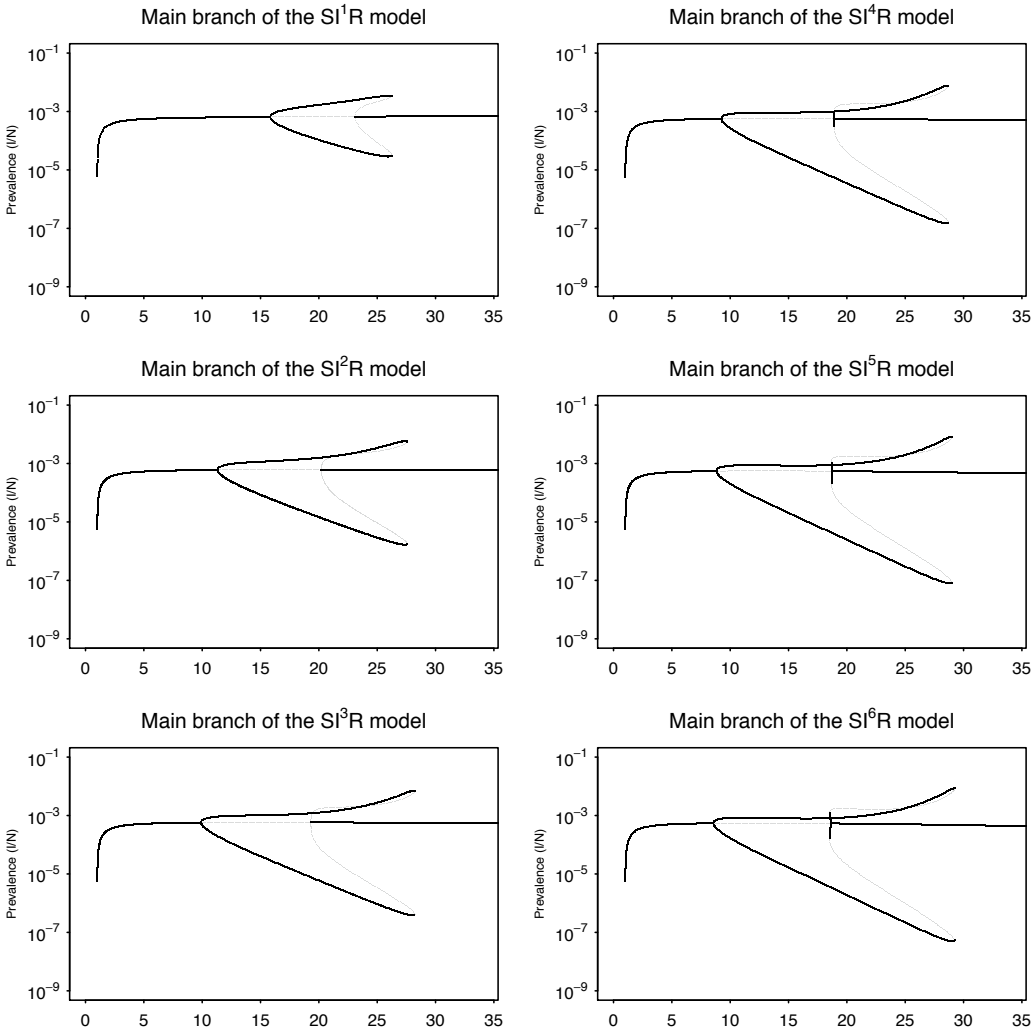
Table 2.2 lists the value of $\mathcal{R}_0$ at the key period doubling point on the main branch of the bifurcation trees of the SE^mI^nR and SI^nR models for the same given mean serial interval of 13 days. The values of $\mathcal{R}_0$ differ at most by 0.26 ($SE^{64}I^1R$ vs SI^1R), which shows that the SI^nR model is a good approximation of the SE^mI^nR model if the mean serial interval is chosen appropriately.

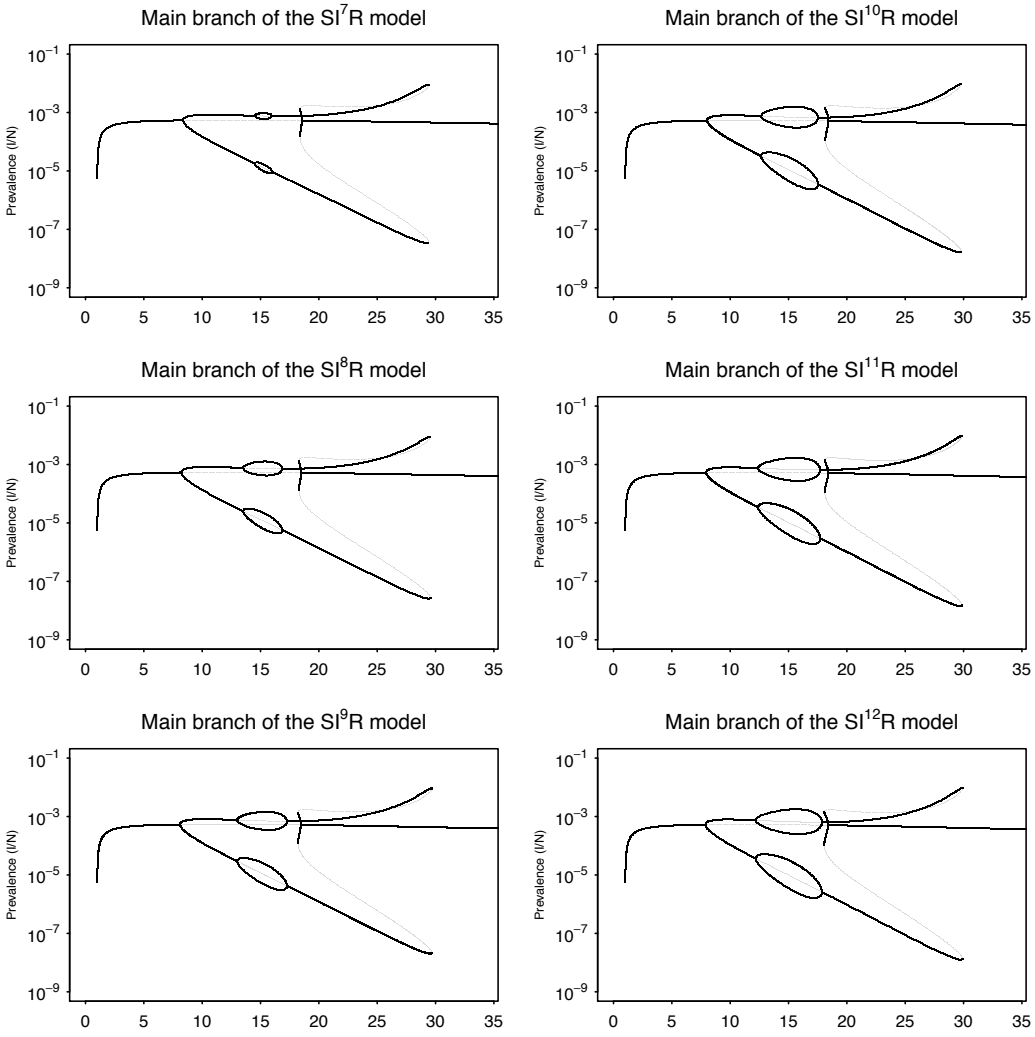
n	SE^1I^nR	SE^8I^nR	$SE^{20}I^nR$	$SE^{64}I^nR$	SI^nR	T_{serial}
1	15.6432	15.5490	15.5418	15.5386	15.7980	13.00
2	13.9431	13.8693	13.8639	13.8613	13.9870	11.75
3	13.3869	13.3186	13.3135	13.3112	13.3999	11.33
5	12.9453	12.8811	12.8764	12.8742	12.9349	11.00
10	12.6159	12.5548	12.5503	12.5482	12.5897	10.75
20	12.4520	12.3921	12.3877	12.3857	12.4180	10.63
50	12.3537	12.2947	12.2904	12.2889	12.3151	10.55
100	12.3211	12.2622	12.2601	12.2592	12.2809	10.53

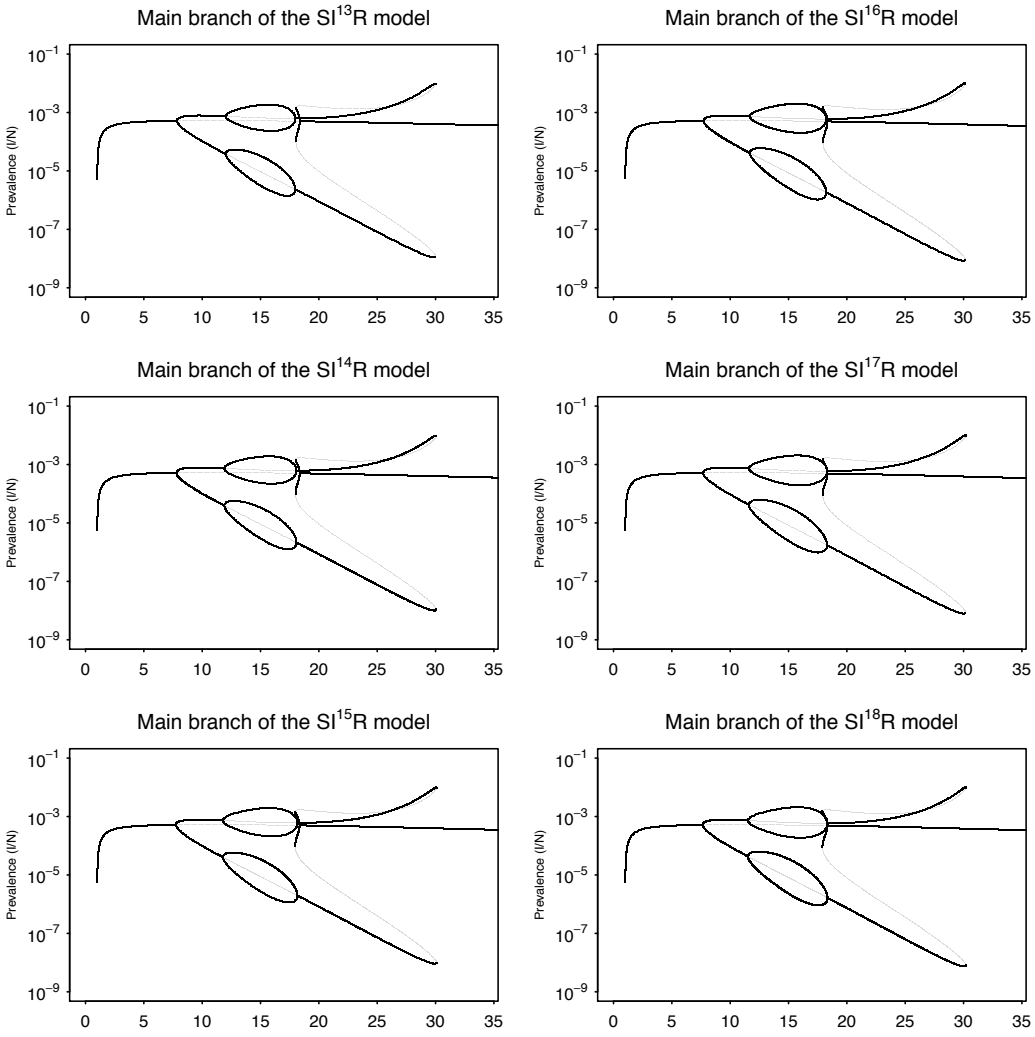
Table 2.2: The value of $\mathcal{R}_0$ at the first period doubling point on the main branch of the bifurcation diagram of SE^mI^nR and SI^nR models with the same mean serial interval, $T_{\text{serial}} = 13$ days. The mean serial interval can be expressed in terms of the mean latent and infectious periods using **equation** (2.23). The same formula (**equation** (2.23)) with $1/\sigma = 0$ gives the mean serial interval in the SI^nR model and allows us to choose a mean infectious period $(1/\gamma)$ that will yield a given serial interval.

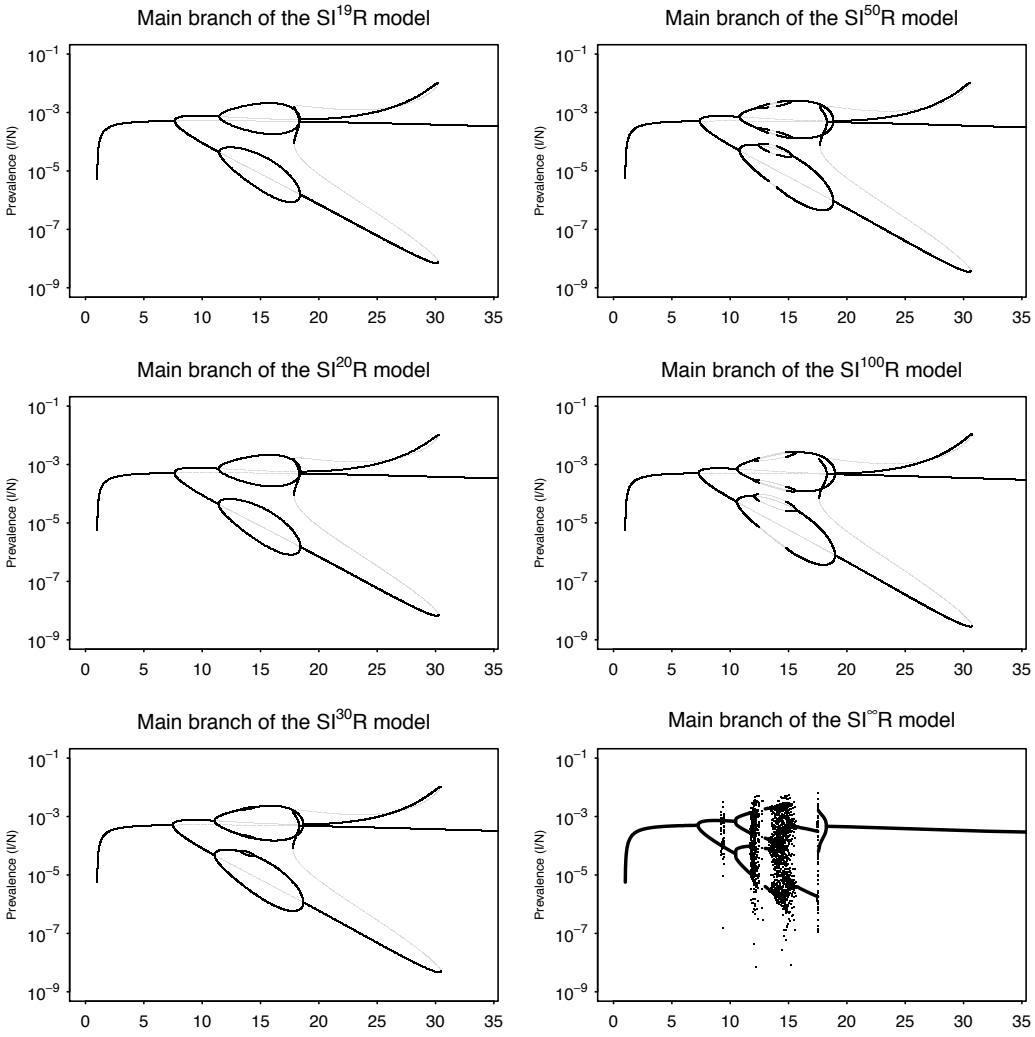
Appendix C: Main branch of the SI^nR model with the fixed mean infectious period, $\frac{1}{\gamma} = 13$ days

We plotted the main branch of the bifurcation tree of the SI^nR model (prevalence vs R_0) for $n = 1,\ 2,\ \ldots,\ 20,\ 30,\ 50,\ 100,\ \infty$


Main branch of the SI[1]R model
Main branch of the SI[4]R model
Main branch of the SI[2]R model
Main branch of the SI[5]R model
Main branch of the SI[3]R model
Main branch of the SI[6]R model
Prevalence (I/N)
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
0 5 10 15 20 25 30 35


Main branch of the SI7R model
Main branch of the SI10R model
Main branch of the SI8R model
Main branch of the SI11R model
Main branch of the SI9R model
Main branch of the SI12R model
Prevalence (I/N)
10−1
10−3
10−5
10−7
10−9
0
5
10
15
20
25
30
35


Main branch of the SI13R model
Main branch of the SI16R model
Main branch of the SI14R model
Main branch of the SI17R model
Main branch of the SI15R model
Main branch of the SI18R model
Prevalence (I/N)
10−1
10−3
10−5
10−7
10−9
0
5
10
15
20
25
30
35


Main branch of the SI19R model
Main branch of the SI50R model
Main branch of the SI20R model
Main branch of the SI100R model
Main branch of the SI30R model
Main branch of the SI∞R model
Prevalence (I/N)
10−1
10−3
10−5
10−7
10−9
0
5
10
15
20
25
30
35

Appendix D: Step by step guide to creating a bifurcation diagram of the seasonally forced SIR model using XPPAUT

This guide extends unpublished notes written by Maystruk (2006) as a supplement to her undergraduate Arts & Science thesis at McMaster University ((Maystruk 2006)).

The model

Consider a seasonally forced SIR model:

$$\frac{dS}{dt} = \mu - \beta(t)S(t)I(t) - \mu S(t), \tag{2.28a}$$
$$\frac{dI}{dt} = \beta(t)S(t)I(t) - \gamma I(t) - \mu I(t), \tag{2.28b}$$
$$\frac{dR}{dt} = \gamma I(t) - \mu R(t), \tag{2.28c}$$

where

$$\beta(t) = \beta_0(1 + \alpha\cos(2\pi t)) \tag{2.29}$$

and S, I, and R are proportions of the population in each epidemiological state.

We want to investigate how dynamics of system 2.28 changes with respect to the changes in the basic reproduction number, $\mathcal{R}_0$. Thus we write **equation** (2.28) in terms of $\mathcal{R}_0$:

$$\mathcal{R}_0 = \frac{\beta_0}{\gamma + \mu}, \tag{2.30}$$

which implies that

$$\beta_0 = \mathcal{R}_0(\gamma + \mu), \tag{2.31a}$$
$$\beta(t) = \mathcal{R}_0(\gamma + \mu)(1 + \alpha\cos(2\pi t)). \tag{2.31b}$$

A bifurcation diagram is a summary of the asymptotic dynamics (attractors) of a dynamical system as a function of a bifurcation parameter. It is helpful to start analysis of a model with a brute force bifurcation diagram, which is constructed strictly by simulations. This diagram will show stable branches of periodic orbits and will identify regions of possible chaotic behavior.

STEP 1: Brute force bifurcation diagram

Creating file *bruteforceSIR_R0.ode*

In order to compute solutions of any model in XPPAUT we need to create a *filename.ode* file that will specify the system of differential equations. This can be done in any text editor and should be saved with an extension .ode. For example, *bruteforceSIR_R0.ode*.

```
############################################
## bruteforceSIR_R0.ode
## Author: Olga Krylova
## Created: 26 Nov 2009
## Brute force bifurcation diagram
##            for seasonally forced SIR model
############################################

## EQUATIONS:
beta=Rzero*(gamma+mu)*(1 + a*cos(2*pi*t))
s'=mu-beta*s*i-mu*s
i'=beta*s*i-(gamma+mu)*i
```

XPPAUT allows us to define initial conditions and different parameters:

```
## INITIAL CONDITIONS:
init S=0.9, I=0.001

## PARAMETERS:
## 1/mu=50 years, 1/gamma=13 days=0.0356 years
par mu=0.02, gamma=28.08, Rzero=17, a=0.08
```

Using this program we can plot solutions of the SIR model for various initial conditions and parameter values in the XPPAUT window (see Ermentrout (2002) for

detailed explanations of working with XPPAUT). Now we want to plot the solution of the system for a given range of values of the bifurcation parameter, $\mathcal{R}_0$, on the horizontal axis and a particular state variable on the vertical axis, $\log_{10} I$. We choose $\log_{10} I$ instead of I since the proportion of infectious becomes very small, and it is convenient to use a logarithmic scale for a better visual representation of the results. XPPAUT allows us to produce such a plot with the help of auxiliary variables.

Auxiliary variables: any quantities in terms of state variables or parameters

```
## AUXILIARY VARIABLES:
## R0 is defined as an auxiliary variable so we can plot
## bifurcation diagram as a function of parameter Rzero
aux R0=Rzero
aux log10s=log10(s)
aux log10i=log10(i)

## XPP SET UP:

## PLOT OPTIONS:
## xp=variable on x axis, yp=variable on y axis
@ xp=R0, yp=log10i
## limits on plot
@ xlo=0, xhi=40, yhi=0, ylo=-25
```

The seasonally forced SIR model exhibits periodic solutions, so if we plot the complete solution for $I(t)$ for each particular value of $\mathcal{R}_0$ we will get straight vertical intervals corresponding to the range of the $I(t)$ at a particular $\mathcal{R}_0$. Such a diagram would be more confusing than useful. It is more convenient to use the Poincaré stroboscopic map. Choosing the strobing interval to be one year, the Poincaré map will return just one point for a period-one orbit, two points for a period-two orbit and so on. This representation of the results is much more intuitive since it allows us to show all periodic solutions of the model and their periods. In the XPPAUT Poincaré map can be set up in the following way:

```
## POINCARE MAP SET UP:
@ poimap=section,poivar=t,poipln=1
## range set up
@ range=1, rangeover=Rzero, rangestep=3000
@ rangelow=0, rangehigh=30, rangereset=no
```

We also should take into account that at the very beginning of the integrations, model solutions may exhibit transient dynamics (i.e., it takes some time to converge to a stable periodic orbit). Hence we choose the total time of integration to be 650 years and the transient to be 600 years, which will allow the solution to converge.

```
## INTEGRATION OPTIONS:
## total time of integration
@ total=650,
## transient time
@ trans=600
## step of integration
@ dt=0.001

## STORAGE and DATA SAVING OPTIONS
## storage info
@ maxstor=2000000
## background color of the plot
@ back=white
## saving data
@ output=bruteforceSIR_R0.dat
done
```

The ode file can be executed from the terminal window using -silent mode with the following command:

```
xppaut bruteforceSIR_R0.ode -silent
```

It will produce a `bruteforceSIR_R0.dat` file with all necessary data:

column 1 $time$

column 2 S

column 3 I

column 4 $\mathcal{R}_0$

column 5 $\log_{10} S$

column 6 $\log_{10} I$

Now we can plot this data using any program we like. For example,  .

Plotting brute force diagram

The following `bruteforceSIR_R0.R` file will produce **Figure 2.12**.

```
############################
## bruteforceSIR_R0.R
## Author: Olga Krylova
## Created: 28 Nov 2009
## Read data file created by XPPAUT and
## create pdf picture of brute force bifurcation diagram
## wrt R0 for the cos forced SIR model
############################
rawdata <- read.table("bruteforceSIR_R0.dat");

## define variables
time <- rawdata[,1];
S <- rawdata[,2];
I <-rawdata[,3];
R0 <- rawdata[,4];
log10S <- rawdata[,5];
log10I <- rawdata[,6];

## plot set up
pdf("bruteforceSIR_R0.pdf",width=8, height=6);
par(mar=c(5,4,0,2));
plot(R0,log10I, type="p",
 xlab=expression(paste("Basic Reproduction Number, ", R[0])),
 ylab=expression(paste(plain(log)[10],"I")),
 col="black", pch=".", cex=1, ylim=c(-8,-1));
dev.off()
```

To run the `bruteforceSIR_R0.R` file execute the following command:

```
R CMD BATCH --vanilla bruteforceSIR_R0.R
```

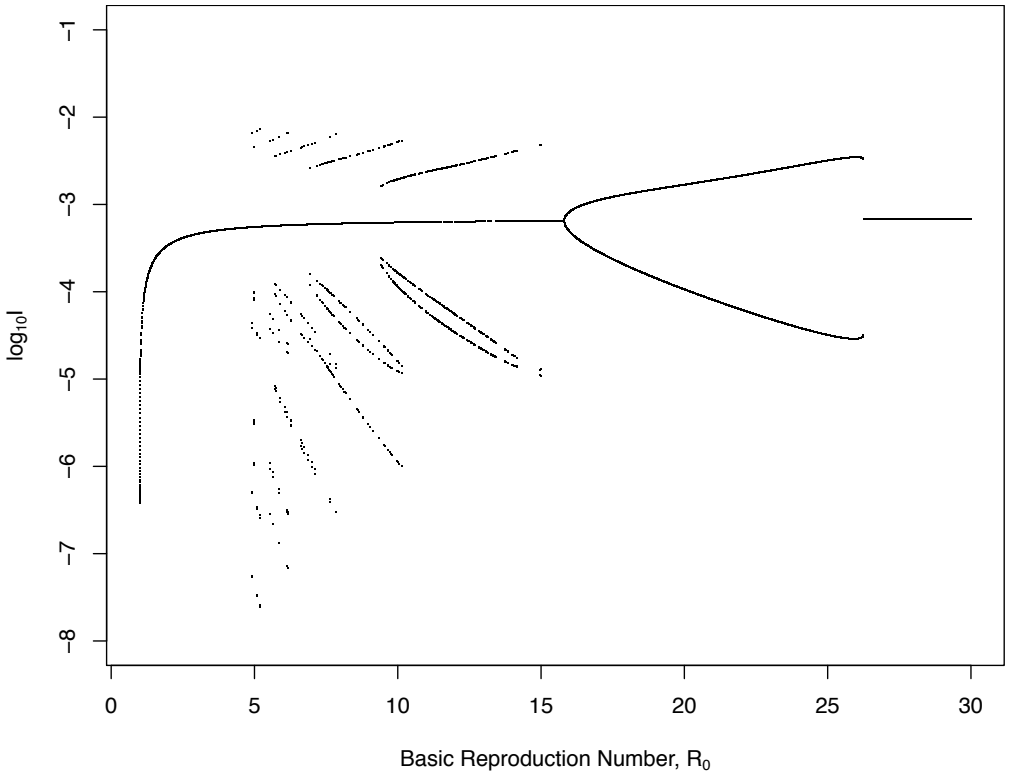


Figure 2.12: Brute force bifurcation diagram of the seasonally forced SIR model

Figure 2.12 clearly shows the existence of periodic orbits of periods one to seven.

STEP 2: AUTO bifurcation diagram

The AUTO part of XPPAUT is a very powerful tool that allows us to track bifurcation curves for steady-state and periodic systems (Doedel 2007). In particular, we can use AUTO to follow stable and unstable branches of periodic orbits. The disadvantage of this method is that AUTO's default settings only allows us to plot either the maximum, minimum, mean or period of the cycle. It is easier to visualize

the bifurcation structure when plotting the diagram as a Poincaré map, similar to the brute force diagram. This way, annual cycles will be displayed as a single solid line, biennial cycles as a double line, cycles of period 3 as triple line, *etc* (Kuznetsov 1995; Wiggins 2003). Hence we will use AUTO to study the discrete time system, or Poincaré map, associated with our original continuous time system. We can analyze the discrete system, keeping in mind that a fixed point of a Poincaré map corresponds to a periodic orbit of the original system. The problem of stability of the periodic orbits is reduced to the problem of stability of fixed points of the map, which is determined by eigenvalues of the map linearized about the fixed points.

Constructing Poincaré map using C-libraries

Using a C-program we can create a function that solves the system of ODEs numerically via the Euler or Runge-Kutta method for a given period of time. In our case we will use a period of one year. Our function uses initial conditions specified in our .ode file or in the XPP window and returns the integrated values after one year as output. This way we are creating a discrete map that can be easily analyzed in XPPAUT. The file `SIRmap.c` will be used in conjunction with the file `SIR.ode`, the file that specifies all initial values and parameters.

```
/***
Poincare map of the seasonally forced SIR model
Created: 02 Dec 2009
Author: Olga Krylova
 ***/

#include <math.h>

/***
 IMPORTANT NOTE:
 We found that the function must be written
 as one entity, without using called functions.
 ***/

#define Time_step 0.0005  /* in units of years */
```

```
#define DAYS_PER_YEAR 365
#define TWO_PI 6.2831853071795864769252867665590
#define Real double

/*** We called our function SIRmap ***/
SIRmap(double *in, double *out, int nin,
           int nout, double *var, double *con)
/*
 in = initial and parameter values we get
        from the ode file (s, i, R0, alpha, gamma, mu)
 out = what we are returning (sp,ip):
        calculated values of S and I after one year
 nin = dimension of in[]
 nout = dimension of out[]
 var and con are arrays that Bard said to include...
 */
{
  /* define starting values in log base 10 */
  double s=in[0], i=in[1];
  /* converting back to the original values,
        not in log */
        s=pow (10,s);
        i=pow (10,i);
  /* define parameter values */
  double R0=in[2], alpha=in[3], gamma=in[4], mu=in[5];
  double ds, di;
  Real seasonal_beta, nonlin_term;
  double time; /* in units of years */
  long istep, nsteps;

  /* number of steps in a year */
  nsteps = (int)( 1/Time_step + 0.5 );

  /* integrating for one year */
  for (istep=0; istep < nsteps; istep++) {
    time=(double)(istep)*Time_step;

    seasonal_beta=R0*(gamma+mu)*(1+alpha*cos(TWO_PI*time));
    nonlin_term=seasonal_beta*s*i;

    /* Compute Xdot */
    ds = mu - nonlin_term - mu*s;
```

```
    di = nonlin_term - (mu + gamma)*i;

    /* Euler method */
    s = s + ds*Time_step;
    i = i + di*Time_step;

  }
  s = log10(s);
  i = log10(i);
  out[0] = s; /* Output in log_10 */
  out[1] = i;
}
```

This C code should be compiled as a shared object library file using the command

```
gcc -dynamiclib -o SIRmap.so SIRmap.c
```

Creating file *AutoSIR.ode*

This file links C-library SIRmap function to XPPAUT.

```
## ODE file for Poincare map of
## the seasonally forced SIR model
## (used in conjunction with SIRmap.c library)
## Created: 05 Dec 2009
## Author: Olga Krylova

## DEFINE LEFT HAND-SIDE
s' = sp
i' = ip
sp = 0
ip = 0

## LINK TO THE C-LIBRARY
## i.e. pass the values {s0, i0, R0, alpha, gamma, mu}
## and ask it to return {sp, ip}
## Note that the order of export must agree with
## the order of in[] and out[]
## in arrays in the C function
export {s, i, R0, alpha, gamma, mu} {sp, ip}

## define a library to be used
```

```
## and a corresponding function
@ dll_lib=SIRmap.so dll_fun=SIRmap

## SET INITIAL CONDITIONS for each periodic orbit
## Note that, for convenience, s and i below
## are really log(s) and log(i)
## nOut sets the number of iterations of the map
## it should be changed for each periodic orbit
set p1 {init s=-1.4139113,
        init i=-3.1684341, R0=27.059999, nout=1}
set p2 {init s=-1.250071,
        init i=-2.776037, R0=19.950001, nout=1}
set p3 {init s=-1.1593947,
        init i=-4.3912416, R0=12.58, nout=1}
set p4 {init s=-0.91435295,
        init i=-4.8151603, R0=9.5299997, nout=1}
set p5 {init s=-0.74093175,
        init i=-5.1411638, R0=5.75, nout=1}
set p6 {init s=-0.71477449,
        init i=-2.272049, R0=5.5300002, nout=1}
set p7 {init s=-0.69836187,
        init i=-2.1395164, R0=5.1900001, nout=1}

## PARAMETER VALUES
## Note the order here determines the
## "main" parameter for AUTO
## alpha = amplitude of seasonal forcing
## R0 = basic reproduction number
## gamma = recovery rate; 1/gamma = mean infectious period
## mu = mean death/birth rate; 1/mu = average life time
par R0=30, alpha=0.08, gamma=28.076923, mu=0.02
aux Rzero=R0

## XPP SETUP
## this is a discrete map not an ODE
@ meth=discrete
## total=20 mean 20 iterations of the map in total
@ total=20, yp=i
## line type = dots
@ lt=0
## plotting options
@ xlo=-1, xhi=21, ylo=-9, yhi=-1
```

```
## AUTO SETUP
## set range for R0, our control parameter:
@ parmin=1.1, parmax=32
## set range of vertical axis variable
##  and set which variable it is:
@ autoymin=-9,autoymax=-1,autovar=i
## set horizontal axis plot range:
@ autoxmin=0, autoxmax=32
## set step size for continuation
##  of the control parameter:
##  (here, ds=standard step size,
##  others are max and min step size)
##  (the sign of ds controls
##  the direction of continuation)
@ dsmax=0.1, ds=0.003, dsmin=0.0000003
## set a few other techinical
##  aspects of the continuation:
@ Nmax=20000, Npr=2000, epsl=1e-6, epsu=1e-6, sepss=1e-4
## above:
## Nmax = maximum number of steps
##  to take along a branch before stopping
## Npr = number of steps
##  before labelling a point (which can help with
##  continuing from points
##  without having to start everything all over)
## eps... = various tolerances

done
```

We need to point out a few important commands that were used in the .ode file.

– Set up initial conditions for the SIRmap.c file and read its output:

```
export {in} {out}
```

where `{in}` indicates initial values and parameters to be passed to the external function and `{out}` defines returning values.

– Specify which C-library (`file.so`) and function to load by including the following line in the ODE file:

```
@ dll_lib=file.so dll_fun=function_name
```

– Define set of initial conditions for each periodic orbit

Initial conditions for the state variables should be specified in the ODE file as well. Note that to compute a periodic orbit in AUTO, we should start running AUTO from a point on this periodic orbit. But how would we know where to start? That is where the brute force diagram becomes very handy. The data from the brute force diagram contains points that lie on stable periodic orbits. Therefore we can extract points on the period 1-, 2-, 3-, 4-, 5-, 6-, 7-year cycles from the brute force data file and then specify them in .ode file.

The following R file reads the data from the previously generated brute force data file and creates the set of initial conditions that will be used to calculate each periodic orbit in AUTO.

```
##########################
## bruteforceSIR_R0_IC.R
## Author: Olga Krylova
## Created: 28 Nov 2009
## Read data generated by
## brute force bifurcation diagram;
## determine the period of each periodic orbit
## and create a set of
## initial conditions to be used in AUTO
##########################
## read data
raw.data <- read.table("bruteforceSIR_R0.dat",
 col.names=c("time", "S", "I", "R0",
 "log10S", "log10I"));

## we use data only for last 15 years
## of each integration from 634 to 649
index <- which(raw.data$time==634|raw.data$time==635|
 raw.data$time==636|
 raw.data$time==637|
 raw.data$time==638|
 raw.data$time==639|
 raw.data$time==640|
 raw.data$time==641|
 raw.data$time==642|
 raw.data$time==643|
```

```
 raw.data$time==644|
 raw.data$time==645|
 raw.data$time==646|
 raw.data$time==647|
 raw.data$time==648|
 raw.data$time==649);
new.data <- as.data.frame(raw.data[index,]);

## for small values of R0 the proportion
## of infectious people is very small.
## It is easier to find stable branches for R0>4
new.data <- new.data[which(new.data$R0>=4),];

## calculating period
period <- array(NA);
i <- 1;
while (i<=length(new.data$time))
{
        ## determine how many unique values are
        ## present for each integration step
        period[i:(i+15)] <-
         length(which(duplicated(
         round(new.data[i:(i+15),6],5))==FALSE));
        i <- i+16;
}

## adding column of period values to the data frame
new.data <- as.data.frame(cbind(new.data, period))

## randomly choosing initial conditions
## corresponding ot each periodic orbit
index1 <- sample(which(period==1),1)
set1 <- new.data[index1,]

index2 <- sample(which(period==2),1)
set2 <- new.data[index2,]

index3 <- sample(which(period==3),1)
set3 <- new.data[index3,]

index4 <- sample(which(period==4),1)
set4 <- new.data[index4,]
```

```
index5 <- sample(which(period==5),1)
set5 <- new.data[index5,]

index6 <- sample(which(period==6),1)
set6 <- new.data[index6,]

index7 <- sample(which(period==7),1)
set7 <- new.data[index7,]

## saving data in a setIC.csv file
setIC <- as.data.frame(rbind(set1, set2, set3, set4,
         set5, set6, set7))
write.table(setIC, "setIC.csv",
 row.names=FALSE, col.names=TRUE, sep=",")
```

The set of points on each periodic orbit is saved in `setIC.csv`:

$time$	S	I	$\mathcal{R}_0$	$\log_{10} S$	$\log_{10} I$	period
639	0.038555704	0.00067852496	27.059999	-1.4139113	-3.1684341	1
647	0.056224938	0.0016748	19.950001	-1.250071	-2.776037	2
…	…	…	…	…	…	…

Notice that we can define many sets of initial data, parameter values and options in the ODE file using the command:

```
set name {parameter values, initial data, options}
```

These sets can be invoked while running XPPAUT with the

```
File→Get par set
```

command. For example, the set for period one orbit is

```
set p1 {init s=-1.41, init i=-3.17, R0=27.06}
```

– It is convenient to set up all necessary AUTO plotting options in the ODE file so that we can plot the bifurcation diagram right away. For example, we can set the range of bifurcation parameters by using commands

```
@ parmin=#, parmax=#
```

or set the axis with

```
@ autoymin=#, autoymax=#, autoyvar=variable_name
@ autoxmin=#, autoxmax=#
```

Following periodic orbits

We have everything ready to compute the bifurcation diagram in AUTO.

- Run XPPAUT

  `xppaut AutoSIR.ode`

- Find period one orbit
  - Choose the first set of initial values called `p1`

    `(F)ile→(G)et par set`
  - Integrate the system to reach an equilibrium

    `(I)nitialconds→(G)o`

    We should get 20 points in a straight horizontal line, indicating that the system is at equilibrium. To assure that the system indeed converged to an equilibrium, check data values by clicking the `Data` button on the top panel of the XPP window. Data values of S and I are still different after the third decimal place, which shows that we have not converged to the fixed point yet. Hence we need to integrate the system a few more times by pressing `(I)nitialconds→(L)ast` until we get repeating values for S and I. After repeating `I L` two more times we find an equilibrium.
- Load AUTO

  `(F)ile→(A)uto`
- Continue the branch in both directions
  - Hit `Run` to follow the period one branch in a positive direction ($\mathcal{R}_0$ increases).

– Change the sign of parameter `Ds` in the `Numerics` menu of AUTO to continue computations in the opposite direction ($\mathcal{R}_0$ decreases). Select the initial point with `Grab` and press `Run` again.

`Numerics`→`Ds:-0.003`

`Grab`→`<Return>`→`Run`

We have computed the period one orbit.

- Save data

  `File`→`All info`→name file `branch1allinfo.dat`

- Clear memory for the new branch

  `File`→`Reset diagram`→hit `OK` and erase data

- Compute period two branch

  – Integrate the system to reach an equilibrium

  `F`→`G`→choose set `p2`→`I`→`G`→`I`→`L`→`I`→`L`...

  Integration results in two horizontal dotted lines indicating that we hit a period two cycle. Check the `Data` window to verify that the values of S and I oscillate between two constant values.

  – Define the period of the orbit. To continue branches of periodic points of the Poincaré map we have to indicate how many times the map should be iterated before returning the value of the map. This can be done using parameter `n(O)utput` from the `n(U)merics` menu. For example, setting `n(O)utput` to two allows us to see all of the branches up to period two. Note that the previous value of `n(O)utput` was one, which allowed us to plot only period one orbits.

  `n(U)merics`→`n(O)utput`→type 2→`<Return>`

  – Run AUTO

  `F`→`A`→`Run`

  We want to have separate data files for each periodic orbit, so that we can plot them later in different colours. Therefore, when computing

period 2 orbit, we have to hit the `ABORT` button very quickly to avoid plotting period one points.

- Save data

  `File`→`All info`→name file `branch2allinfo.dat`

- Compute period three branch

  - Integrate the system to reach an equilibrium

    `F`→`G`→choose set `p3`→`I`→`G`→`I`→`L`→`I`→`L`...

    Three horizontal dotted lines are plotted in the XPP window indicating that we hit a period three cycle. In the `Data` window, values of S and I oscillate between three constant values.

  - Define the period of the orbit

    `n(U)merics`→`n(O)utput`→type 3→`<Return>`

  - Run AUTO

    `F`→`A`→`Run`

    You will see that AUTO computed only one of the three branches of the period three orbit. Since the branches of the period three orbit are disconnected (we see that from the brute force diagram), AUTO can detect only a branch based on the initial conditions. Hence to compute the other two branches we must start running AUTO from points on these branches. This can be achieved by using `Data` recorded in the main XPP window from our integrations. Back in the main XPP window, click the `Data` button. That window contains three distinct values of S and I. To set one of them as the initial point, make that line the first line in the XPP `Data` window and press `Get`. While this operation will assign appropriate values to S and I, the value of the bifurcation parameter $\mathcal{R}_0$ will not be changed accordingly and that must be done manually. Click on the `Param` button in the XPP top menu to change the value of $\mathcal{R}_0$ to the one saved under the column `Rzero` in the `Data`

window. Do not forget to click `ok` in the `Parameter` window, which is necessary to set parameter values to the original value.

- Save data

  `File`→`All info`→name file `branch3allinfo.dat`

- Repeat procedure to compute periodic cycles of period four, five, six and seven.

  1. in the XPP window: `F`→`G`→choose set `p#`→`I`→`G`→`I`→`L`→`I`→`L`...
  2. in the XPP window: `U`→`O`→specify `n_out #`→`<Return>`→`<ESC>`
  3. in the AUTO window: `R`→hit `ABORT`
  4. in the XPP window:
     (a) open the `Data` window→ scroll down so that the second line is the first line of the window→press `Get`
     (b) copy the value of `Rzero` from the `Data` window
     (c) open the `Param` window→set the value of $\mathcal{R}_0$ as in (b)→ hit `Ok`
  5. in the AUTO window: `Run`→`No` to destroy diagram → `ABORT` to stop integration
  6. repeat steps 4-5 for each branch of the orbit
  7. in AUTO: `F`→`A`→save file `branch#allinfo.dat`→`Ok`
  8. in AUTO: `F`→`R`→`YES`

Plotting AUTO bifurcation diagram

The following R file can be used to create a nice plot as in **Figure 2.13** using all the data we computed in XPPAUT.

```
## plot AUTO output

pdf("AutoSIRBifDiag.pdf", width=8, height=6)
#quartz()
plot(seq(1:10), type="n", main="",
 xlim=c(0,33), ylim=c(-9,0),
 xlab=expression(paste
```

```
 ("Basic Reproduction Number, ", R[0])),
 ylab="Incidence I/N",
 yaxt="n");
y.ticks <- c(seq(0, -8));
y.label <- c(0, expression(10^-1), expression(10^-2),
 expression(10^-3), expression(10^-4), expression(10^-5),
 expression(10^-6), expression(10^-7), expression(10^-8));
axis(2, at=y.ticks, las=2, label=y.label);
auto<-lapply(1:7, function(.indx)
 read.table(paste("branch",.indx,"allinfo.dat", sep="")));
for (i in 7:1)
{
        # type of point
        n <- auto[[i]][,1];
        # branch number
        b <- auto[[i]][,2];
        # first active parameter
        R0 <- auto[[i]][,3];
        # second active parameter
        alpha <- auto[[i]][,4];
        period <- auto[[i]][,5];
        shi <- auto[[i]][,6];
        ihi <- auto[[i]][,7];
        slo <- auto[[i]][,8];
        ilo <- auto[[i]][,9];
        # real part of first eigenvalue
        ev1.re <- auto[[i]][,10];
        # imaginary part of first eigenvalue
        ev1.im <- auto[[i]][,11];
        ev2.re <- auto[[i]][,12];
        ev2.im <- auto[[i]][,13];
        stable.pts <- which(n==1);
        unstable.pts <- which(n==2);
        points(R0[unstable.pts],
         hi[unstable.pts],col=i,pch=".");
        points(R0[stable.pts],
         ihi[stable.pts],col=i,pch=19, cex=0.4);
}
segments(17,0, 17,-8, col="red", lty=2);
dev.off()
```

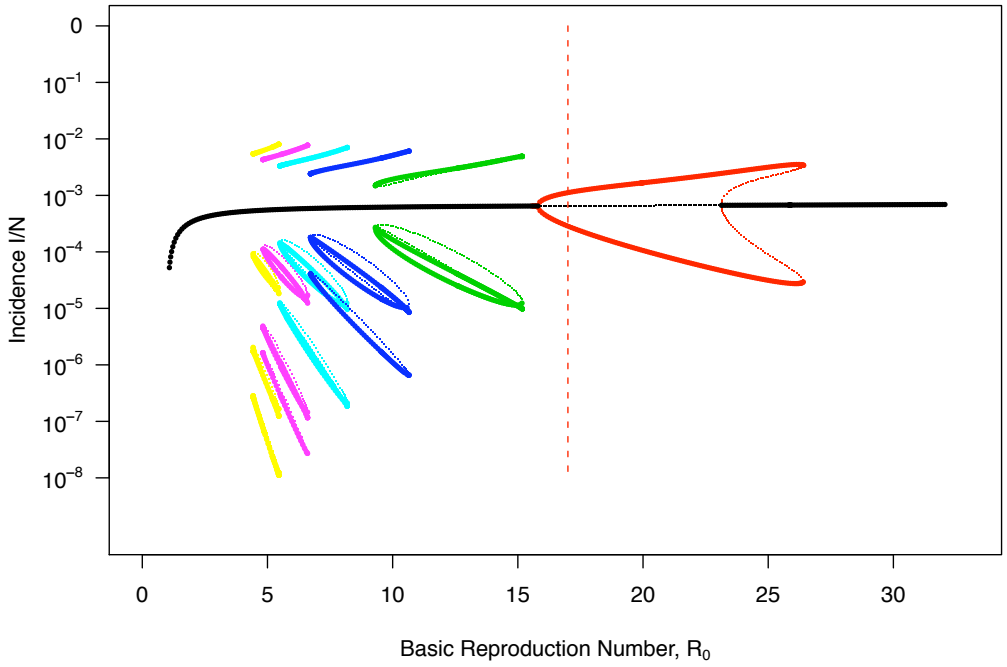


Figure 2.13: AUTO bifurcation diagram of the seasonally forced SIR model

Chapter 3

The temporal pattern of smallpox mortality in London, England, 1664-1930

Abstract

Smallpox is one of the most terrifying and devastating infectious diseases that has affected humans. It was declared eradicated by the World Health Organization in 1979 (WHO, 2011 a). However, smallpox still presents a potential threat as a biological weapon because of its high fatality rate, the lack of population immunity and the lack of an effective treatment (Kemp 2005). Modelling public health responses to the potential release of smallpox virus has been in the spotlight recently. Nevertheless, relatively little attention has been given to the analysis of historical epidemics of smallpox. The London Bills of Mortality, which provide weekly records spanning several centuries, present a unique opportunity to study historical epidemics of many infectious diseases. The smallpox records are likely among the most accurate in the Bills, because smallpox was easily recognized. Moreover smallpox was constantly present in London with frequent and often devastating epidemics since its first recorded outbreak in 1610 (Creighton 1965). Our

study presents detailed analysis of London's weekly smallpox mortality records from 1664 until the end of 1930. We describe these data using standard statistical tools and correlate shifts in the temporal structure of smallpox outbreaks with a number of historical events and government interventions related to implementation of various control strategies.

Keywords: smallpox; London Bills of Mortality; smallpox in London; wavelet diagram; seasonality.

3.1 Introduction

Smallpox is an extraordinary research subject. During its existence smallpox had been directly responsible for many history changing events and was probably accountable for more deaths than any other illness, including bubonic plague and cholera (Hopkins 1983; Macaulay 1866; McNeill 1998; Razzell 1977). According to the London Bills of Mortality, over the last three centuries smallpox killed over 320,000 people in London alone.

This paper presents a detailed analysis of smallpox death records from the Weekly London Bills of Mortality over almost three hundred years. This data set starts from the early time of smallpox appearance in London, when only the naturally acquired virus existed and no control measures were available, and runs through the periods when various preventative measures and government regulations regarding smallpox control were implemented. Our time series ends in 1930, after which smallpox deaths were extremely rare in London. Using methods of time series analysis, we identify major characteristics of smallpox dynamics in historical London. We also propose mechanisms that could have been responsible for the persistence of smallpox epidemics, their severity, and changes in their patterns. Previous studies of smallpox mortality data from the London Bills of Mortality were carried out by Duncan et al. (1996). However their work was based on aggregated

yearly data, which limit the epidemiological details that can be estimated. The weekly data that we examine provide a unique opportunity to investigate in much greater details mortality from smallpox and the slopes of epidemic curves including seasonal changes.

In the following subsections, we briefly discuss the origin and early history of smallpox, invention of preventative measures against smallpox and the story of eventual eradication success, which made smallpox the first infectious disease eradicated by human efforts. We finish the introduction with a description of smallpox virology and its natural history. **Section 3.2** outlines the history of smallpox in England, highlights important historical events related to smallpox control, and estimates variolation and vaccination uptake levels. We provide a detailed description of the available data set and also examine changes in London's population over the period for which we have smallpox data in **Section 3.3**. Description of the methods we have used for the time series analysis of the data, our principal findings and conclusions are presented in **Section 3.4**. We summarize our analysis and future research directions in the last section of the paper (**Section 3.5**).

3.1.1 Origin and early history

The origin of the smallpox virus is not known. It probably coexisted with human beings for thousands of years. It has been hypothesized that smallpox first appeared in Asia or Africa sometime after 10,000 BC and then spread to India and China (Hopkins 1983). Egyptian mummies from the Eighteenth and Twentieth Dynasties (1570 to 1085 BC) are probably the earliest credible evidence of smallpox existence in the ancient world (Fenner et al. 1988; Hopkins 1983). The mummified head of Egyptian Pharaoh Ramses V (died 1157 BC) clearly shows scars that could have been caused by smallpox (**Figure 3.1**). It is generally assumed that the pharaoh died of smallpox, but scientists have not been able to prove or disprove this assertion (Hopkins 1983). DNA analysis of a tissue sample that would be relatively easy to

Figure 3.1: The mummified head of Egyptian pharaoh Ramses V depicts pock marks that could have been caused by smallpox virus. The pharaoh presumably died of smallpox in 1157 BC. Photo source: WHO, 2011 (b)

perform in modern laboratories might reveal the presence of smallpox. However, the mummy is located in the Cairo museum of Egypt and is considered a national treasure by Egyptian authorities, hence any examinations of the mummy are not permitted (Geddes 2006; Hopkins 1983).

After its establishment in Egypt, India and China, smallpox spread to Athens and Persia. By the eighth century AD the virus had reached Japan in the East and Europe in the West (Hopkins 1983). In the fifteenth century, smallpox was widespread throughout Europe (Fenner et al. 1988). The Spanish invasion of Mexico in the sixteenth century brought smallpox to the previously unexposed populations of the Aztec and Inca civilizations. Devastating epidemics killed nearly half of the native population of Mexico in less than six months after its first appearance in April of 1520 (Hopkins 1983). Two centuries later, smallpox had become a major endemic disease everywhere in the world.

3.1.2 Invention of preventative measures (variolation and vaccination)

Smallpox was directly responsible for the death of hundreds of thousands of people each year during the Early Modern period (1500-1800) in Europe (Razzell 1977). In many cases surviving victims were left blinded or disfigured for life (WHO, 2011 a). Smallpox also could have been a major cause of infertility in male survivors (Hopkins 1983; Razzell 1977). The only defense against this "speckled monster" was a procedure called *inoculation*. It was introduced in Europe only in 1721 despite being successfully used in China and India for centuries before. Inoculation is now known by the more specialized term *variolation* and can be described as an injection of the smallpox virus (taken from a pustule or dried scabs of a person suffering from smallpox) into a healthy individual (Razzell 1977). The discovery of *vaccination* by Edward Jenner in 1796 provided a safer and much cheaper alternative to variolation. Vaccination was originally an injection with cowpox virus, which also provided immunity to smallpox. This discovery was a major milestone in modern medicine as it laid the ground for all future vaccination strategies. Even the word *"vaccine"* takes its origin from the latin *"vacca"* meaning cow and was first used by Jenner to describe his new method of "vaccine inoculation" (Jenner 1801).

3.1.3 Eradication success

The existence of a vaccine was the key factor that made eradication of smallpox an achievable goal. Other important factors were the absence of an animal reservoir, easy recognition of the disease from symptoms, and non-existence of asymptomatic cases (Center for Global Development). The World Health Organization launched its eradication campaign in 1967 (Fenner et al. 1988; Geddes 2006). Ten years later the world's last endemic smallpox case was registered in Somalia (**Figure 3.2**).

Figure 3.2: 23-year-old Ali Maow Maalin, Merka, Somalia (1977). The last recorded smallpox case in the world. Photo source: CDC, 2009 (b)

In 1979, a Global Commission declared the eradication of smallpox (WHO, 2011 b). The only remaining viral samples were stored in laboratories in Russia and the United States. The world was finally free of "the greatest killer" (Hopkins 1983). This was the first disease to be eradicated entirely by human efforts.

3.1.4 Modern challenges

Unfortunately, the tragic events of September 11 2001 and the anthrax scare of 2001 led to widespread concern about bioterrorism and reminded us that smallpox still presents a potential threat as a bioweapon. In fact, its high case fatality rate, person-to-person transmission and limited immunity in the population make smallpox more dangerous than anthrax. It is classified as a Category A bioweapon agent by the Centers for Disease Control and Prevention (CDC), which confirms its potential hazard (CDC, 2009 a). Vaccination is the only available measure to protect against this virus, but it is not effective after 4 days of infection (WHO, 2011

a). Currently the majority of the world's population is completely unprotected. As routine vaccination was stopped in the 1970s, everybody born since then is fully susceptible. In addition, the level of immunity in people vaccinated before 1970 is probably very low. While naturally acquired smallpox and variolation in most cases produce life-long immunity, vaccinia virus provides complete immunity only for a short period of time, 3 to 5 years, after which the immunity wanes rapidly (Fenner et al. 1988; Ryan and Ray 2004). Long term vaccine-induced immunity is uncertain. However, even after 20 years the case fatality rate in vaccinated persons is much smaller than in the unvaccinated (Fenner et al. 1988).

Currently available vaccine is not considered safe because of very rare but extremely dangerous complications that include severe eczema, progressive vaccinia that could lead to death, and postvaccinal encephalitis (inflammation of the brain). During routine vaccination in the United States six to eight children died each year because of various vaccine complications (Hopkins 1983, p.294). Moreover smallpox vaccination is not safe for people with weakened immune systems (e.g. pregnant women, HIV-positive individuals and cancer patients). Consequently, mass vaccination is not performed because the risks of complications are perceived to be much greater than the risk of a bioweapon attack.

3.1.5 Smallpox virology

In this section we review major characteristics of the smallpox virus and present the natural history of smallpox infection. Proper understanding of a disease virology is very important for any infectious disease study and particularly for modelling work. For example, Kaplan et al. (2002) and Halloran et al. (2002) aimed to determine optimal control measures in the case of smallpox bioterrorism. However they used inaccurate biological assumptions and hence their final conclusions were considered to be dubious (Cooper 2006; Eichner and Dietz 2003). Most recent papers make more realistic assumptions (Eichner 2003; Longini et al. 2007).

Variola virus

Smallpox is caused by the *Variola* virus (**Figure 3.3**), which takes its name from the latin word *varius* meaning spotted or *varus* meaning pimple. The commonly accepted term "small-pox" was first used in England at the end of the 15th century to distinguish it from syphilis, which was known as "great-pox" (Hopkins 1983, pp.22-29).

Variola virus is a member of the genus of orthopoxviruses, which also includes cowpox, monkeypox, buffalopox, vaccinia and many other members (Hopkins 1983, p.6; Fenner et al. 1988, pp.69-120). There are two distinct variants of *Variola* virus that can cause smallpox: *Variola major* and *Variola minor* (Fenner et al. 1988, pp.1-68; Hopkins 1983, pp.3-9; Ryan and Ray 2004, pp.525-527). These two smallpox variants differ substantially in severity of symptoms and case fatality rate.

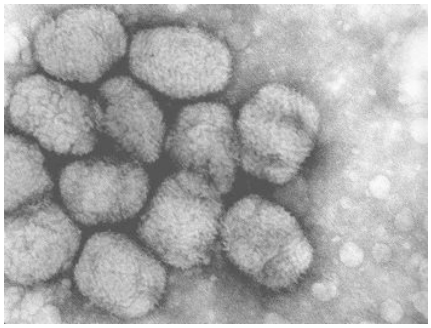


Figure 3.3: *Vriola major* virus. Photo source: CDC, 2009 (b)

Variola major, which has a case fatality rate of 5-25% and occasionally higher (Fenner et al. 1988, p. 4), was the only known smallpox type until the beginning of the 20th century. Depending on strain virulence and host response, the following clinical types of *variola major* were defined (Fenner et al. 1988, pp.1-121):

- **ordinary-type** was the most common clinical type of *variola major* (about

80% of cases), with a case fatality rate of about 20%;

- **modified-type** was milder than ordinary-type with an accelerated course of infection and occurred mostly in previously vaccinated persons;
- **variola sine eruption** was a rare type that occurred in vaccinated persons through contact with smallpox patients and was characterized by high fever without rash eruption;
- **flat-type** was characterized by lesions that remained flat. It was a rare type and usually fatal;
- **haemorrhagic type** is distinguished from other types by very severe symptoms, short incubation period and high fatality rate of about 96%. It was rare, occurred mostly in adults, and at the early stage of infection was hard to distinguish from the ordinary-type or modified-type smallpox.

Variola minor was first recognized by Korte in 1904 (Hopkins 1983). Later investigations showed that it appeared around 1894 in North America and then spread to South Africa, Europe and Australia (Fenner et al. 1988, pp.1-121). *Variola minor* causes a milder, less virulent form of smallpox and has a mortality rate of about 1% or less. *Variola minor* was the only endemic type of smallpox present in England after 1920 (Hopkins 1983, pp.8, 97; Fenner et al. 1988, pp.243). In 1935 England became smallpox-free for the first time since its introduction, and further cases of smallpox arose only from importation.

Natural history of smallpox infection

Smallpox is very acute, highly contagious and frequently fatal. It is characterized by high fever and a distinctive skin rash that often leaves pock scars after scabs fall off (WHO, 2011 a). Infection usually occurs via the respiratory tract through air droplets if exposure occurs from face-to-face contact with an infectious person. Direct contact with virulent rash, bodily fluids, or bedding, blankets, or clothing used by an infectious individual can also, in rare cases, result in smallpox infection

(Hopkins 1983, p.3; Fenner et al. 1988, pp.121-168).

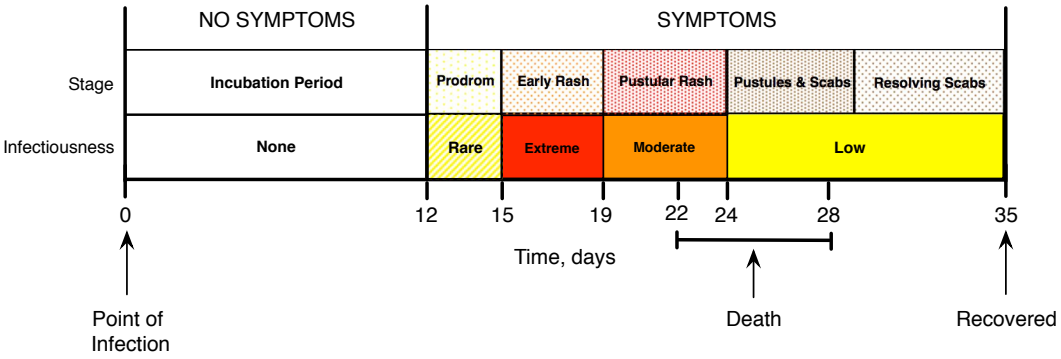


Figure 3.4: The natural history of smallpox infection. The *prodrom stage* begins with fever but the patient is very rarely contagious. *Early rash* is the most contagious stage, when the rash develops and transforms into bumps. During the *pustular rash* stage bumps become pustules, which then turn into scabs during the *pustules and scabs* stage and fall off during the *resolving scabs* stage. The infected person is contagious until the last scab falls off.

The course of a single smallpox infection (its natural history) depends on variant type, clinical type, and vaccination status of the host individual. Since the ordinary type of *Variola major* was the most common type of smallpox, we will describe its natural history here (**Figure 3.4**).

There is an *incubation period* during which the infected person has no symptoms and is not contagious. The duration of this stage can vary from 7 to 19 days but in most cases is about 12 days. The *prodrom* stage begins with the onset of fever and sometimes includes vomiting and diarrhea, and is rarely contagious. The rash appears 2-4 days after the onset of fever. It starts as small red spots on the tongue and in the mouth that grow into sores that break open within 24 hours of their appearance. At this point, a large amount of virus is contained in the mouth and throat of the infected host, making him/her extremely contagious. Then the rash spreads rapidly all over the body and in a few days transforms into bumps filled with thick fluid. This *early rash* stage continues for about 4 days.

It is followed by a *pustular rash* stage (average duration is 5 days) during which bumps become pustules. Over the next 5 days (*pustules and scabs* stage), pustules turn into scabs. Scabs fall off during the *resolving scabs* stage (6 days duration), often leaving pock marks on the skin. The overall duration of the illness is about 23 days (CDC, 2004). In fatal cases, the majority of deaths occur on the 10th-16th day from the beginning of symptoms (Fenner et al. 1988, p.22). The usual cause of death is severe toxemia, which produces a lethal concentration of circulating immune complexes and viral antigens (Fenner et al. 1988, p.130; Gantz et al. 2005, p.345).

3.2 Smallpox in England

The London Bills of Mortality used in this study provide smallpox mortality records spanning an extremely long period of time (276 years). Such an extensive data set requires not only comprehensive statistical analysis but also a thorough examination of the smallpox history in England. Understanding the historical background, which is presented in the following subsections, will allow us to create a timeline of significant historical events related to implementation of control measures, population movement, wars, *etc.*, that may have influenced smallpox dynamics (**Figure 3.5**). Creating this timeline is also necessary for estimating the level of variolation and vaccination uptake during the 17th-20th centuries in England.

3.2.1 Control-free era (1610-1721)

Smallpox first appeared in Europe around the sixth century AD. By the end of the sixteenth century it was well established in almost all of Europe (Hopkins 1983). Shortly thereafter it had become one of the major causes of death, exceeding in numbers plague, leprosy and syphilis (Hopkins 1983; Landers 1993; Mercer 1990). The first recorded outbreak in England is dated to August 12, 1610 (Creighton 1965,

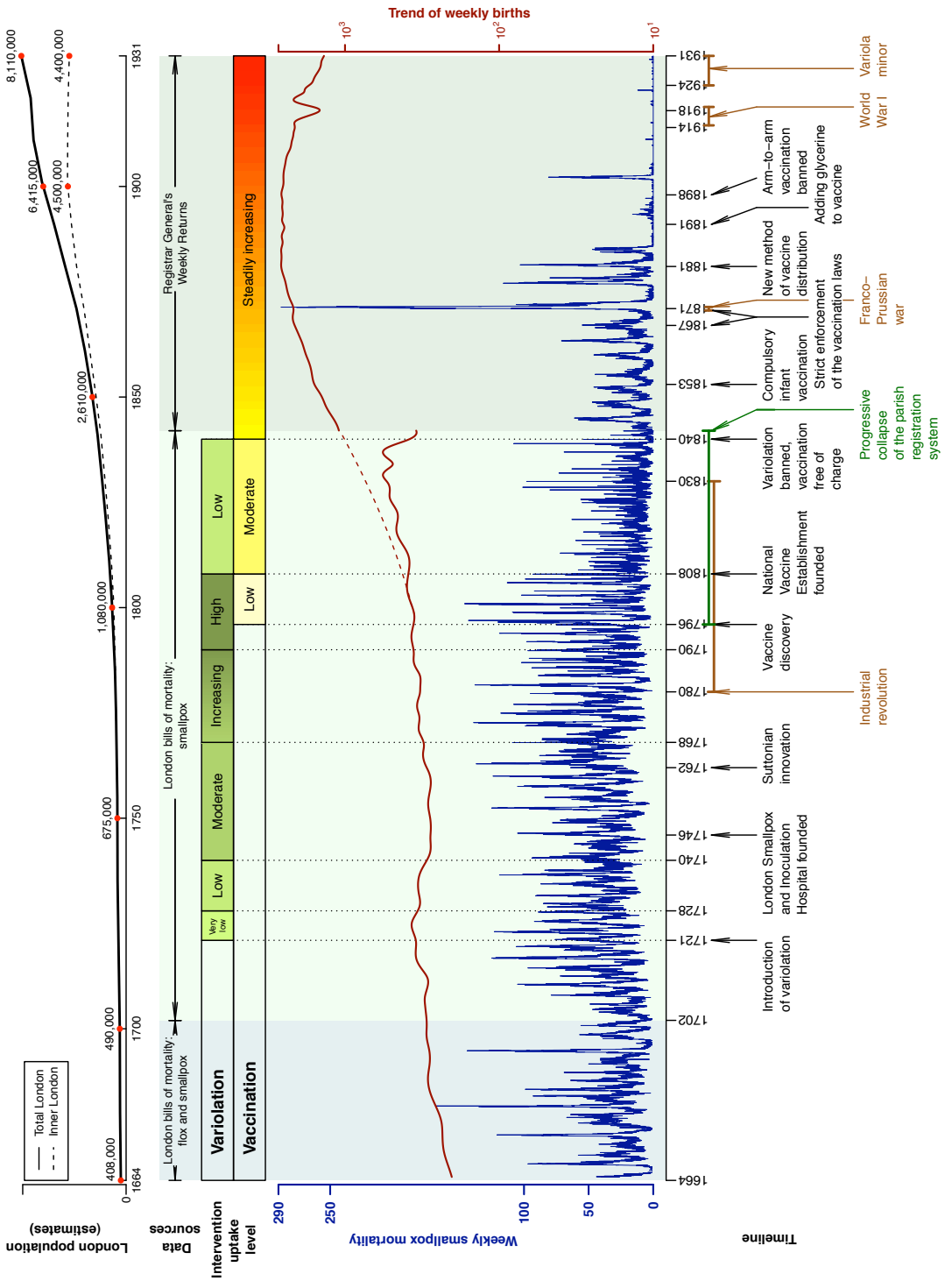


Figure 3.5: London's population and weekly smallpox mortality from 1664 until 1930 against the timeline of historical events related to the smallpox history in England. **(continued on the next page)**

Figure 3.5: The **top panel** shows estimates of the London population for inner London (dashed black line) and total London (solid black line). The **main panel** starts with indication of data sources. Smallpox mortality data were collected from two sources: i) London bills of mortality, where smallpox deaths were recorded from 1664 until 1701 under the name "flox and smallpox" (light blue background colour) and under the name "smallpox" (light yellow background colour) from 1701 until 1841; ii) Registrar General's Weekly Returns (light grey background colour) used from 1842 until 1931. Intervention uptake levels are shown as colour bars: yellow green-dark olive for variolation with yellow green indicating the lowest level and dark olive the highest; and yellow-red for vaccination with light yellow indicating the lowest level and red the highest level. Digitized records of weekly smallpox mortality from the London Bills of Mortality are presented in dark blue. The trend of weekly births are plotted in dark red. The transition from one registration system to another during 1796-1842 resulted in reduced accuracy of the data records. Dashed dark red line shows fitted birth trend during this period. The **bottom panel** shows the timeline of historical events related to the smallpox history in England: text in black colour indicates events that influenced uptake of control measures; brown text shows events that influenced people behaviour; dark green shows the period when data accuracy is reduced.

p.435), after which it became endemic.

The registration system introduced in some European countries around the end of the 16th century presented indisputable evidence of the increasing impact of smallpox on population dynamics. Statistical records of smallpox mortality were first used in 1580 by the authorities in Geneva. In England, the weekly reports of church burials by cause of death were recorded by parish clerks from the end of the 16th century. In London, weekly bills of mortality were first recorded in 1592.

3.2.2 Variolation era (1721-1808)

There were no available preventative measures in use against smallpox until the introduction of variolation, which came to England from Turkey in the beginning of the 18th century. Lady Mary Wortley Montagu, one of the most influential women of her century, a writer and a poet (Grundy 2001), had her daughter professionally

variolated in London in April 1721. She is given all the credit for introducing variolation to Great Britain (Fenner et al. 1988; Geddes 2006; Hopkins 1983; Razzell 1977). Since its first appearance and until the 1740s the practice of variolation was not very popular among the British population. Only 857 people were variolated in the whole of Great Britain from 1721 to 1727 and only 37 in 1728 (Creighton 1965). After 1728 the numbers are unknown but assumed to be low. English medical practitioners implemented variolation very crudely with deep incisions (deep injection technique) that caused severe symptoms, morbidity and high mortality (up to 2%). It took more than two decades for the general public to overcome its fear of this new technique and understand the benefits it offered. At the beginning of the 1740s variolation was affordable only to rich people. Nevertheless the number of variolations started to grow. Charitable variolation began with the establishment of the London Smallpox and Inoculation Hospital in 1746. This hospital was one of the first places to provide free variolation and care for poor patients with natural smallpox and for variolated persons during the time of their infectiousness.

Before the 1760s, variolation was usually preceded by four to six weeks of preparation, which included purging, bleeding and a restricted diet with limited amount of food, very little meat and low alcohol consumption. Variolation was followed by an isolation period of two or more weeks. Throughout all this time variolated persons were placed in inoculation houses specifically built for this purpose (Fenner et al. 1988; Razzell 1977, p.255). The lengthy preparation period was proven to be unnecessary after Robert Sutton's innovation of variolation through light incisions in 1762. Sutton's new method dramatically decreased the severity of symptoms and death and reduced the cost due to the shorter preparation period. Consequently, variolation was offered more often free of charge to the poor population. Robert Sutton also believed that isolation after variolation was unnecessary (Razzell 1977, p.28). He was not aware of the fact that variolated individuals could also spread the infection to the unprotected population (Fenner et al. 1988, p.255).

As a result, elimination of the isolation period, while reducing the cost of the procedure, also had the potential to seed new smallpox outbreaks.

Sutton's new variolation technique spread quickly around England and became very popular in rural areas. When the threat of a new epidemic became highly probable, "general variolation" of entire villages and communities was performed. In large towns and cities the situation was quite different. In London, the use of variolation was very irregular and attempts to perform "general variolation" were sporadic and very rare. The only possibility for poor Londoners to be variolated was through Smallpox Charities. They performed public variolation in batches, separately for males and females, approximately 8 to 12 times a year (Creighton 1965, p.506), but this was insufficient to control smallpox. The charities did not admit children under 7 years of age, despite the fact that the vast majority of smallpox cases at that time were in children under 3 years of age. The full extent of variolation in London after the Suttonian innovation is unclear, though figures from London Hospital show that the number of variolated individuals increased dramatically from 29 in 1750 to 653 in 1767 and almost doubled the following year to 1084 (Creighton 1965, p.506). This was in large part due to the increasing practice of the Suttonian method. Studying various historical reports, Razzel concluded that variolation gained considerable popularity in London at the very end of the 18th century and the beginning of the 19th century (Razzell 1977, p.72).

The Industrial Revolution, which started in the 1780s, brought thousands of immigrants to London and the population grew rapidly. In London and other densely populated cities of England, outbreaks of smallpox were recurrent and the disease never died out completely. Rural areas were relatively smallpox free between epidemics (Duncan et al. 1994a; Fenner et al. 1988). Hence young adults migrating from rural cities were at great risk of acquiring the infection and contributed continuously to the recruitment of susceptibles into London.

3.2.3 Vaccination era (1796-1931)

Despite the growing popularity of variolation and its widespread practice in Europe, by the end of the 18th century smallpox still remained endemic in London with increasingly severe outbreaks (**Figure 3.5**). The discovery of vaccination by Edward Jenner in 1796 was the crucial first step towards smallpox eradication. Jenner was practicing medicine in Berkley, Glucostershire. Being a rural doctor he knew that milkmaids, who had blisters on their hands from cowpox, did not catch smallpox. This fact gave him the idea that infection with cowpox may possibly give immunity to smallpox. On May 14, 1796 he injected a small boy from his village, James Phipps, with cowpox virus taken from a milkmaid, who had recently acquired the infection from a cow. The boy developed a pustule around the area of injection but had no other reaction. A couple of weeks later Jenner inoculated the boy again, but this time with smallpox virus, which had no effect on the boy. Jenner submitted his findings to the Royal Society in 1797. However, his work was accepted with great skepticism and even ridicule. Despite this reaction from his colleagues, Edward Jenner continued his experiments again in 1798 with a larger group of individuals and found the same results. This time his work was accepted more widely (Fenner et al. 1988; Hopkins 1983).

Offering almost no risk to the vaccinated person, no preparation period and reduced cost, the new prevention method against smallpox became widely adopted by the public much faster than variolation (Fenner et al. 1988). Initially vaccination had many problems associated with ineffective methods of vaccine distribution and storage, shortage of cowpox virus, vaccine efficacy and religious and philosophical objections. Despite these difficulties, more and more people were vaccinated, which resulted in a dramatic decline of smallpox mortality by the end of the 19th century (**Figure 3.5**).

Unfortunately, vaccinations were poorly recorded until the end of the 19th century. Available data are misleading, incomplete, uncertain and inconsistent. For

example, the figures from the London Smallpox and Inoculation Hospital show the percentage of vaccinated patients admitted to the hospital increasing steadily from 32% in 1825 to 73% in 1856 (Hardy 1983, p.114). Mercer (1990) referenced the Royal Commission on Vaccination, which found that only 25% of newborns were vaccinated by 1820 and about 70% in some parishes by 1840. Mooney (1997) states that from 1854-1856 the percentage of vaccinated infants might have ranged from 28% to 81%. Another source indicates that infant vaccination rates for London during the period 1845-1890 were much lower than the national average and never increased above 500 per 1,000 live births (i.e. 50%) (Mooney 1997). Data for vaccinations of older age groups during this period do not exist. It is hypothesized that many adults escaped vaccination, and thus, the numbers appear to be very inconsistent (Hardy 1983). Still, some rough inferences can be made about the vaccine uptake level from these sources.

State involvement in the control of smallpox in England began with the foundation of the National Vaccine Establishment in 1808. It provided free vaccination in its London stations and distributed vaccine to other parts of England (Hennock 1998). Around this time, the London Smallpox and Inoculation Hospital ceased variolation of its out-patients and began vaccination in greater numbers. The next decade was characterized by very mild smallpox outbreaks, which were believed to be due to the growing popularity of vaccination. Striking epidemics occurred in London in 1817-1819 and 1837-1838, the latter growing into a Europe-wide pandemic. Both large smallpox epidemics coincided with massive typhus outbreaks (Creighton 1965; Hopkins 1983, p.87), which were presumably a coincidence. The authorities in England realized that some radical measures had to be taken, which led to the first Vaccination Act of 1840, providing vaccination free of charge and banning variolation. It was followed by the Vaccination Act of 1853, which made the vaccination of every child during the first four months of life compulsory. The Vaccination Act of 1867 introduced penalties for not complying with

the compulsory vaccination.

The Franco-Prussian war started in late July 1870 and is believed to have initiated the worst pandemic of smallpox in all of Europe in the 19th century, which resulted in at least half a million deaths. England lost more than 40,000 people. Thanks to compulsory vaccination, the fatality rates in England were three times lower than in Prussia, Austria and Belgium (Hopkins 1983, pp.87-91). The immediate response of the English government to this devastating pandemic was the Vaccination Act of 1871. It enforced very strict control (through the courts) of the implementation of previous Acts.

In the second half of the 19th century, many problems that arose in the early stages of vaccination had been resolved. Arm-to-arm vaccination [1], which was used as the main method of vaccine distribution at the beginning of the century, was replaced by the new technique of passing cowpox from cow to cow. Arm-to-arm vaccination was in many cases very dangerous since it could transmit other diseases such as syphilis. The new method of vaccine distribution was first introduced by Negri of Naples in 1843. However it arrived in England only in 1881. Another important discovery was made in 1891 by Monckton Copeman. He showed that adding glycerine to smallpox vaccine reduces bacterial contamination, making it more efficient and reliable. The dangers associated with arm-to-arm vaccination finally resulted in its complete outlaw in 1898 (Hopkins 1983).

The main objection to Jennerian vaccination was uncertainty in the duration of immunity it provided. Evidence of smallpox cases in previously vaccinated individuals was noticed as early as 1804. It brought up the question of vaccine efficacy and the probable need of revaccination. Unfortunately Jenner's false belief of life-long vaccine induced immunity was very well accepted in England and was not questioned at all. Hence revaccination there started much later than in other

[1] Arm-to-arm vaccination was the method of vaccine distribution, which involved vaccine transfer from the infectious pustule of vaccinated individual to a non-vaccinated individual, and so on (Fenner et al. 1988)

European countries. Germany was the first country to recognize the necessity for revaccination and began so in 1829. Revaccination came to England only at the end of the 19th century.

Until the beginning of the 20th century, smallpox was endemic in London. The last large outbreak of *Variola major* occurred in London in 1901-1902, after which very small outbreaks occured very rarely and were mostly imported from other countries. The new type of smallpox virus, *Variola minor*, was first noticed in England in 1901 and was well established by 1919. It appeared in London for the first time in 1928 and was then endemic until 1934 (Fenner et al. 1988).

In 1967 the World Health Organization launched its global smallpox eradication campaign, and by 1979, smallpox was certified as the first infectious disease to be eradicated by human efforts.

3.3 Data description

In this study, we use digitized records of the weekly London bills of mortality. We examine weekly births, smallpox and all-cause mortality in London from the week of October 18, 1664 until January 1, 1931. The data were collected from the Guildhall Library, British Library, Wellcome Library, and London Metropolitan Archive. The bills of mortality were used for the years 1664 to 1841. The Registrar General's Weekly Returns for London had been printed since 1837, but the records found in the Guildhall Library begin in 1842. They represented the data source from 1842 until 1931 (**Figure 3.5**). The following subsections present a more detailed description of the available data.

3.3.1 London mortality data

A death registration system was introduced in England at the end of the 16th century. It was created to keep track of deaths from plague and alert the city authorities

at the onset of plague epidemics. Bills of mortality were collected by the Parish Clerk Company and included information about weekly baptisms and church burials, as well as causes of death, gathered from the individual Anglican parish registers (Finlay 1981; Mercer 1990). In London, weekly mortality records were published regularly from 1604 (Creighton 1965). In addition, annual summary bills were published from 1629 (Creighton 1965). Annual all-cause mortality (1629-1837) as well as smallpox mortality (1629-1893) from these bills can be found in Creighton (1965) and is reproduced in **Figure 3.6**.

From the first available records until 1701, the original disease category used to record smallpox deaths in the London bills of mortality was *"flox and smallpox"* (**Figure 3.5**). *"Flox"* is an old term for the haemorrhagic type of smallpox (p.436 Creighton 1965; Schmidt 2011). After 1701 the name *"smallpox"* was used consistently to distinguish it from all other causes of death (**Figure 3.5**). Some confusion was created by the presence of *"bloody flux"* and *"flux"* as the cause of death in the records, which could have also been related to smallpox. Razzell (1977) suggested that *"bloody flux"* was a name used for haemorrhagic smallpox and was considered a separate disease (Razzell 1977, p. 104). However Creighton (1965) described it as an old name for dysentery and not as something related to smallpox (Creighton 1965, p.774). A Glossary of Archaic Medical Terms also defines *"bloody flux"* as dysentery and *"flux"* as diarrhea (Schmidt 2011). In any case, the mortality from *"bloody flux"* and *"flux"* is negligible compared to smallpox (total of 4679 deaths from "bloody flux" and "flux" compare to 322,219 total smallpox deaths). Therefore, even if they were related to smallpox deaths, they would not significantly influence our findings. Hence we used the sum of *"flox and smallpox"* and *"smallpox"* records and did not include *"bloody flux"* and *"flux"* in our data.

The bills of mortality were the only official registration system used in England until the introduction of national records, the Registrar General's Weekly Returns, in 1837. The necessity for improvement of the methods of data collection

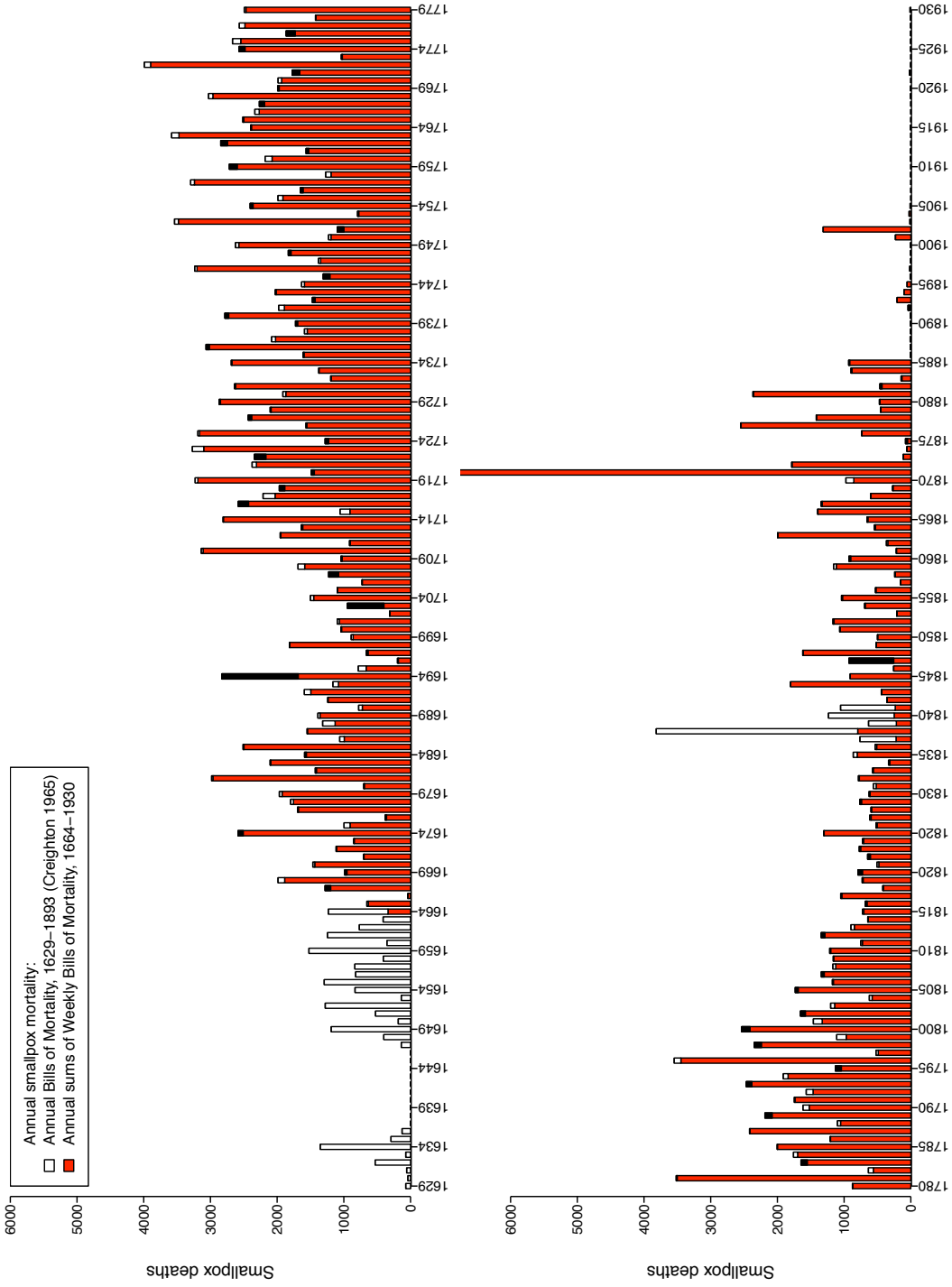


Figure 3.6: Annual smallpox mortality data in London, England: 1629-1779 (**top panel**) and 1780- 1930 (**bottom panel**). **(continued on the next page)**

Figure 3.6: The data were gathered from two sources: annual smallpox mortality records from the Annual Bills of Mortality (Creighton 1965) and annual sums of weekly digitized mortality records from the Weekly Bills of Mortality. The white bars show the data from the Annual Bills, which were missing in the Weekly Bills. The differences between the two data sets are shown as white stack bars if values in the Annual Bills are larger, and black stack bars if values in the Weekly Bills are larger. Note that smallpox mortality data for the period 1637-1646 were not available (Creighton 1965). Also there is a significant difference between the two sources of information during the period of transition from one registration system to another (1837-1841).

became evident at the end of the 18th century. The accuracy of the old system of parish records had been compromised by the rapid growth of London's population from the beginning of the industrial revolution and a lack of expansion in Anglican parishes. Parish clerks simply could not keep up with the increasing flow of information. As a result, the General Register Office was created with the main purpose of keeping birth and mortality records more complete and covering all sectors of the population (Mercer 1990).

Even a cursory examination of the weekly pattern of deaths from smallpox in the bills of mortality reveal that for the period 1796-1842, the last week of the reported year, which at that time was the first week of December, had an unusually high number of reported deaths (**Figure 3.5**). The reason for this appears to be a backlog. As mentioned earlier, this was also the beginning of the tremendous population growth in London. Parishes became overwhelmed and, as a result, the bills of mortality became increasingly inaccurate. To address this problem, we replaced the records showing strikingly high mortality with the average of the previous and following weeks. Then the difference between original and replaced values was uniformly distributed throughout the year.

The continuous historical records available to us go back as far as 350 years to 1662. Not surprisingly reports from some weeks have been lost. Fortunately, all

gaps are very small (typically 1-5 weeks) with the largest gap of 9 weeks. These gaps were replaced with linearly interpolated values to obtain time series without missing values.

3.3.2 London population data

London population data (**Table 3.1**) was gathered from many different sources. Finlay and Shearer (1986) estimated the total population of London for every half of the century from 1500 until 1700, based on information about the population North and South of the River Thames. They also took into account inaccuracies in data collection. We noticed that (Finlay and Shearer 1986, Tables 2 and 3) contain population data for the North and South of the River not just for every 50 years, but also for some other years. Hence we repeated calculations of Finlay and Shearer (1986) to estimate the population of London for these extra years. Note that Finlay and Shearer (1986) (Table 5) estimate for 1650 appears to be inaccurate due to an arithmetic error (it should be ≈400,000 instead of 375,000 (**Table 3.1**)). Population data for 1700-1800 were reproduced from Landers (1993), who estimated the population for every decade from 1730 until 1800. Census data, which is available for each decade since 1801, were the source of remaining values (Census). The resulting data (**Table 3.1** of **Appendix 3.5**) are plotted in the top panel of **Figure 3.5**.

3.4 Time series analysis

In this section we identify changes in the temporal pattern of smallpox dynamics in London using methods of time series analysis. We also correlate these changes in smallpox dynamics with historical events outlined earlier.

3.4.1 Time plot

The time plot of the raw data is presented in the **Figure 3.5**. It shows significant differences in the structure, amplitude and temporal frequency of epidemics over different periods of time. However, the raw data can be misleading since it does not account for changes in population growth and inconsistency of data sources. For example, the epidemic of 1871 appears to be the largest one over the whole period. However it was not the most significant relative to the population size. Surprisingly, the epidemic of 1838, a major outbreak in the 19th century that is frequently mentioned in the literature (Creighton 1965; Hopkins 1983), is not easily identifiable in the raw data. The reason may be the use of different data sources: our annual totals of smallpox mortality differ substantially from data in Creighton (1965) for 1838-1841.

3.4.2 Normalized data

To control for changes in population size, city boundaries and data collection methods, we normalized the smallpox data by the trend of the weekly all-cause mortality in London (**Figure 3.7**).

To calculate the trend, we used the recently developed method of Empirical Mode Decomposition (EMD), which is considered to be the most appropriate approach for finding trends in nonlinear and highly non-stationary time series (Huang et al. 1998; Vatche and Sharple 2008; Wu et al. 2007). It was developed to overcome the flaws of other commonly used methods such as moving averages or other linear filters, or linear regression, which may perform poorly on non-stationary data. EMD decomposes a signal (in our case the time series) into several components with a well defined instantaneous frequency, termed "intrinsic mode functions" (IMF). IMFs are basically zero-mean oscillation modes present in the data: the first IMF captures the high frequency (shorter period) oscillations while all subse-

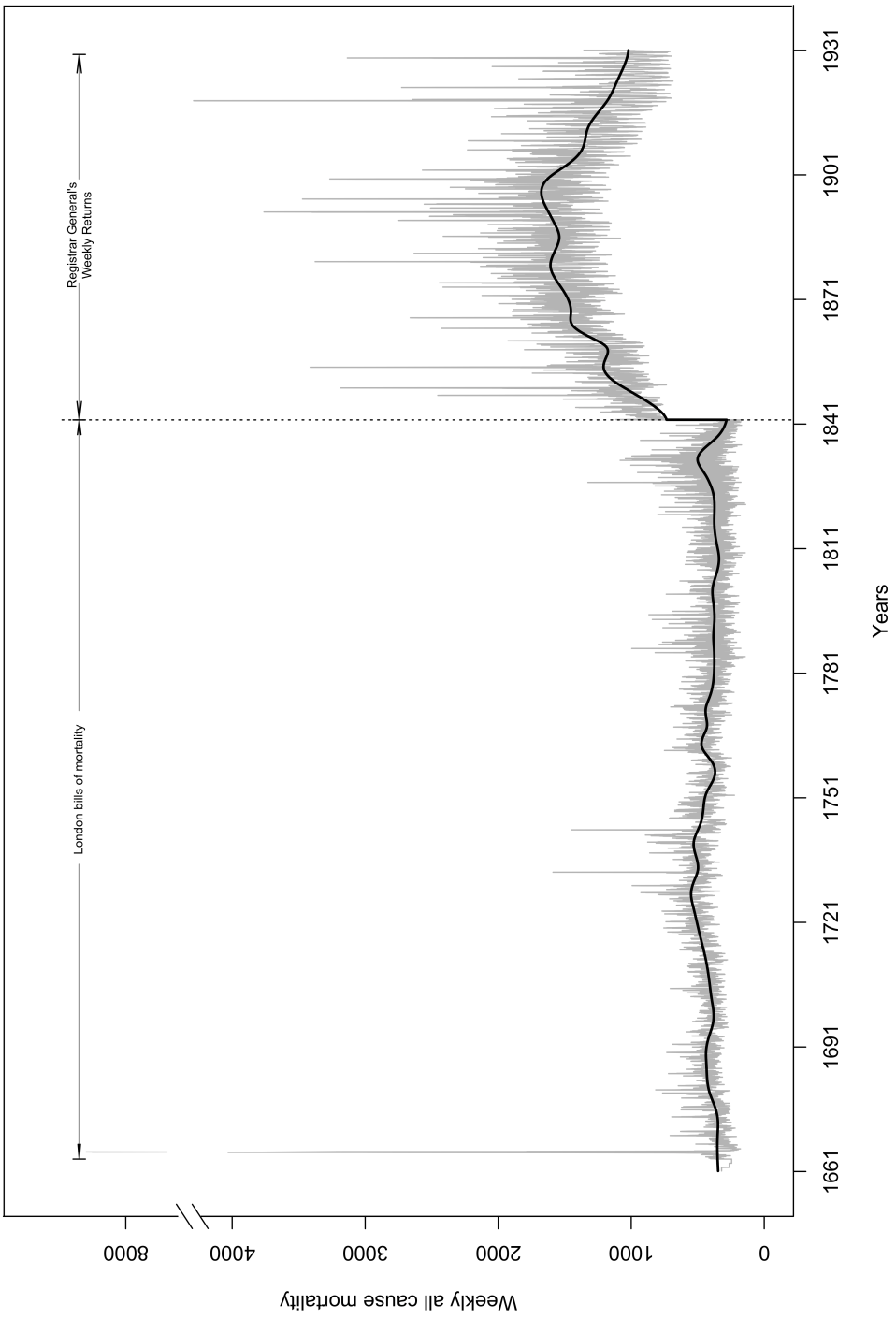


Figure 3.7: Weekly all-cause mortality and its trend, London, England (1661-1930). The trend was estimated by Empirical Mode Decomposition applied separately to the periods 1661-1842 and 1842-1930, which correspond to different data sources. The largest peak of all-cause mortality occurred during the Great Plague of London (1664-1665), which killed over 8,000 people in one week.

quent IMFs have lower average frequency (longer period). Each IMF is extracted recursively starting from the original time series until there are no more oscillations in the residue. The last residual component of this process can be considered as an estimate of the trend (Wu et al. 2007). **Figure 3.7** shows the trend for the weekly all-cause mortality in London (1661-1930) computed by EMD. (See **Appendix 3.5** for a detailed description of the EMD method.)

The normalized weekly smallpox time series (**Figure 3.8**) provides a more consistent representation of smallpox dynamics. For example, the epidemic of 1838 is now easily identifiable, and the epidemic of 1871-1872 is similar in magnitude to typical epidemics of the 18th century (though still the largest of the 19th century). **Figure 3.8** also shows the general trend in smallpox mortality calculated by the EMD method. The trend is nonlinear and shows that the smallpox mortality rate was increasing steadily from 1664 until its peak in approximately 1770, after which we see a gradual decline until its complete elimination. The decline can be associated with the growing popularity of variolation in the late 1760s inspired by the Suttonian innovation. Despite the rapidly growing London population, smallpox mortality continued to decrease even further and at a much greater rate after the discovery of vaccination at the end of the 18th century.

The time plot of the normalized smallpox deaths clearly shows the presence of different time frequency components in the series. From the earliest data in the series, the epidemics appear to be very severe and exhibit regular oscillations. The changes in variolation uptake after 1770 seem to modify the general pattern and the epidemics occur with almost constant periodicity. After the introduction of vaccination in 1796, the amplitude of epidemics is dramatically reduced and periodicity of the time series changes. During the period when variolation and vaccination were both in use (1796-1840), the data appear to be noisier and outbreaks occurred more frequently. After the outlawing of variolation in 1840 there is another change in the pattern: the inter-epidemic period increases and epidemic peak heights decline,

except for the large epidemics in 1871, 1876 and 1902.

3.4.3 Spectral analysis

Further analysis was performed to quantify periodicity in the pattern of smallpox epidemics and how the frequency structure changed over time. By comparing this spectral timeline with estimated levels of variolation and vaccination uptake, we are able to frame hypotheses concerning how these preventative measures influenced smallpox dynamics.

Fourier (spectral) analysis is the classical technique used to determine the frequency components of observed time series (Anderson and May 1991; Bauch 2008; Bauch and Earn 2003a; Chatfield 1989; Earn 2009; Shumway and Stoffer 2006). It helps to identify major periodic modes present in the data. The Fourier spectral density of the normalized smallpox mortality is plotted in **Figure 3.9** as a function of period (inverse of frequency). It reveals the existence of major periods at 1, 2.2, 2.4, and 3 years. A larger period of about 5.1 and 6 years also appear to be detected.

The Fourier transform determines all major periodic components that are present in the data but does not reveal any information about the time of their occurrence. The wavelet transform (Addison 2002; Cazelles et al. 2007, 2008; Torrence and Compo 1998) is a much more sophisticated method of time series analysis that overcomes this problem and is more appropriate for the analysis of non-stationary data. It provides a detailed picture of how the periods change over time. This technique has recently become very popular in the analysis of epidemiological data (Bauch 2008; Cazelles et al. 2007; Grenfell et al. 2001; Magney et al. 2007).

The method of wavelet analysis can be briefly summarized as follows. First, we select the basic shape function called the *wavelet function* or *analyzing wavelet*, which depends on two parameters: time and scale. The scale parameter allows us to narrow the wavelet function to represent high-freuqency modes or widen for low-

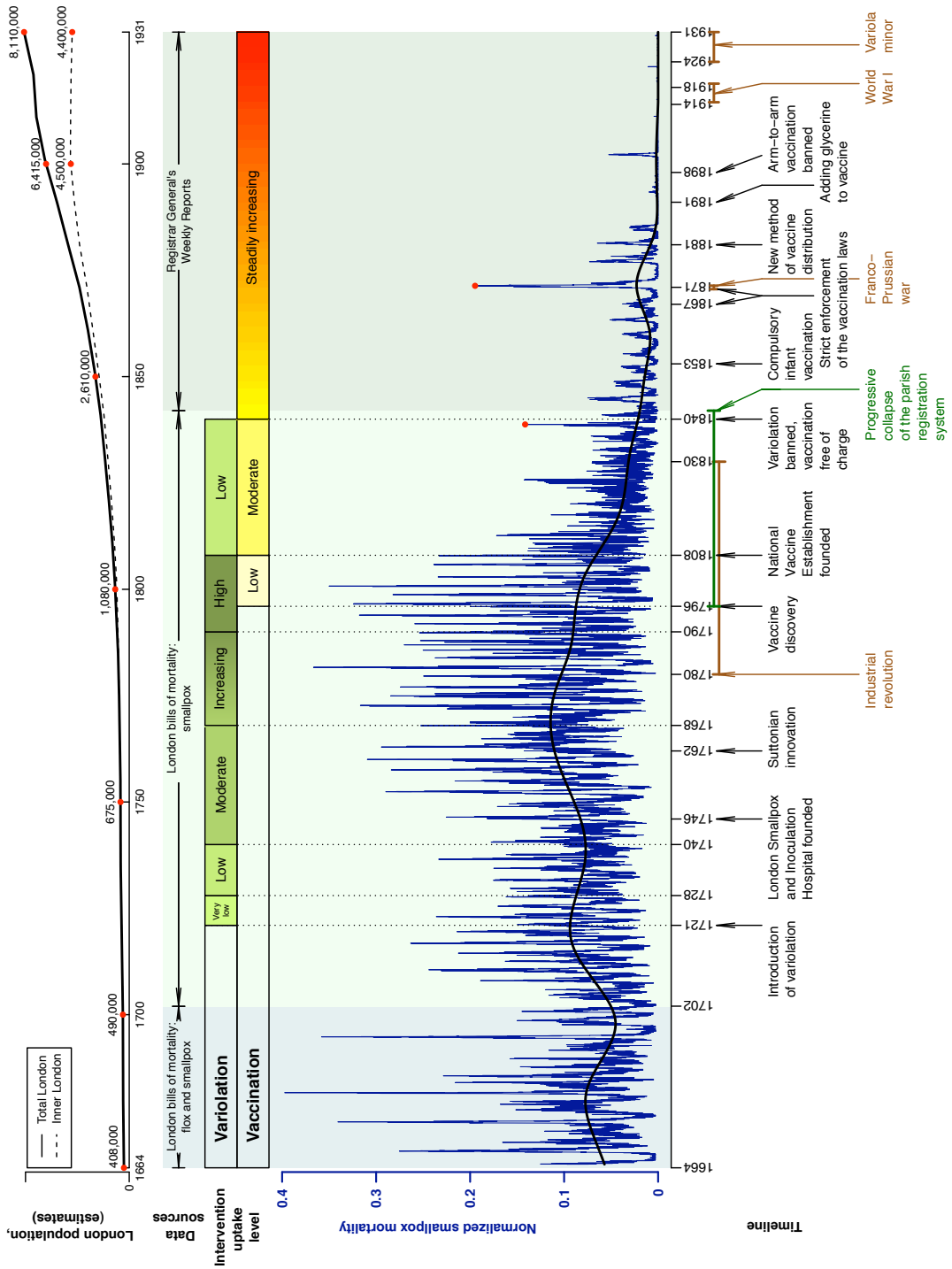


Figure 3.8: Time plot of weekly smallpox mortality normalized by the trend of weekly all-cause mortality. The trend (solid black curve) was estimated by Empirical Mode Decomposition. Red dots denote the peaks of the epidemics of 1838 and 1871-1872, the most significant smallpox epidemics of the 19th century. Annotation is as in **Figure 3.5**.

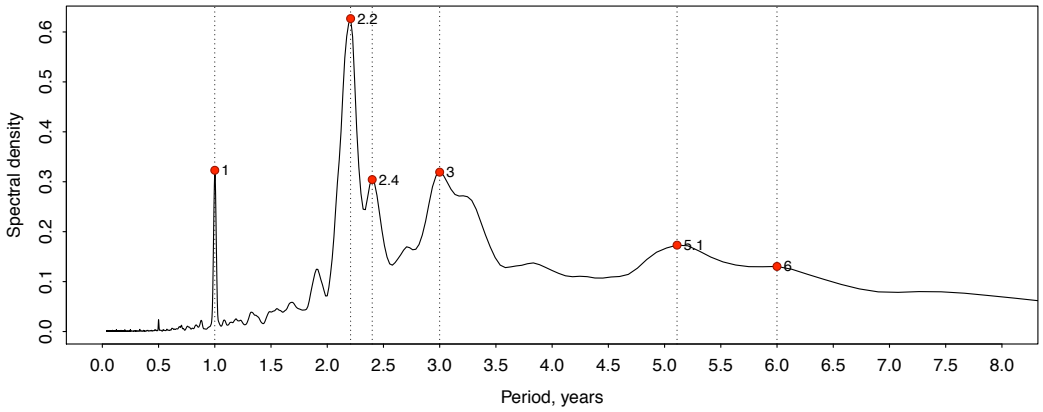


Figure 3.9: Classical Fourier transform of the normalized weekly smallpox mortality time series for London, England (1664-1930). Before computing the Fourier spectrum the time series was detrended and square root transformed (Chatfield 1989). The Fourier spectrum was smoothed using a Daniel window (weighted moving average transformation) (Chatfield 1989).

frequency structure. The analyzing wavelet is localized at a particular scale and time and then matched with the data by computing the convolution of the wavelet function with the time series. If the wavelet correlates well with the signal at a specific time and location, a larger value of the wavelet transform is obtained. Otherwise it does not correlate well with the signal and one obtains a smaller value of the wavelet transform. Moving the wavelet function along the time-series and over a continuous range of scale parameter allows us to obtain a 2-dimensional diagram (often called a wavelet scalogram) (e.g. **Figure 3.10**), which shows a time series decomposed into the time and frequency domain.

The resulting wavelet diagram of the London smallpox data is shown in **Figure 3.10**. The wavelet transform values are colour coded: blue represents the lowest power and dark red the highest. The algorithm used to compute the wavelet power spectrum requires that the length of the analyzed series is a power of two. Therefore, zero-values are added ("zero padding") to bring the number of time points in the data to the closest power of two. This creates an artificial discontinuity at the ends of the data. Hence the accuracy of the wavelet transform is reduced at the edges

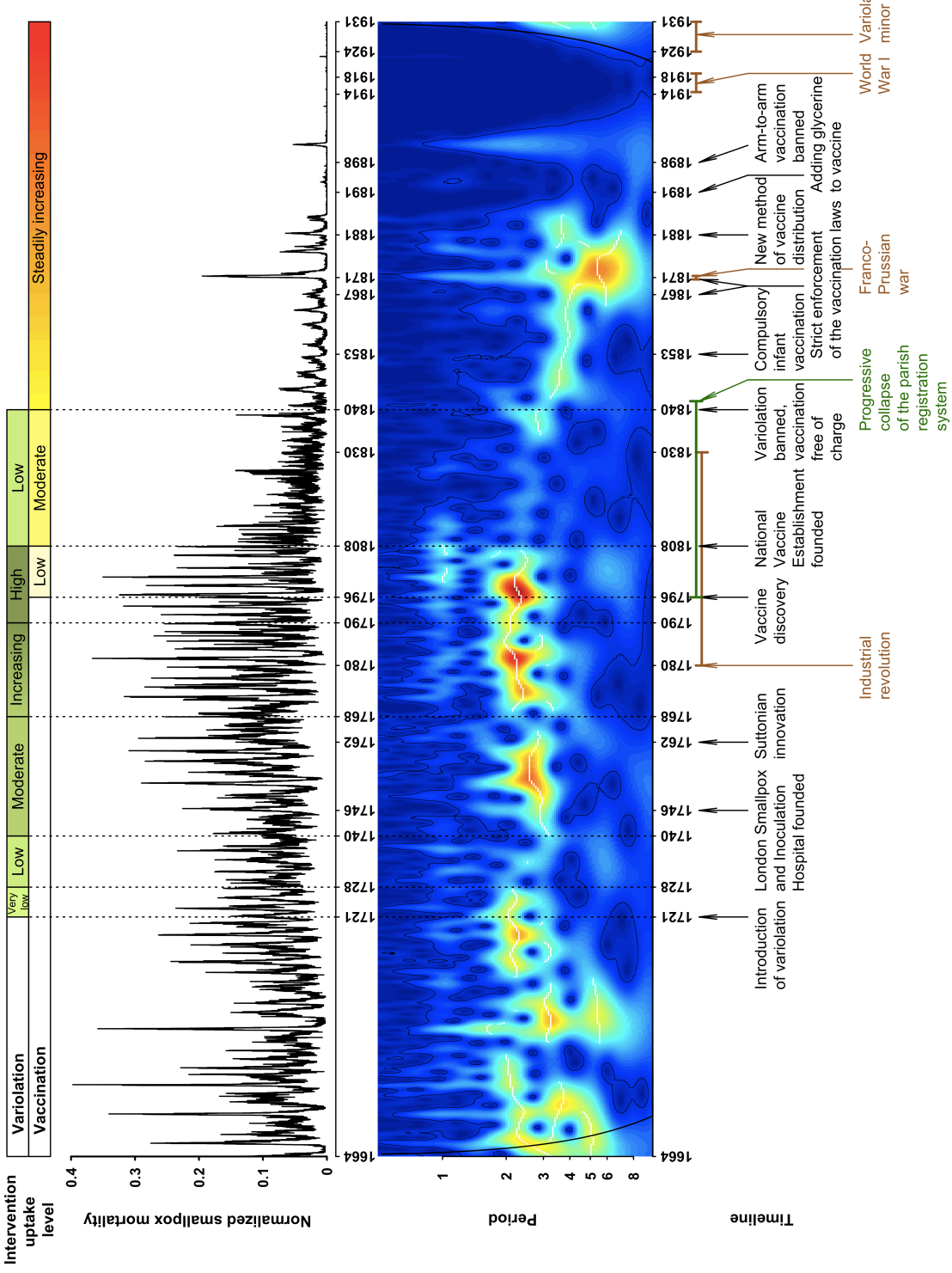


Figure 3.10: Wavelet transform of the normalized weekly smallpox mortality time series (square root-transformed and normalized to unit variance) for London, England (1664-1930) and its correlation with the historical timeline. **(continued on the next page)**

Figure 3.10: The white curves show the local maxima of wavelet power (squared modulus of wavelet coefficients (Cazelles et al. 2008, p.291) at each time. The colours of the wavelet diagram vary from dark blue, which corresponds to low power, to dark red for high power. The thin black line indicates the 95% confidence region, estimated from 1000 bootstrapped time series generated by the method of (Cazelles et al. 2008, p.292-293). Below the cone of influence (Cazelles et al. 2008; Torrence and Compo 1998), the calculation of wavelet power is less accurate because it includes edges of the time series that have been zero-padded to make the length of the series a power of 2. The wavelet spectrum was computed using MATLAB code kindly provided by Bernard Cazelles.

of the time series. To show the regions with lower accuracy a "cone of influence" is drawn. The data outside this cone should be interpreted with caution. Statistical significance of the results is computed based on 1,000 Markov bootstrapped series (Cazelles et al. 2007, 2008) and the 95% confidence region is presented as a thin black line on the diagram.

The wavelet transform of the smallpox mortality data is presented in **Figure 3.10**. It shows how the periodicity of smallpox epidemics changed over the centuries (the white curves highlight the period with greatest power at each timepoint). From 1664 until 1700 the dominant period is 3–4 years. Around 1705 dominant period shifts to 2–3 years. After the introduction of variolation, changes in periodicity are evident: the wavelet power weakens and the dominant period lengthens. Around 1740 a 3-year mode dominates until 1770. We can recognize the coexistence of multiple cycles of 1 and 2-3-years in the period between 1770 until around 1810. After 1820 the annual cycle disappeared and after 1840 the major period smoothly evolves into a longer cycle of 3-4 and then 4-8 years. Interestingly, these changes coincide with the onset of the vaccination era in 1800 and later The Vaccination Act of 1840 that banned variolation and made vaccination free of charge. The increase in vaccination uptake could have been responsible for the increase in inter epidemic intervals during the vaccination era and for the disappearance of the annual cycle. Such effects of vaccination on infectious disease dynamics have been

suggested before in the theoretical modeling work (Anderson and May 1991; Earn et al. 2000b) as well as in time series analysis of different epidemiological data sets (Earn et al. 2000b; Grenfell et al. 2001). For example, Grenfell et al. (2001) described similar features in measles data and explained less frequent epidemics with lower amplitude by the introduction of vaccination .

3.4.4 Seasonality

Seasonality of disease outbreak patterns is of great interest from a strictly biological perspective (Nelson et al. 2002). It also has dynamical significance since seasonal forcing can be responsible for the recurrence of epidemics and changes in their periodicity (Bauch and Earn 2003a; Earn et al. 2000a). Previous work on seasonality of smallpox (Duncan et al. 1996; Fenner et al. 1988; Nishiura and Kashiwagi 2009) concluded that in temperate climates the majority of smallpox incidence occurred in winter and spring, but in tropical climates seasonality was not as distinctive. The general conclusion was that smallpox incidence always increases when the weather is cool and dry, which influenced the planning of the eradication campaign in India and helped to improve its efficiency (Fenner et al. 1988, p.179-181). Previous seasonality studies were mainly based on data from the 19th and the 20th centuries when preventative measures were already implemented to various degrees (Fenner et al. 1988; Nishiura and Kashiwagi 2009). Our data set is of particular interest since it includes a period when only naturally acquired smallpox immunity existed. Consequently, we can compare early and later periods and access the impact of intensive preventive measures on smallpox seasonality.

Figure 3.11 shows the seasonal structure of smallpox mortality in London. It suggests the presence of seasonality in the observed time series and illustrates how it changed over the centuries. During the period when mostly naturally acquired smallpox immunity existed (1664-1740), the maximum number of smallpox deaths occurred in the summer and autumn. Then the peaks of the outbreaks shifted

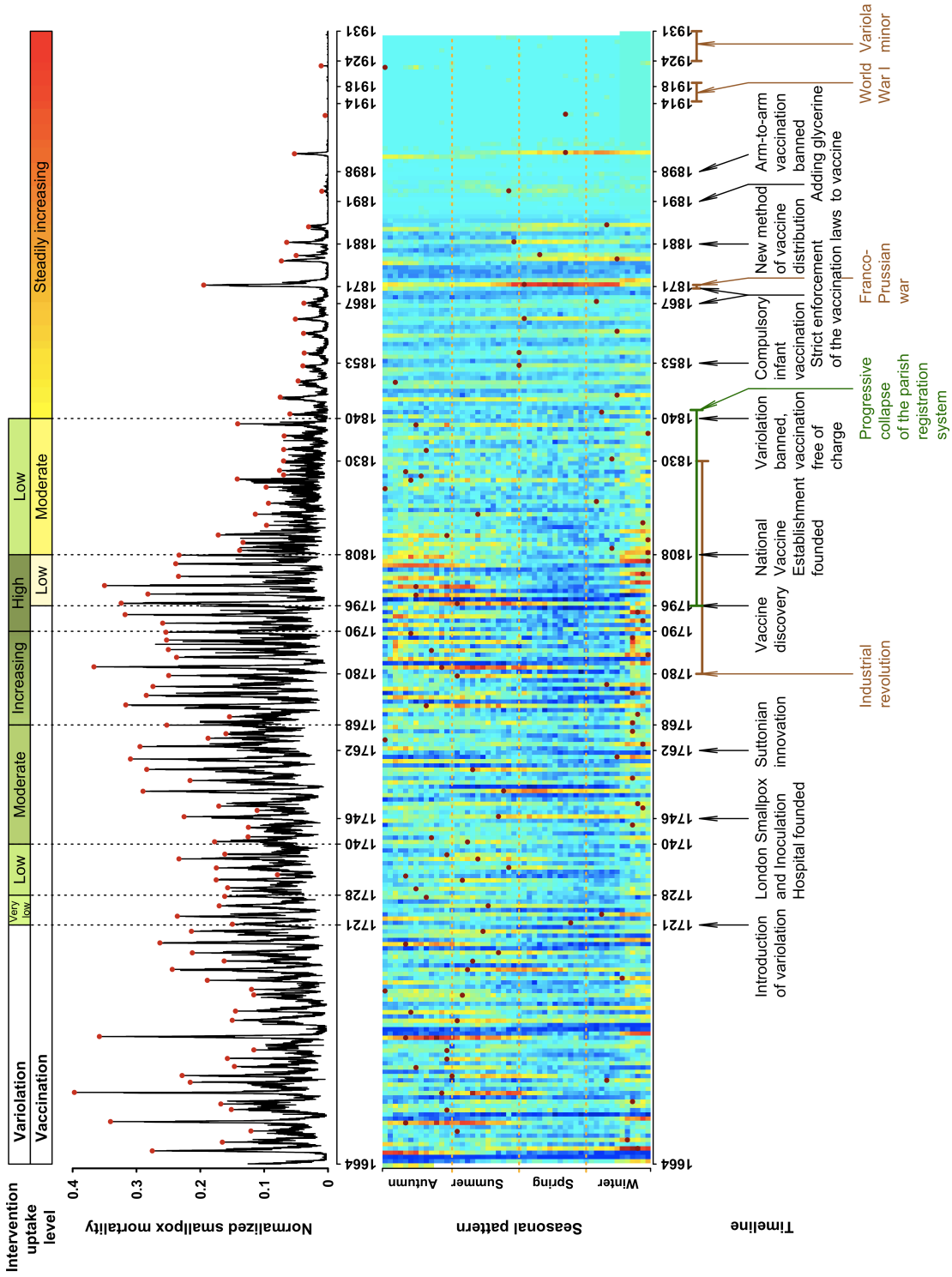


Figure 3.11: Seasonal variations of smallpox mortality in London, England (1664-1930). **(continued on the next page)**

Figure 3.11: The **top panel** shows weekly smallpox mortality normalized by the trend of all-cause mortality together with variolation and vaccination uptake levels. The red circles show the major peaks in the data as detected by visual inspection. The **main panel** shows the seasonality of smallpox epidemics colour coded so that dark red represents the week with the highest mortality and dark blue with the lowest. The **bottom panel** shows the timeline of historical events related to the smallpox history in England.

to the winter months until about 1770. After that the majority of deaths mostly occurred in autumn and winter. Seasonal patterns became very irregular during the period 1808-1840. After 1840, as vaccination levels gradually increased, epidemics occurred regularly with the majority of deaths occurring in winter and spring. The peaks of epidemics are hard to detect and to correlate with seasonal changes because of the noisiness of the data and the large variance in the amplitude of the outbreaks. However, the low points of the outbreaks are much more visible in **Figure 3.11**. From the earliest data until about 1853, spring was the period when the fewest deaths occurred. But after 1853, the year when compulsory vaccination was introduced by the Vaccination Act, spring and occasionally winter became the seasons of highest mortality.

Our analysis of the weekly London smallpox time series reveals a strong correlation between changes in smallpox dynamics and historical events that influenced variolation and vaccination levels. Even before these measures succeeded in reducing the severity of epidemics, our wavelet analysis shows they affected the periodicity of smallpox outbreaks (**Figure 3.10**) and our seasonality analysis (**Figure 3.11**) shows they affected the seasonal pattern of smallpox outbreaks.

3.5 Discussion

This study presents a statistical description of weekly smallpox mortality in London, England, over almost three hundred years. Time series analysis revealed the

simultaneous coexistence of multiple periodic modes as well as the evolution of the relative importance of different periodic components. We also detected changes in the seasonal pattern of smallpox epidemics. We established a timeline of important historical events related to changes in variolation and vaccination uptake and correlated them with changes in smallpox dynamics.

Previous studies of historical smallpox in London analyzed annual smallpox mortality data from the London Bills of Mortality over the period 1647-1893 (Duncan et al. 1996, 1994a,b). Those studies computed major periodic components of the annual data using spectral analysis. To identify changes in periodicity they divided the time series into six different time intervals. However, those intervals are not defined by any historical events, which may be why the authors were led to conclude that changes in smallpox periodicity were caused by changes in birth rate and nutrition status. Our study shows direct correlation between changes in smallpox periodicity and the implementation of preventative measures and legislation regarding those measures. Another difference between previous studies and ours is the data set used. The annual data analyzed by Duncan et al. (1996) masks the presence of annual periodicity and seasonality in the data. Our analysis detected the presence of weak annual cycles from 1664 to 1768 and strong annual cycles from 1768 to 1820. We were also able to identify seasonal changes in the peak times and to correlate them with historical events.

In a companion paper, we use a mathematical model to investigate how implementation of different preventative strategies, rapid growth of the population, improvement of sanitary conditions and increases in life expectancy influenced the dynamics of smallpox epidemics in London. This study provides a good foundation for modeling work by identifying time periods during which model parameters were relatively constant. Our historical timeline allows us to approximate parameter values (e.g. rates of birth, variolation and vaccination) in successive time intervals and thereby reduce uncertainty in model specifications.

Appendix A: Population of London (1550-1931)

Year	Total London	Outer London	Inner London	Source
1550	120,000	120,000		Finlay and Shearer (1986)
1560	140,800	140,800		Estimated[1]
1580	180,400	180,400		Estimated[1]
1600	200,000	200,000		Finlay and Shearer (1986)
1620	297,000	297,000		Estimated[1]
1640	390,500	390,500		Estimated[1]
1650	391,450	391,450		Estimated[1]
1660	392,400	392,400		Estimated[1]
1680	468,600	468,600		Estimated[1]
1700	490,000	490,000		Finlay and Shearer (1986)
1735	660,000	660,000		Landers (1993)
1745	670,000	670,000		Landers (1993)
1750	675,000	675,000		Finlay and Shearer (1986)
1755	680,000	680,000		Landers (1993)
1765	730,000	730,000		Landers (1993)
1775	780,000	780,000		Landers (1993)
1785	859,234	826,502		Landers (1993)
1795	1,007,703	909,507		Landers (1993)
1801	1,096,784	959,310	137,474	Census
1811	1,303,564	1,139,355	164,209	Census
1821	1,573,210	1,379,543	193,667	Census
1831	1,878,229	1,655,582	222,647	Census
1841	2,207,653	1,949,277	258,376	Census
1851	2,651,939	2,363,341	288,598	Census
1861	3,188,485	2,808,494	379,991	Census
1871	3,840,595	3,261,396	579,199	Census
1881	4,713,441	3,830,297	883,144	Census
1891	5,571,968	4,227,954	1,344,014	Census
1901	6,506,889	4,536,267	1,970,622	Census
1911	7,160,441	4,521,685	2,638,756	Census
1921	7,386,755	4,484,523	2,902,232	Census
1931	8,110,358	4,397,003	3,713,355	Census

Table 3.1: Population of London, England (1550-1931).

[1]Estimated based on (Finlay and Shearer 1986, Tables 2, 3) data.

Appendix B: A note on computing trends of demographic and epidemiological data by Empirical Mode Decomposition

Why should we use EMD?

Identifying temporal trends is very important in time series analysis. In many instances we are interested in long term smooth changes in data that can be seen from its trend. For example, the all cause mortality trend or the total births trend are important quantities we rely on while analyzing or modeling epidemiological time series. We also often require the removal of a trend from the data (detrending), for example when computing wavelet or Fourier power spectra.

The observed demographic and epidemiological time series that we usually work with while studying epidemics are often non-linear, noisy and highly non-stationary (frequency and amplitude change over time). Empirical Mode Decomposition (EMD) (Wu et al. 2007) is a recently developed tool that can be effectively used to find a non-linear trend in non-stationary data. It overcomes the flaws of other commonly used methods such as moving average, filtering and regression analysis, which may perform poorly when applied to noisy and non-stationary time series.

Moving average is argued not to be very useful for non-stationary processes since it requires a predetermined time scale, but a local time scale is unknown a priori for non-stationary signal (Wu et al. 2007). Moreover, the moving average method has a "boundary problem", i.e. the trend cannot be computed for some portions of the data at the beginning and the end of the time series. That can potentially be a problem. If the aim is to have a very smooth trend, the moving average window may be too long and we might lose many data points because of this. In the example I have considered below, moving average (MA) produces a trend similar to

the EMD trend. However the MA trend is noisier and loses 12 years of trend data.

Regression analysis and filtering, which are based on linear and stationary assumptions, use specific models and curve fitting to determine a trend. However the driving mechanisms in epidemiological and demographic data are complex and often unknown a priory. Therefore a non-parametric and adaptive EMD method is preferable.

EMD Approach

The main idea behind the EMD method is that any signal (or time series) consists of several components with a well defined instantaneous frequency. EMD decomposes a signal into a set of "intrinsic mode functions" (IMFs). IMFs are basically zero mean oscillation modes present in the data: the first IMF captures the higher frequency oscillations while all subsequent IMFs have a lower average frequency. Each IMF is extracted recursively starting from the original time series until there are no more oscillations in the residue. The residual components of this process can be considered to be estimates of the trend (Wu et al. 2007).

The algorithm to find IMFs starts with computing successive extrema of the signal (min and max). These extrema are then connected by cubic splines to form the upper and lower envelopes. The mean of these envelopes is then computed. Ideally after subtraction of the mean from the original data we would get the IMF; however that is not always the case, the extracted component may not satisfy the IMF criteria. The repetition of the above process, which is called *sifting*, is needed to find the right IMF (Huang et al. 1998; Rilling et al. 2003). The number of siftings used to extract each IMF should be chosen with caution, since over-iteration may result in over-decomposition (Wu et al. 2007).


package "EMD" allows to compute trends by EMD method for any data (EMD). There also exists a MATLAB package for this purpose (Rilling et al. 2003).

EMD Algorithm:

To find the trend of a time-series $x(t)$ the following steps are performed (Rilling et al. 2003; Wu et al. 2007):

1. all local extrema of $x(t)$ (min and max) are identified;
2. interpolating using cubic spline between minima and maxima, the envelopes $e_{\min}$ and $e_{\max}$ are created;
3. low-frequency part is then computed as the mean, $m_j(t) = (e_{\min} + e_{\max})/2$;
4. high-frequency part is computed as the difference, $h_1(t) = x(t) - m_1(t)$;
5. steps 1 to 4 (sifting process) are iteratively repeated on the component $h_j(t)$, until the stopping criterion is met. Note $h_{j+1}(t) = h_j(t) - m_j(t)$;
6. when stopping criterion is met the sifting process is repeated some finite number of times, k. The last $h_k(t)$ is defined as the Intrinsic Mode Function;
7. the residue is then $r_i(t) = x(t) - \text{IMF}_i$, $i \in 1, \ldots, n$;
8. above steps are iteratively repeated again on the residue, $r_i(t)$, until another IMF is obtain and so on until the residue becomes a monotonic function from which no further IMFs can be extracted.

At the end the data is decomposed into several components: from the highest frequency, IMF_1, to the lowest frequency, IMF_n and residue, r_n.

$$x(t) = \sum_{i=1}^{n} \text{IMF}_i(t) + r_n(t) \tag{3.1}$$

the total number of IMFs is close to $\log_2 N$ where N is the total number of data points.

The following is based on the help page of the R EMD Package (function `emd`) to which I've added some information.

Function `emd`

There are many useful functions inside of the EMD package. For computing the trend of the time series, use the `emd` function (EMD).

Description

This function performs empirical mode decomposition.

Usage

```
emd(xt, tt=NULL, tol=sd(xt)*0.1^2, max.sift=20,
 stoprule="type1",boundary="periodic", smlevels=c(1),
 sm="none", spar=NA, weight=20, check=FALSE,
 max.imf=10, plot.imf=TRUE, interm=NULL)
```

Arguments

`xt`	observation or signal observed at time `tt`
`tt`	observation index or time index
`tol`	tolerance for stopping rule of sifting
`max.sift`	the maximum number of sifting (the sifting process stops if the `max.sift` is reached or stopping criteria for obtaining IMF is met)
`stoprule`	stopping rule of sifting has two options: `type 1`: ($\lvert h_i(t)\rvert <$ `tol`) and `type 2`: $\sum_t \left(\frac{h_i(t)-h_{i-1}(t)}{h_{i-1}(t)}\right)^2 <$ `tol`
`boundary`	used to eliminate boundary effect of the data (specifies the adjusting method of the boundary from: "`none`" - no boundary adjustments, "`wave`" - constructs a wave, "`symmetric`" or "`periodic`" - assumes data is symmetric or periodic, "`evenodd`" - see Kim and Oh (2009) for explanations)
`smlevels`	specifies which level of the IMF is obtained by smoothing other than interpolation
`sm`	specifies whether the envelope is constructed by smoothing spline
`spar`	specifies user-supplied smoothing parameter of spline
`weight`	the smoothness of spline is determined by weight times smoothing parameter of GCV
`check`	specifies whether the sifting process is displayed (if `check=TRUE`, click the plotting area to start the next step)
`max.imf`	the maximum number of IMFs
`plot.imf`	specifies whether each IMF is displayed (if `plot.imf=TRUE`, click the plotting area to start the next step)
`interm`	specifies vector of periods to be excluded from the IMFs

Value

`imf`	IMFs
`residue`	residue signal after extracting IMFs from observations `xt`
`nimf`	the number of IMFs

Examples

As an example we will identify the trend in **all-cause mortality** data from London, UK (1661-1930) by EMD.

```
## read data
data <- read.table
    ("your_data_file", header=TRUE, sep = ",");
## define time verctor
time <- data$numdate;
## observed acm
acm <- data$acm;
## plot
plot(time, acm,
    type="l", xlab="Time", ylab="ACM");
```

Visual analysis of the data plotted in **Figure 3.12** gives the following observations:

- There is an enormous spike in mortality in 1665, which corresponds to the time of Great Plague of London epidemic.
- There is an obvious discontinuity in observations in 1842, which occurred due to using two different data sources: London Bills of Mortality (1661-1842) and Registrar Generals Weekly Returns (1842-1931).

It seems that the best approach to calculate the trend of this time series would be to compute it separately for periods 1661-1842 and 1842-1931. The spike in 1665 should also be removed and replaced with, for example, interpolated data, so that trend calculation would not be effected by abrupt discontinuities of the original data.

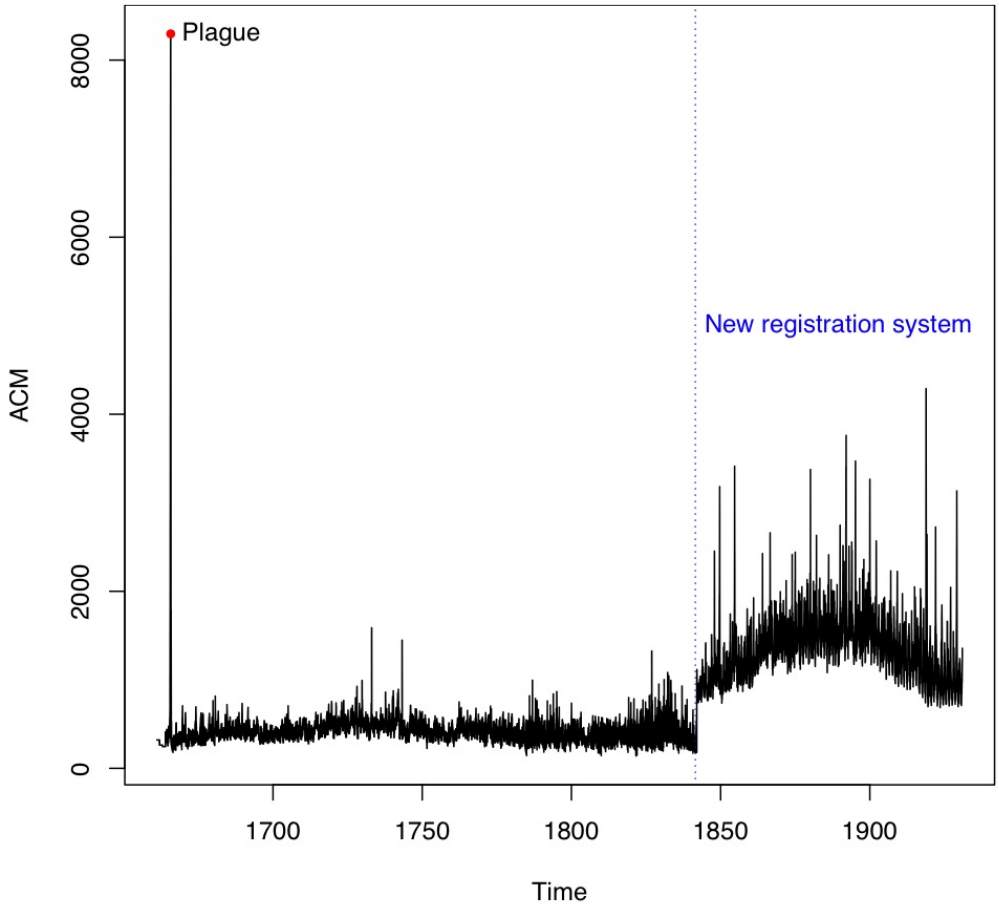
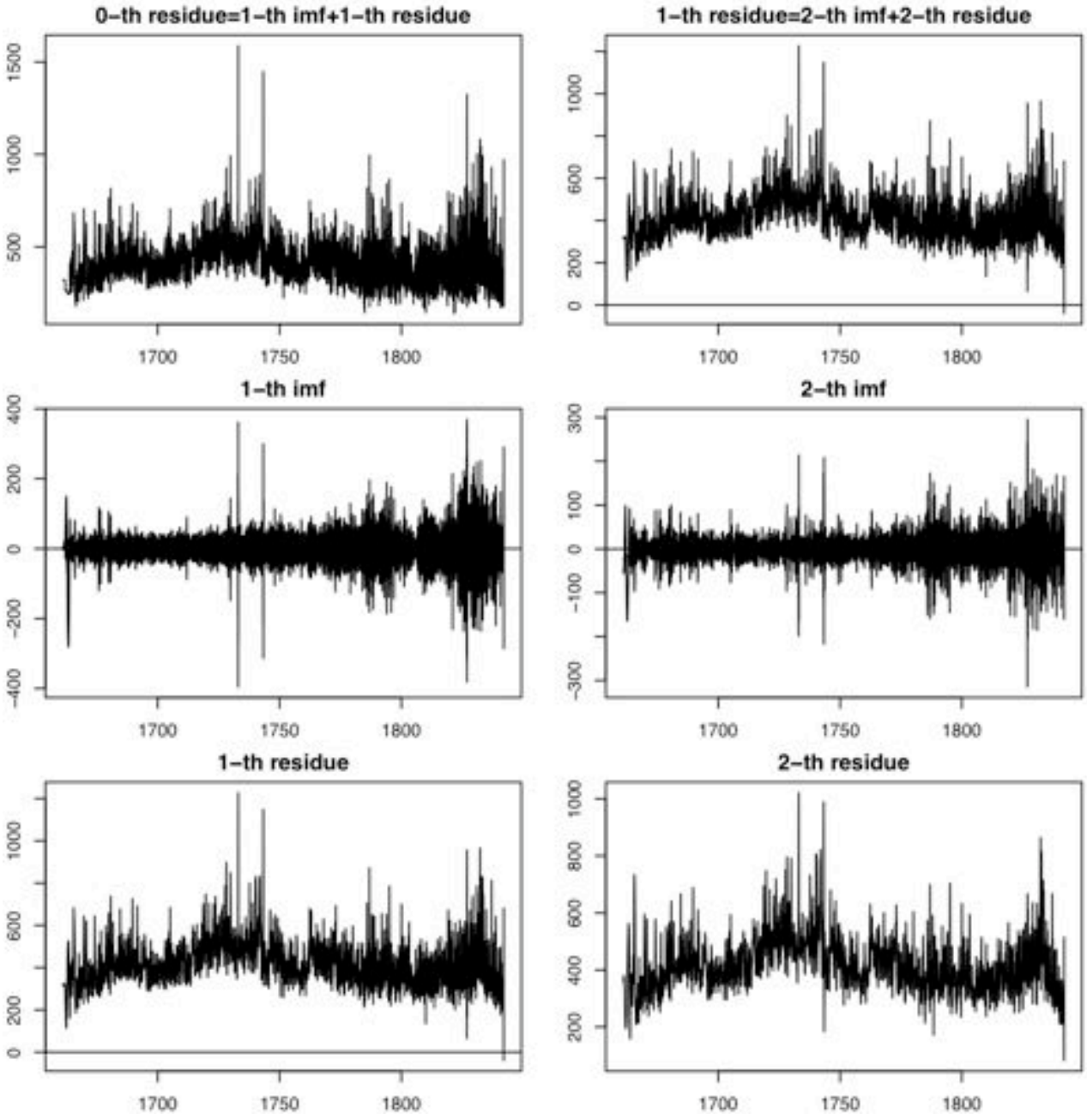


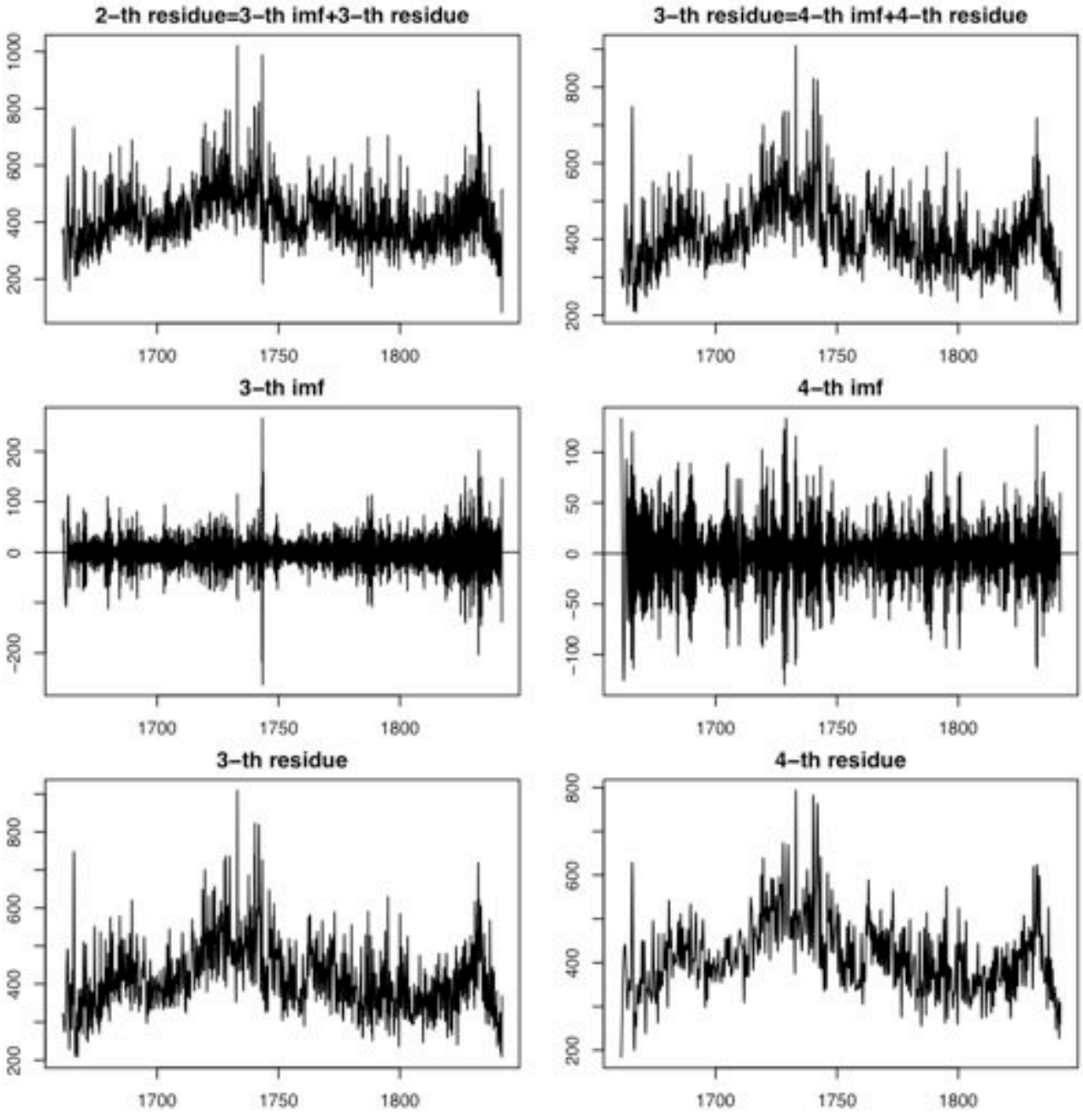
Figure 3.12: London all-cause mortality, 1661 - 1930.

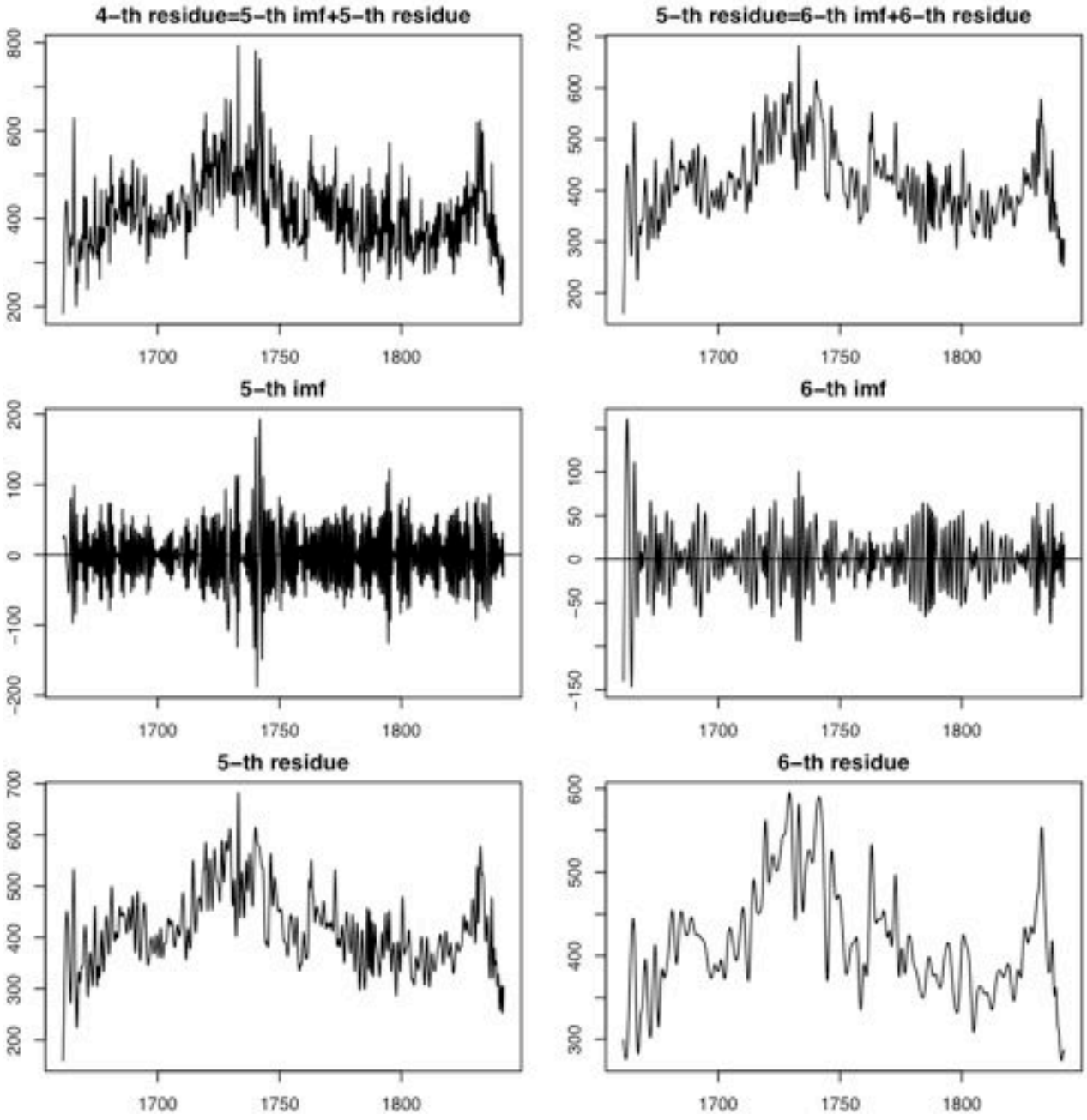
The following code produces **Figure 3.5**, which then will be used to determine a trend.

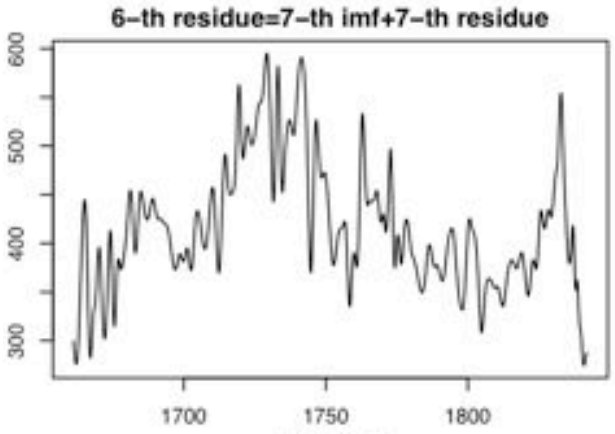
```
## spike during 1665 plague epidemic was removed
## and replaced with linearly interpolated values
## new data:
acm <- data$acm.no.plague;
## identify beginning of new registration system
index2 <- which(round(time,0)==1842)[1];
```

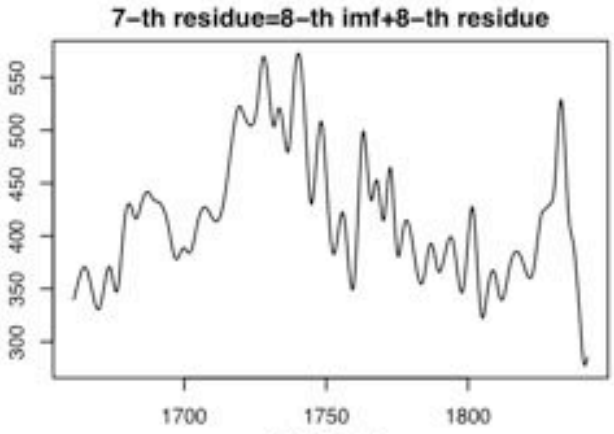
```
#######################
## COMPUTING TREND  ###
#######################
## load EMD package
library(EMD);
##
## Compute trend for the part 1 of the data (1661-1841):
## save all produced pictures in pdf
pdf("file_name.pdf");
## set plotting window
par(mfcol=c(3,2), mar=c(2,2,2,2));
## compute emd
emd.out.part1 <- emd(acm[1:index2], time[1:index2],
    boundary="wave", max.imf=20, max.sift=20, plot.imf=TRUE);
## close pdf
dev.off();
## compute trend as 8th residue
acm.trend.part1 <-
    acm[1:index2]-rowSums(emd.out.part1$imf[ ,1:8]);
## plot original data
plot(time[1:index2], acm[1:index2],
    type="l", ylab="ACM", xlab="Time");
## plot trend
lines(time[1:index2], acm.trend.part1, col="red", lwd=2);
```

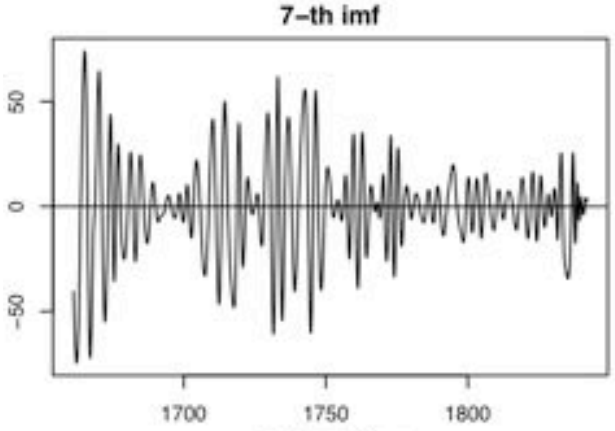

0−th residue=1−th imf+1−th residue
1−th residue=2−th imf+2−th residue
1−th imf
2−th imf
1−th residue
2−th residue
1700
1750
1800

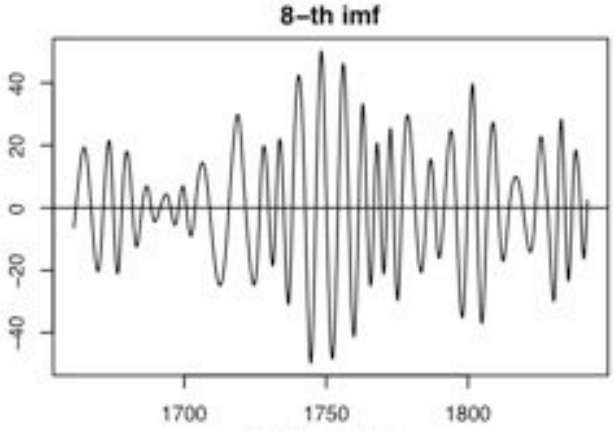

2−th residue=3−th imf+3−th residue
3−th residue=4−th imf+4−th residue
3−th imf
4−th imf
3−th residue
4−th residue
1700
1750
1800

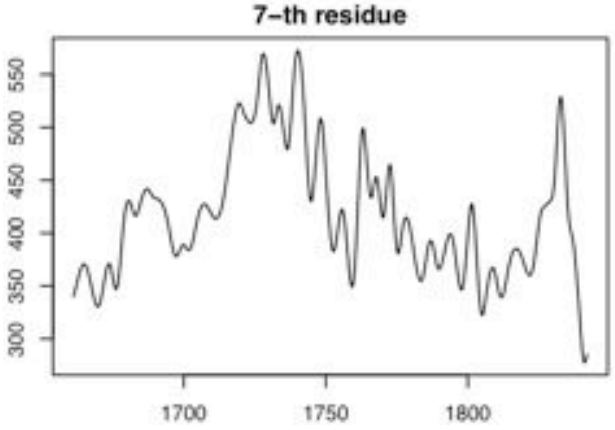

4−th residue=5−th imf+5−th residue
5−th residue=6−th imf+6−th residue
5−th imf
6−th imf
5−th residue
6−th residue
1700
1750
1800

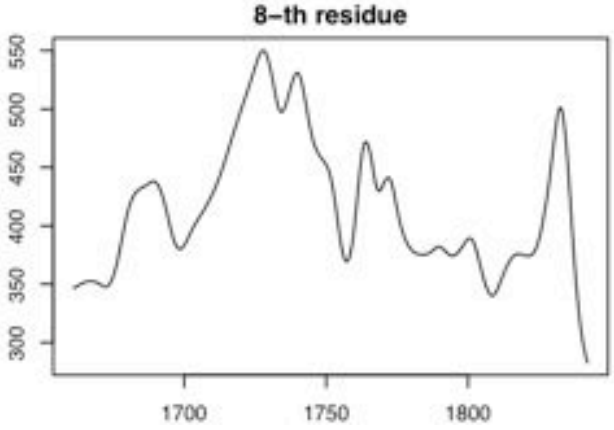

6−th residue=7−th imf+7−th residue
600
500
400
300
1700
1750
1800

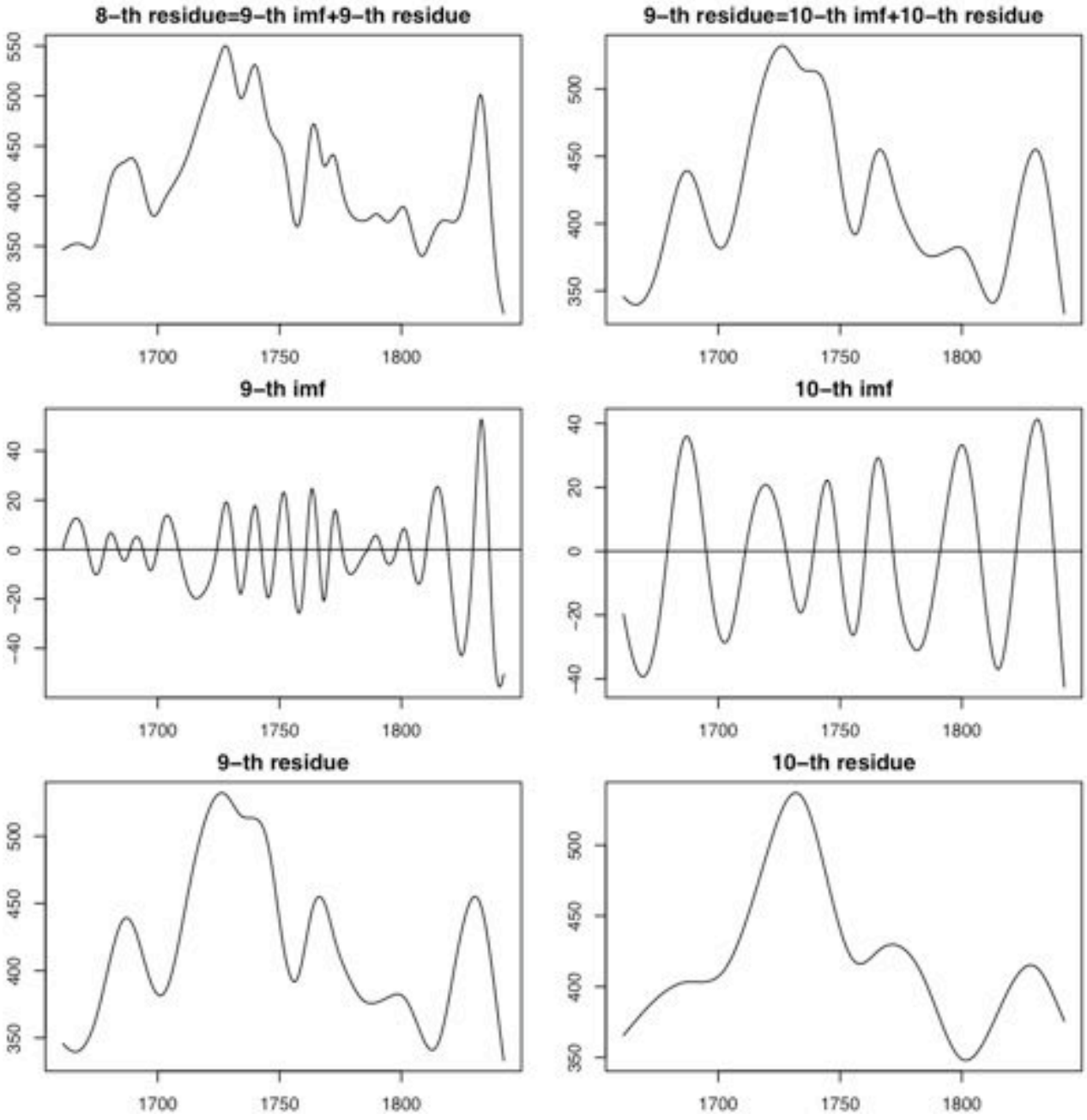

7−th residue=8−th imf+8−th residue
550
500
450
400
350
300
1700
1750
1800

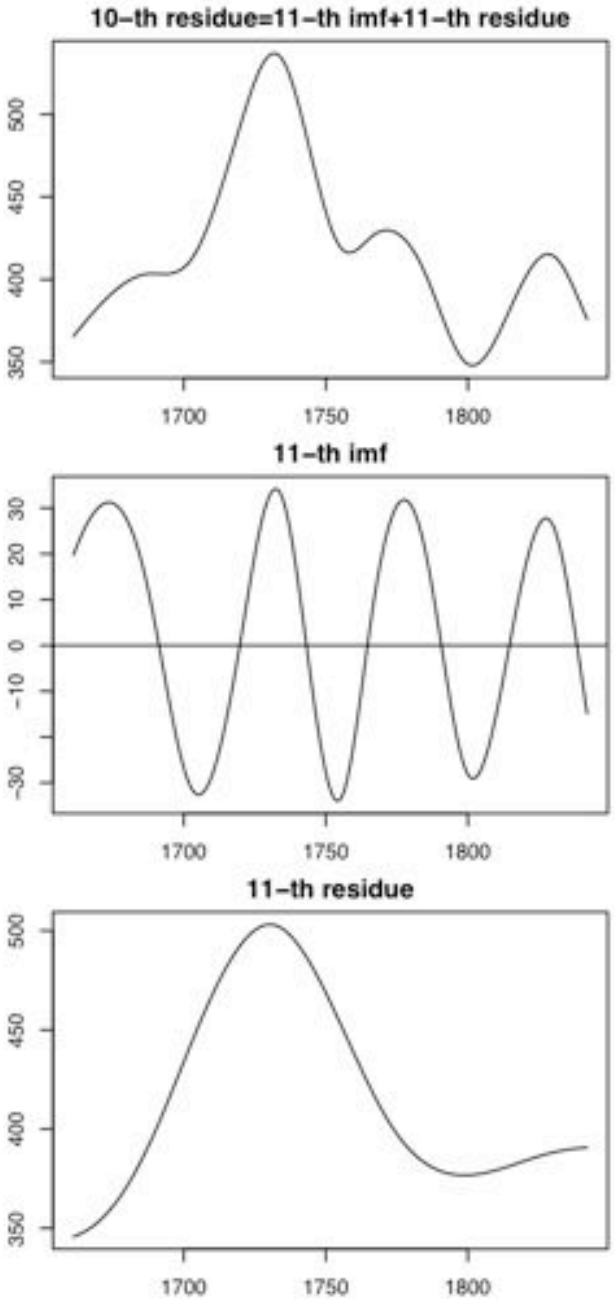

7−th imf
50
0
−50
1700
1750
1800


8−th imf
40
20
0
−20
−40
1700
1750
1800


7−th residue
550
500
450
400
350
300
1700
1750
1800


8−th residue
550
500
450
400
350
300
1700
1750
1800


8−th residue=9−th imf+9−th residue
9−th residue=10−th imf+10−th residue
9−th imf
10−th imf
9−th residue
10−th residue
1700
1750
1800


10−th residue=11−th imf+11−th residue
350
400
450
500
1700
1750
1800
11−th imf
−30
−10
0
10
20
30
1700
1750
1800
11−th residue
350
400
450
500
1700
1750
1800

From visual examination we conclude that all 8th to 11th residues can be considered as reasonably good trends. We would like the trend to capture some variability of the data and choose residue 8th as a trend for part 1 (**Figure 3.13**).

Note: It is not trivial to choose the right boundary option. We have tried all possible choices for the parameter `boundary` of the `emd` function from "`none`" to "`evenodd`" and concluded that the "`wave`" option works the best in this case.

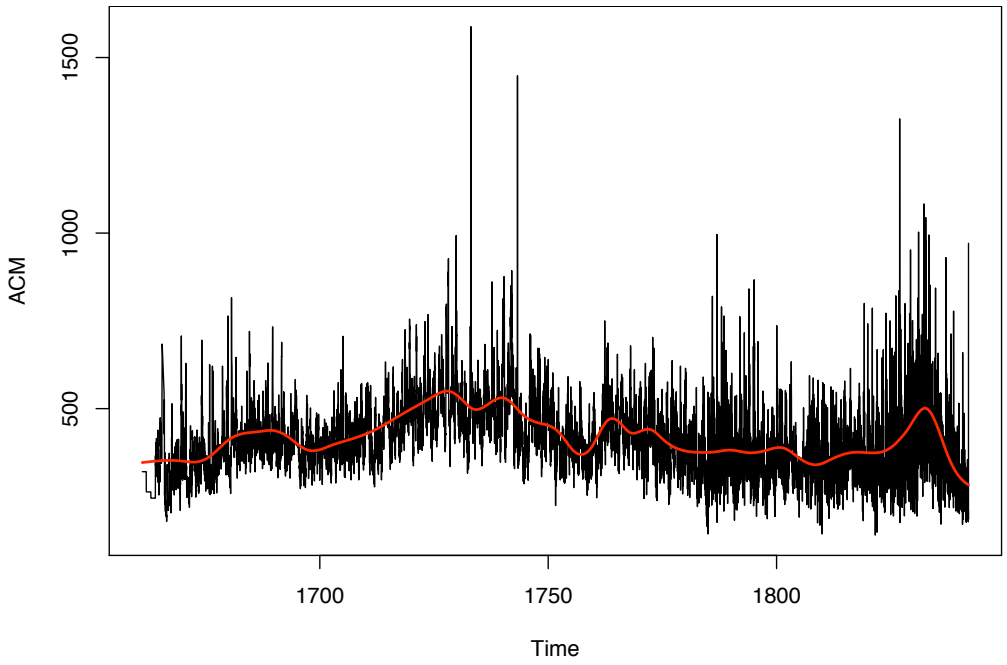


Figure 3.13: London all-cause mortality, 1661 - 1842.

Similarly we have computed the trend for part 2 of the data (1842-1931).

```
## Compute trend for part 2 of the data (1842-1931):
## save all produced pictures in pdf
pdf("file_name.pdf");
## set plotting window
par(mfcol=c(3,2), mar=c(2,2,2,2));
```

```
## compute emd
emd.out.part2 <- emd(acm[(index2+1):length(time)],
    time[(index2+1):length(time)], boundary="evenodd",
    max.imf=20, max.sift=20, plot.imf=TRUE);
## close pdf
dev.off();
## compute trend as 8th residue
acm.trend.part2 <-
    acm[(index2+1):length(time)]-rowSums(emd.out.part2$imf[ ,1:8]);
## plot original data
plot(time[(index2+1):length(time)],
    acm[(index2+1):length(time)],
    type="l", ylab="ACM", xlab="Time");
## plot trend
lines(time[(index2+1):length(time)],
    acm.trend.part2, col="red", lwd=2);
```

The `boundary` parameter option was chosen to be "`evenodd`" and the trend was identified as 8th residue (**Figure 3.14**).

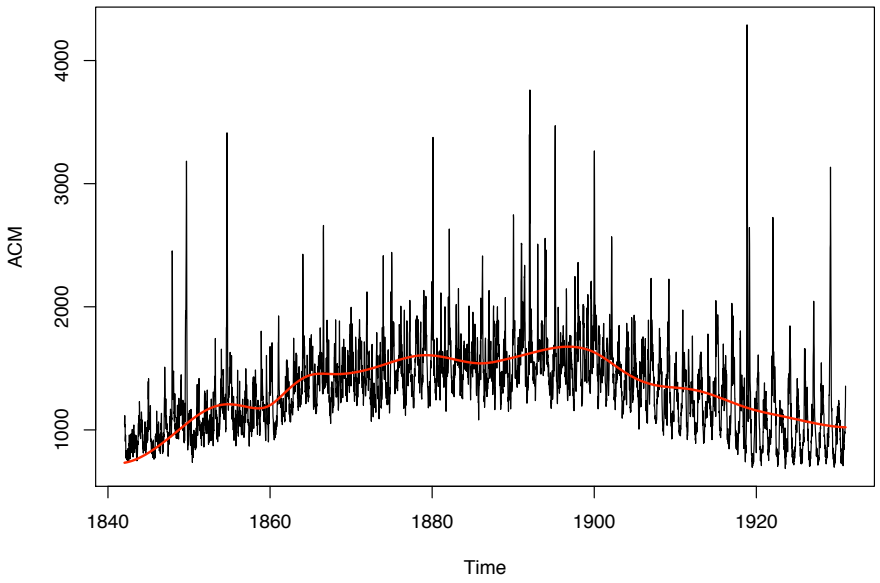


Figure 3.14: London all-cause mortality, 1842 - 1930.

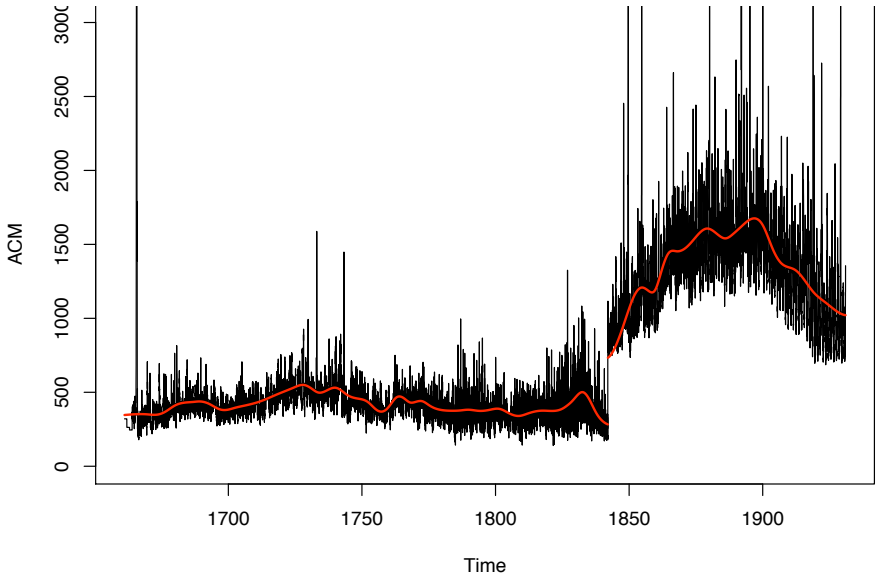


Figure 3.15: London all cause mortality (1661 - 1930) and its trend computed by EMD.

Figure 3.15 shows the end result: the trend for complete time series.

I have also compared this result with the moving average of 6 years in **Figure 3.16**. The EMD method was better at producing a smother trend and not losing any info at the boundaries. Notice that since we computed the trend separately for two parts of the data set, the MA method produced 12 years of missing values.

```
#########################
## TREND by MA Method
#########################
## 6 years (or 2*156 weeks) ma
## for part 1
ma1 <-  filter(acm[1:index2],
    filter=rep(1/(2*156+1),2*156+1));
## for part 2
ma2 <- filter(acm[(index2+1):length(time)],
    filter=rep(1/(2*156+1),2*156+1));
## plot original data
plot(time, data$acm.no.uhp, type="l",
```

```
    ylab="ACM", xlab="Time", ylim=c(0,3000));
## plot trend
lines(time[1:index2], ma1, col="blue", lwd=2);
lines(time[(index2+1):length(time)], ma2, col="blue", lwd=2);
```

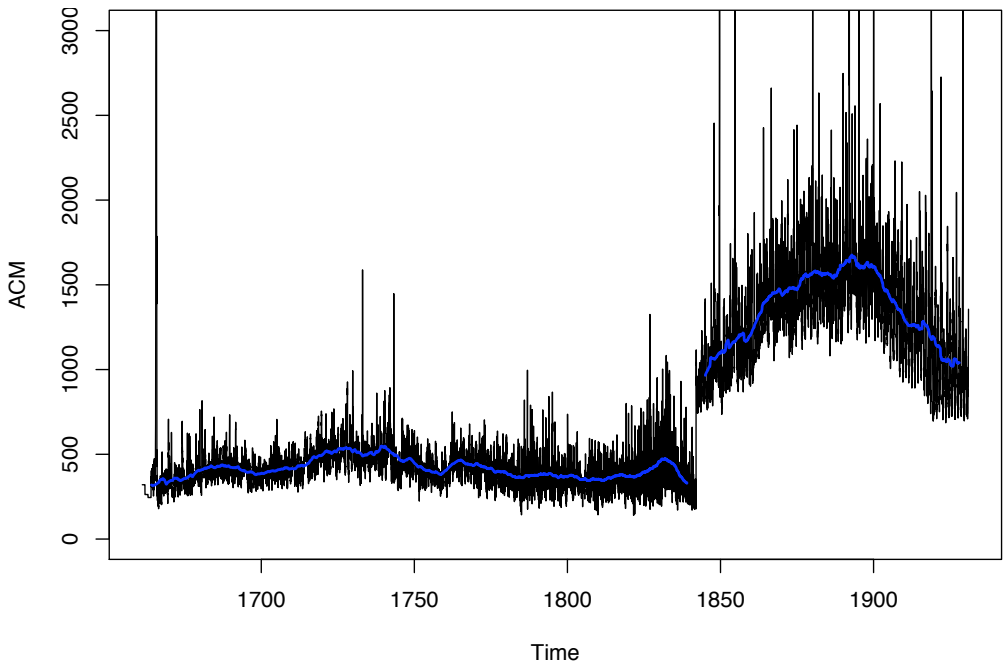


Figure 3.16: London all-cause mortality (1661 - 1930) and its trend computed by moving average (6 years window).

Where to look for more info

A more detailed description of EMD can be found in the following sources:

- Wu et al. (2007) - great starting point. This short paper will give you a good idea what is EMD and how it can be used. Authors also propose a definition of the trend of nonlinear and non-stationary data and define the process of data detrending;

- EMD - EMD package in  . Package description provides great examples of how to use different functions inside the package;
- Kim and Oh (2009) - offer more detailed explanations of the EMD package in  . It is very useful;
- Rilling et al. (2003) - description of the numerical algorithm for computing EMD used in MATLAB package;
- `http://www.commsp.ee.ic.ac.uk/~mandic/research/emd.htm` - various recently published papers about EMD;
- Huang et al. (1998); Vatche and Sharple (2008) - more on analytical approach used for developing EMD.

Chapter 4

Estimating the seasonal pattern of smallpox transmission in London, England, 1664-1930

Abstract

Seasonal variation in infectious disease transmission is a key determinant of epidemic dynamics. Estimation of the amplitude and seasonal pattern of the transmission rate is necessary for the development of accurate epidemiological models. In this paper we use a simple and computationally efficient method to estimate the transmission rate, $\beta(t)$, for smallpox in London, England, from 1664 until 1930. Our analysis reveals that seasonality of $\beta(t)$ changed significantly over the centuries. The events associated with lowest transmission change from the annual harvest in the 17th century to the summer school holiday in the 19th century. After the introduction of public schools at the end of the 19th century, variations in the smallpox transmission rate are similar to the variations in the measles transmission rate in England and Wales (1950-1977), which were previously associated with school terms (Fine and Clarkson 1982; Finkenstadt and Grenfell 2000). Our results suggest that the contact rate among individuals, particularly children, who became

the primary victims of smallpox in the 18th century (Creighton 1965; Fenner et al. 1988; Razzell 1977), was the driving force of smallpox transmission in London.

Keywords: smallpox; seasonal pattern; transmission rate; differential equation model

4.1 Introduction

The study of epidemiological time series via mathematical modeling opened a route for understanding the biological processes that drive the spread of infectious diseases in populations. The complicated mechanism of transmitting infection between susceptible and infected individuals, which depends on the type of population mixing, contact pattern between individuals, the type of contact and various other factors, is typically approximated in S(usceptible)-I(nfectious)-R(ecovered)-type models by a simple term, βSI. The transmission rate, β, is the product of probability of a susceptible individual acquiring infection if a contact with an infected person occurs and the rate at which pairwise contacts occur in the population. For many infectious diseases the transmission rate varies seasonally (Altizer et al. 2006) and for childhood diseases it often mimics the school-term pattern (Conlan et al. 2010, 2011; Eames et al. 2011; Fine and Clarkson 1982; Hooker et al. 2011). It has been shown that the shape of the seasonal variation of β determines the dynamical structure of the SIR-type models (Earn et al. 2000a) and therefore influences inferences and predictions derived from the models. Hence obtaining adequate estimates of the seasonal pattern of the transmission parameter, β, is very important. Moreover, estimated seasonality of the transmission rate could help to identify the driving forces of the spread of infectious diseases (e.g. changes in the contact rate between individuals, seasonal variations in temperature and humidity, *etc.*).

Most studies that aim to estimate the disease transmission rate seasonality are based on relatively short data sets (20-50 years) (Fine and Clarkson 1982;

Finkenstadt and Grenfell 2000; He et al. 2010; Hooker et al. 2011; King et al. 2008) and therefore cannot provide information about long-term changes in the seasonality of disease transmission over centuries. Such information could help to clarify the fundamental epidemiological processes that affect epidemic patterns. The Weekly London Bills of Mortality, analyzed in this study, present a unique opportunity to examine changes in the seasonality of smallpox transmission over 267 years. The main purpose of the current work was to use this extremely rich data set to estimate and analyze the seasonal pattern of the smallpox transmission rate and make some inferences about major forces that drive smallpox seasonality.

Over the last three decades various methods of estimating the transmission rate from observed data have been proposed (Cauchemez and Ferguson 2008; Fine and Clarkson 1982; Finkenstadt and Grenfell 2000; Hooker et al. 2011; King et al. 2008; Metcalf et al. 2009; Xia et al. 2005). One of the earliest methods was developed by Fine and Clarkson (1982). They obtained plausible estimates of the measles transmission rate (**Figure 4.1**) using a crude discrete-time approximation of the deterministic SIR model. This method was later modified by Finkenstadt and Grenfell (2000), who developed a discrete time stochastic version of the SIR model termed the TSIR (time series SIR) model and estimated the transmission rate by fitting this model to measles data. The resulting estimated seasonal pattern of measles transmission (**Figure 4.3**) was very similar to the one obtained earlier by Fine and Clarkson (1982) (**Figure 4.1**). Recently, more sophisticated methods that take into account observation error, demographic noise, non-linearity and non-stationarity of the underlying biological process, have been developed (He et al. 2010; Hooker et al. 2011; Ionides et al. 2006; King et al. 2008).

While these methods represent state-of-the-art statistical techniques they are often computationally demanding and time consuming to implement. Since our main purpose was to estimate seasonality of the smallpox transmission from an extremely long time-series (13889 data points) we required a computationally ef-

ficient method that could be implemented in a reasonable time frame. Therefore, we developed a simple approach of estimating the transmission parameter based on the continuous-time SEIR model. Our resulting formulas resemble the ones used by Fine and Clarkson (1982). However, our method takes into account reporting delay, which was ignored in Fine and Clarckson's approach, and can be applied not only to case notification, but also to disease mortality data. It is also free of the assumption that the generation time is equal to the observation interval, and therefore does not require unnecessary temporal aggregation of the data (tri-weekly in the case of smallpox). The method is straightforward in implementation and very efficient. It produces accurate results when tested on simulated data, and when applied to the data sets used previously in the literature for estimating disease transmission rates (it reconstructs the same seasonal pattern as other more sophisticated methods).

In the next section we describe our approach. We apply the method to smallpox mortality data and present our principal findings in **Section 4.3**. We summarize our results and future research directions in the last section of the paper (**Section 4.4**).

4.2 Methods

Consider the standard SEIR model:

$$\frac{dS}{dt} = (1 - p(t))\nu(t) - \beta(t)S(t)I(t) - \mu(t)S(t)\,, \tag{4.1a}$$

$$\frac{dE}{dt} = \beta(t)S(t)I(t) - (\sigma + \mu(t))E(t)\,, \tag{4.1b}$$

$$\frac{dI}{dt} = \sigma E(t) - (\gamma + \mu(t))I(t)\,, \tag{4.1c}$$

$$\frac{dR}{dt} = \gamma I(t) - \mu(t)R(t)\,, \tag{4.1d}$$

where $S(t)$, $E(t)$, $I(t)$, and $R(t)$ are continuous stable variables that represent the proportions of individuals who are *Susceptible*, *Exposed*, *Infectious* and *Recovered*,

and $\beta(t)$ is the (per capita) transmission rate. All other parameters are explained in **Table 4.1**.

Symbol	Definition
$1/\sigma$	Mean latent period
$1/\gamma$	Mean infectious period
β	Per capita transmission rate
ν	Birth rate
μ	Death rate
p	Proportion of newborns immunized, $p = p^{\text{var}} + p^{\text{vacc}}$
p^{var}	Proportion of newborns successfully immunized by variolation
p^{vacc}	Proportion of newborns successfully immunized by vaccination
Δt	Time interval between successive incidence samples
Z_t	Number of new infections at time t
S_t	Number of susceptible individuals at time t
S_1	Initial number of susceptible individuals in the population
B_t	Number of births at time t
C_t	Number of reported cases at time t
M_t	Reported disease mortality at time t
T	Mean time from initial infection to recovery, $T = k\Delta t$, where $k = 1,\ 2,\ 3,\ \ldots$
T_{report}	Mean time from initial infection to its reporting, $T_{\text{report}} = m\Delta t$, where $m = 1,\ 2,\ 3,\ \ldots$
η	Case fatality proportion
ρ	Reporting rate (proportion of cases reported)

Table 4.1: Notation used in this paper.

If we assume that the disease dynamics in a population are governed by the mechanisms captured in this model, we can estimate the transmission rate β from incidence data using **equation** (4.1) in the following manner. Let Z_t be the number of new infections aggregated over some time interval, Δt. This number can be derived from the rate of increase in the latent class (**equation** (4.1b)) as:

$$Z_t = \int_{t-\Delta t}^{t} \beta(\tau) S(\tau) I(\tau) d\tau, \tag{4.2}$$

or, if we take into account that individuals who where infected at time t leave the infectious class at time $t + T$ on average (where T is the mean time from initial infection to recovery, $T = \frac{1}{\sigma} + \frac{1}{\gamma}$), the number of new infections, Z_t, aggregated over the time interval, Δt, can be derived from the rate of decrease in the infectious class (**equation** (4.1c)):

$$Z_t = \int_{t-\Delta t}^{t} (\gamma + \mu(\tau + T))I(\tau + T)d\tau. \tag{4.3}$$

Assuming that $\beta(t),\ S(t),\ I(t)$ and $\mu(t)$ are almost constant over the short time interval, Δt, **equations** (4.2) and (4.3) can be rewritten as:

$$Z_t \cong \beta_t S_t I_t \Delta t, \tag{4.4}$$

and

$$Z_t \cong (\gamma + \mu_{t+T})I_{t+T}\Delta t. \tag{4.5}$$

Equation (4.5) yields a formula for I_t:

$$I_t = \frac{Z_{t-T}}{(\gamma + \mu_t)\Delta t}, \tag{4.6}$$

which can be inserted into **equation** (4.4) to obtain:

$$Z_t = \beta_t S_t \frac{Z_{t-T}}{\gamma + \mu_t}. \tag{4.7}$$

Hence the transmission rate β_t can be obtained from the incidence time series (Z_t) via

$$\beta_t = \frac{1}{S_t}\frac{Z_t}{Z_{t-T}}(\gamma + \mu_t), \tag{4.8}$$

where S_t, based on **equation** (4.1a), is:

$$S_{t+\Delta t} = S_t + (1 - p_t)B_t - Z_{t+\Delta t} - \mu_t S_t. \tag{4.9}$$

Note that **equation** (4.8) reconstructs the transmission parameter, β_t, from the number of new infections, Z_t (i.e. infections that occurred between $t - \Delta t$ and t). In general, infections (Z_t) are not observable and only case notification data or disease mortality data are collected. Therefore we have to account for a delay (T_{report}) between the time of infection and the time of case notification or death and treat only a proportion $\rho < 1$ as reported. Assuming that both parameters, T_{report} and ρ, are constant, reported cases, C_t, can be estimated as:

$$C_t = \rho Z_{t-T_{\text{report}}}, \quad T_{\text{report}} \in \{\Delta t, 2\Delta t, \ldots\}. \tag{4.10}$$

Similarly, if the case fatality proportion, η, is constant, **equation** (4.10) can be altered for mortality data, M_t:

$$M_t = \rho\eta Z_{t-T_{\text{report}}}, \quad T_{\text{report}} \in \{\Delta t, 2\Delta t, \ldots\}. \tag{4.11}$$

Equations (4.10)mortality allow us to rewrite **equations** (4.8) and (4.9) in terms of reported cases, C_t,

$$\beta_t = \frac{1}{S_t}\frac{C_{t+T_{\text{report}}}}{C_{t+T_{\text{report}}-T}}(\gamma + \mu_t), \tag{4.12a}$$

$$S_{t+\Delta t} = S_t + (1-p_t)B_t - \frac{1}{\rho}C_{t+\Delta t+T_{\text{report}}} - \mu_t S_t, \tag{4.12b}$$

or reported disease induced deaths, M_t:

$$\beta_t = \frac{1}{S_t}\frac{M_{t+T_{\text{report}}}}{M_{t+T_{\text{report}}-T}}(\gamma + \mu_t), \tag{4.13a}$$

$$S_{t+\Delta t} = S_t + (1-p_t)B_t - \frac{1}{\rho}\frac{1}{\eta}M_{t+\Delta t+T_{\text{report}}} - \mu_t S_t. \tag{4.13b}$$

Before estimating the transmission rate, β, from the reported data (**equations** (4.12) and (4.13)) we need to determine the initial number of susceptibles in the population, S_1. We assume that the average level of susceptibles should be roughly con-

stant over a 5–10 year period. Therefore we choose S_1 (from the range sN_1, where $0 < s < 1$ and N_1 is the initial population size) so that the reconstructed time series of S_t (**equations** (4.12b) and (4.13b)) for the 5–10 years period has no long-term trend.

Using **equations** (4.12) and (4.13) we estimate the transmission parameter, β_t, for the whole time series. To identify the seasonal component of β_t, we remove the long-term trend using the EMD method (Wu et al. 2007), which we have found to be the best for determining the trend in noisy, non-stationary data. The seasonal pattern changes from year to year, but is roughly constant for periods of many years. Hence we divide the time series into a sequence of time intervals during which the seasonal pattern is approximately constant. For each of these time intervals, we estimate the annual seasonal pattern to be given by the mean across all years in the interval (e.g. the annual pattern of β_t on 1 January is $\beta_{1\text{ Jan}} = \frac{1}{n}\sum_{i=1}^{n}\beta_{1\text{ Jan},i}$, where n is the number of years in the focal time interval). We use the same approach to estimate a biennial and a triennial pattern, taking averages across biennia or triennia, respectively. We then compare the estimated annual, biennial and triennial patterns to examine whether there is evidence for seasonal forcing on a timescale longer than one year. If the seasonal pattern we estimate is still noisy, we smooth it using a moving average or EMD. When plotting the estimated seasonal pattern, we also show the upper and lower quartiles (calculated across years) to indicate the degree of variation in the seasonal pattern with each time interval with roughly constant seasonality.

Note that **equations** (4.12a) and (4.12b) are similar to the ones used by Fine and Clarkson (eq. 2–3). However, **equation** (4.12a) includes the term $(\gamma + \mu_t)$ and accounts for a reporting delay, which were omitted in Fine and Clarkson (1982). Our method does not rely on the assumption that the generation time is equal to the sampling interval, and therefore does not require unnecessary temporal aggregation of the data. However there is a limited flexibility in the choices of T_{report} and T,

which must be constant multiples of the sampling interval, Δt.

To check the validity of our method, we tested it on simulations of the standard deterministic SEIR model (**equation** (4.1)) as well as its stochastic version (Gillespie simulations) with fixed and sinusoidally forced $\beta(t)$. The reconstructed $\beta(t)$ was identical to the one used in simulations. We also applied the method to the weekly measles case notification data from England and Wales (1950-1965), used previously by Fine and Clarkson (1982) and Finkenstadt and Grenfell (2000), to check that our method produces comparable results. Our estimated seasonality of β_t presented in **Figure 4.2** (annual seasonal pattern (blue curve) with its upper and lower quartiles (grey region)) are very similar to the results obtained by Fine and Clarkson (1982) (panel B of **Figure 4.1**) and Finkenstadt and Grenfell (2000) (**Figure 4.3**). Note that we plot β/γ, rather than β, so the vertical scale has the units of the basic reproduction number, $\mathcal{R}_0$, for easier interpretation. We also were able to reproduce the seasonal pattern of β_t obtained by Hooker et al. (2011) for weekly measles incidence data for Ontario, Canada (1939-1965). (**Figure 4.4** shows results from Hooker et al. (2011) and **Figure 4.5** shows the estimated β/γ obtained by our method). Note that there is a significant difference between the estimated mean of β/γ (the red dotted line in **Figure 4.5** and **Figure 4.2**) for measles in Ontario (mean $\beta/\gamma \approx 27$) and measles in England and Wales (mean $\beta/\gamma \approx 11$). This can be explained by different reporting rate (ρ) used in these studies ($\rho \approx 27\%$ in Hooker et al. (2011) vs. $\rho \approx 66\%$ in Fine and Clarkson (1982)). If we choose the same reporting rate, the estimated mean β/γ is almost identical in both cases. The above comparisons indicate that despite its simplicity, our method produces results that are consistent with the outputs of other methods.

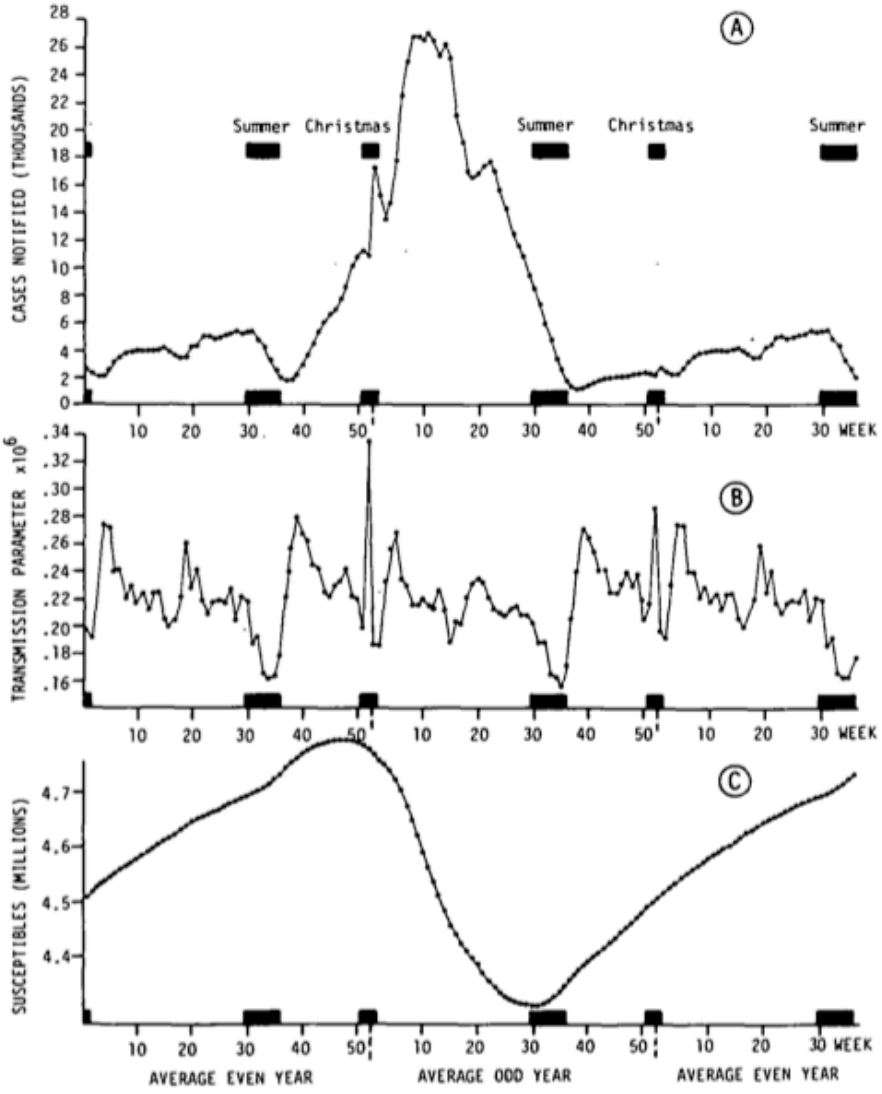


FIG. 3

Analysis of average biennial measles pattern, based on data from 1950 to 1965: (a) Average number of cases notified per week; (b) Calculated weekly transmission parameters; (c) Estimated number of susceptibles per week. Shaded blocks indicate school summer and Christmas holiday periods.

Figure 4.1: Seasonality of measles transmission, reprinted from Fine and Clarkson (1982)

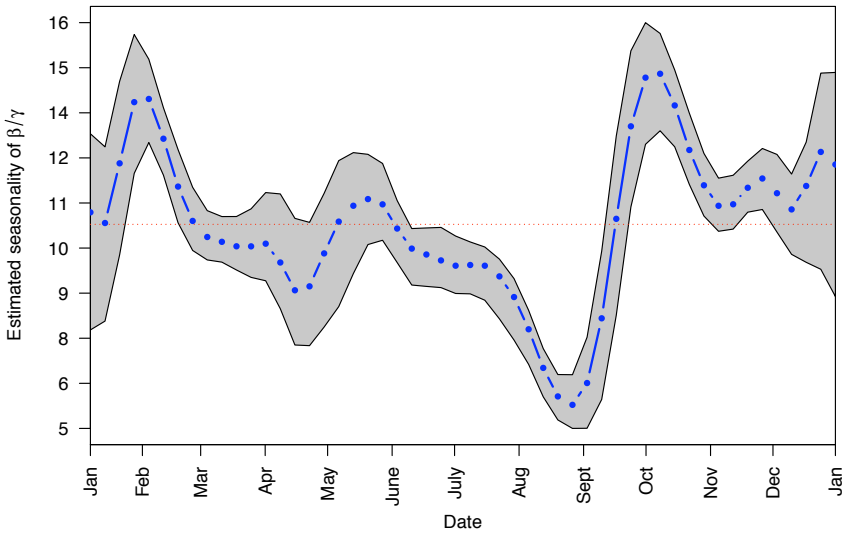


Figure 4.2: Estimated seasonality of β/γ for measles in England and Wales (1950-1965). Blue dots show the annual seasonal pattern of β/γ estimated by our method while the grey region shows the upper and lower quartiles. The red dotted line shows the mean value of β/γ. Parameter values were chosen as follows: $S_1 = 4.5 \cdot 10^6$, $B_t = 12,300\,\text{week}^{-1}$, $\mu = 0\,\text{year}^{-1}$, $p_t = 0$, $\rho = 66\%$ (the above parameters are as in Fine and Clarkson (1982)), $T_{\text{report}} = T = 2$ weeks, $1/\gamma = 5$ days.

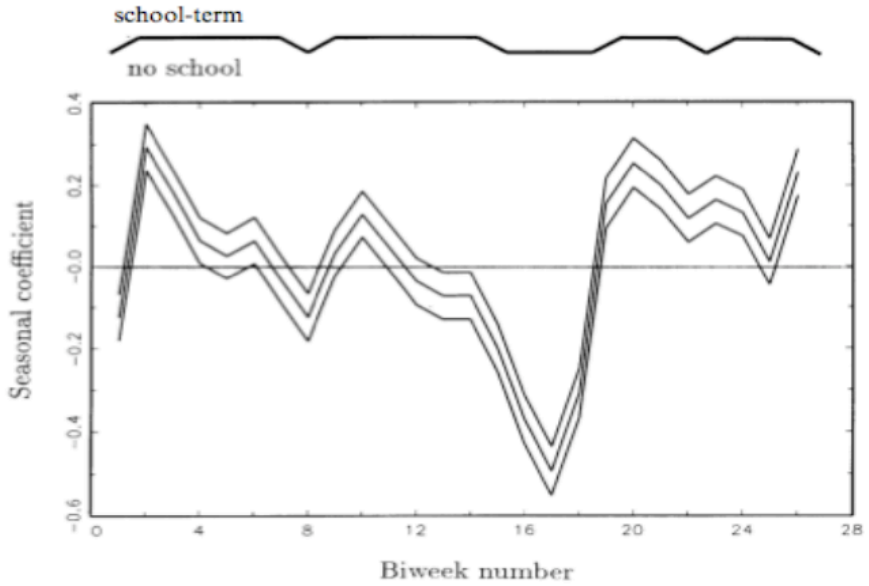


Fig. 7. Estimates of seasonal coefficients (mean centred) of measles in England and Wales within a confidence band of width two standard deviations: the uppermost graph shows the school-term indicator function (high level, school-term; low level, no school) for England and Wales

Figure 4.3: Seasonality of measles transmission, reprinted from Finkenstadt and Grenfell (2000).

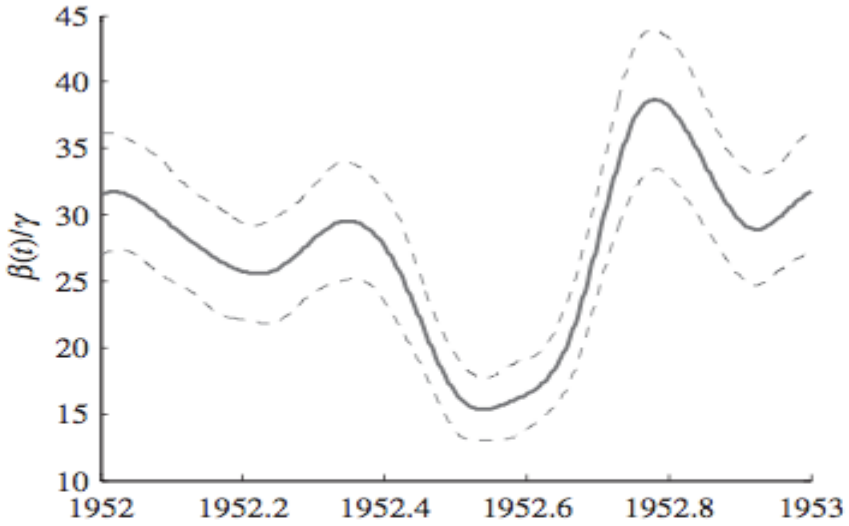


Figure 3. Estimated seasonal variation in the transmission rate (plotted for the year 1952; other years will be shifted slightly upwards or downwards owing to the long-term linear trend $\hat{\beta}_1$). Solid lines show the generalized profiling estimate, while the dashed lines show pointwise 95% CI.

Figure 4.4: Seasonality of measles transmission, reprinted from Hooker et al. (2011).

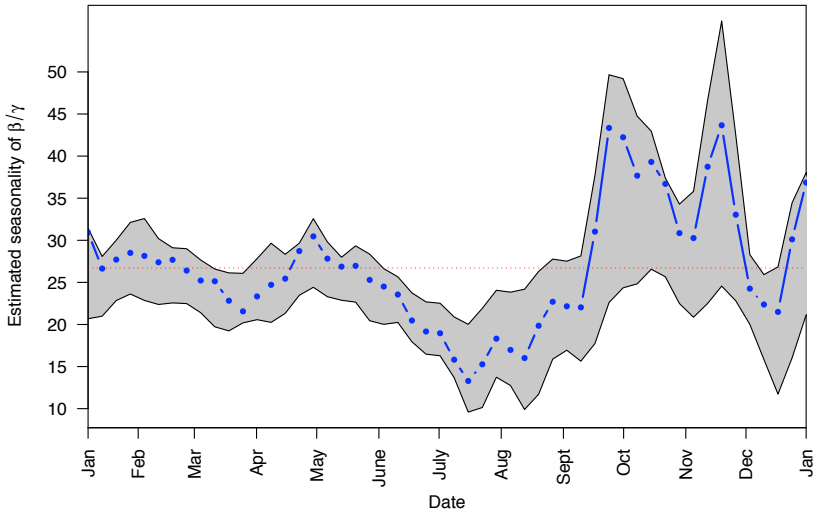


Figure 4.5: Estimated seasonality of β/γ for measles in Ontario, Canada (1939-1950). Blue dots show the annual seasonal pattern of β/γ estimated by our method while the grey region shows the upper and lower quartiles. The red dotted line shows the mean value of β/γ. Parameter values were chosen as follows: B_t - monthly birth data from the province of Ontario, Canada, $S_1 = 4.5 \cdot 10^5$, $1/\gamma = 5$ days, $\mu = 0\,\text{year}^{-1}$, $p_t = 0$, $\rho = 30\%$ (the above parameters are as in Hooker et al. (2011)), $T_{\text{report}} = T = 2$ weeks.

We also carried out a sensitivity analysis to determine the robustness of our method with respect to changes in the number of births (B_t), the initial number of susceptibles in the population (S_1), the death rate (μ_t), the reporting rate (ρ), the case fatality proportion (η) and the mean time from initial infection to its reporting (T_{report}). This analysis has shown that estimated seasonal variation of the transmission rate, β, are not significantly affected by changes in these parameters. However the *mean* of the estimated transmission rate strongly depends on the initial number of susceptibles in the population (S_1), the reporting rate (ρ) and the case fatality proportion (η). Since usually there is great uncertainty in these parameters, the estimated mean value of β should be treated with caution.

In the next section we apply the method described above to the weekly London smallpox mortality data and estimate the seasonality of the smallpox transmission rate for the period 1664-1930.

4.3 Seasonality of the smallpox transmission rate

The London Bills of Mortality represent a remarkably rich data set on epidemics of many infectious diseases. Our particular focus is smallpox. In this section we use smallpox mortality data to estimate the seasonal pattern of smallpox transmission and investigate how it changed across the centuries.

Data and parameter estimates

The London Bills of Mortality (1664-1930) provide weekly data on three processes: birth (total christened), death (all-cause mortality) and smallpox mortality (that we consider deaths due to smallpox). Given smallpox mortality data (M_t), we used **equation** (4.13) to estimate the seasonality of the smallpox transmission rate, β_t. The number of weekly births, B_t, was estimated as a trend of reported weekly births using Empirical Mode Decomposition (Wu et al. 2007). The death rate, μ_t,

was calculated as a trend of reported weekly death rate (Total deaths$/N$, where N is London population's size, as estimated in **Chapter 3**). Variolation and vaccination rates (p_t^{var} and p_t^{vacc} respectively) were estimated based on the historical analysis discussed in **Chapter 3** and are summarized in **Table 4.2**.

Year	p^{var}	p^{vacc}	$p = p^{\text{var}} + p^{\text{vacc}}$
1721	0.01	0	0.01
1728	0.03	0	0.03
1740	0.10	0	0.10
1768	0.15	0	0.15
1790	0.20	0	0.20
1797	0.25	0.05	0.30
1808	0.30	0.10	0.40
1835	0.13	0.40	0.53
1840	0.10	0.50	0.60
1841	0	0.60	0.60
1850	0	0.65	0.65
1870	0	0.75	0.75
1880	0	0.80	0.80
1930	0	0.80	0.80

Table 4.2: Variolation (p^{var}) and vaccination (p^{vacc}) uptake level for smallpox in London, England (1664-1930). The variolation and vaccination rates were estimated based on the historical analysis presented in **Chapter 3**. Linear interpolation method was used to obtain estimates for the rest of the years.

We estimated the initial number of susceptible individuals (S_1) as $0.3 \cdot N_{1664}$ (for 1664), which is plausible, assuming the system is close to equilibrium ($S_{\text{equilibrium}} = 1/\mathcal{R}_0$, where $3 \leq \mathcal{R}_0 \leq 10$ (Anderson and May 1991; Eichner and Dietz 2003; Gani and Leach 2001) implying that $0.1 \leq S_{\text{equilibrium}} \leq 0.3$). The low number of susceptibles during 1665–1675 could have resulted from the Great Plague of 1665-1666, if the plague killed a disproportionate number of young people.

Smallpox mortality records are likely to be accurate due to the fact that smallpox was an easily recognizable disease (Hopkins 1983). Therefore we assume that the death reporting rate, ρ, was close to 100%. All other parameter values and

their sources are described in **Table 4.3**.

Parameter	Description	Value	Source
$1/\sigma$	Mean latent period	15 days	**Chapter 3**
$1/\gamma$	Mean infectious period	7 days	**Chapter 3**
Δt	Sampling interval	1 week	
T	Mean time from initial infection to recovery	3 weeks	$T = 1/\sigma + 1/\gamma$
T_{report}	Mean time from initial infection until death due to smallpox	3 weeks	**Chapter 3**
η	Smallpox fatality rate	20%	(Fenner et al. 1988)
ρ	Reporting rate	100%	(Fenner et al. 1988)
S_1	Initial number of susceptible individuals in the population	$0.3 \cdot N_{1664}$	Estimated

Table 4.3: Parameters and their values.

Results

Figures 4.6–4.7 summarize the results of our analysis and show the estimated seasonal pattern of the smallpox transmission rate (blue solid line in subfigures (a)-(i)) together with 25% and 75% quartiles (the grey region) for the period 1664-1930. Unlike the measles transmission pattern in the 20th century (**Figure 4.2**), which clearly shows three distinct peaks associated with the school terms, the smallpox transmission rate does not have such an obvious and easily identifiable pattern. Seasonality is much harder to detect. To the eye the reconstructed transmission rate is very close to the mean (red dotted line) and its pattern can be argued to be generated by noise and not seasonality. **Figure 4.8** shows the estimated β/γ (grey lines) for each year corresponding to the time intervals in **Figures 4.6–4.7**. The pattern of the curves show considerable stochastic variations. To understand if estimated smallpox transmission rate is indeed seasonal we computed its wavelet spectrum. The wavelet spectrogram (**Figure 4.9**) shows a periodic component of one year, most

prominently for 1768–1820. The spectral peak at one year indicates that smallpox transmission varied seasonally. Clearly, more careful investigation of this question is required in the future. For our present purpose we take the median of $\beta(t)$ (blue lines in **Figures 4.6–4.7**) to be our best estimate of seasonal variations in smallpox transmission.

Figures 4.6–4.7 suggest that the seasonal pattern of the smallpox transmission rate changed over time and the following features can be observed.

- Throughout the period when only naturally acquired immunity for smallpox existed (**Figure 4.6**, panels (a)-(b)), the transmission rate exhibits a decline during August-September, which could have been caused by seasonal migration of Londoners to the countryside for harvesting (Clark 1979). After 1721 this decline can no longer be detected.
- During 1768-1840 the transmission rate exhibits a deep trough during the first two weeks of December (**Figure 4.7**, panels (e)-(g)). Such trough is not evident during other periods. Another trough is noted in February and then an annual rise in March.
- After 1840, the rise in transmission occurs in the beginning of September (**Figure 4.7**, panels (h)-(i)), unlike previous years. The obvious question is: could it be the effect of school opening in September? There was very limited schooling prior to the mid 19th century, when only wealthy families sent their children to schools. However, during the second half of the 19th century, the number of schools started to grow. Since by the 18th century smallpox have become primarily a childhood disease (Creighton 1965; Fenner et al. 1988; Razzell 1977), contacts among school children could have been a key driver of the smallpox transmission.
- During the period 1875-1895, the seasonal pattern of smallpox transmission shows three peaks (**Figure 4.7**, panel (i) and **Figure 4.10**) similarly to the seasonal pattern of measles in England and Wales, 1950-1965, which was driven

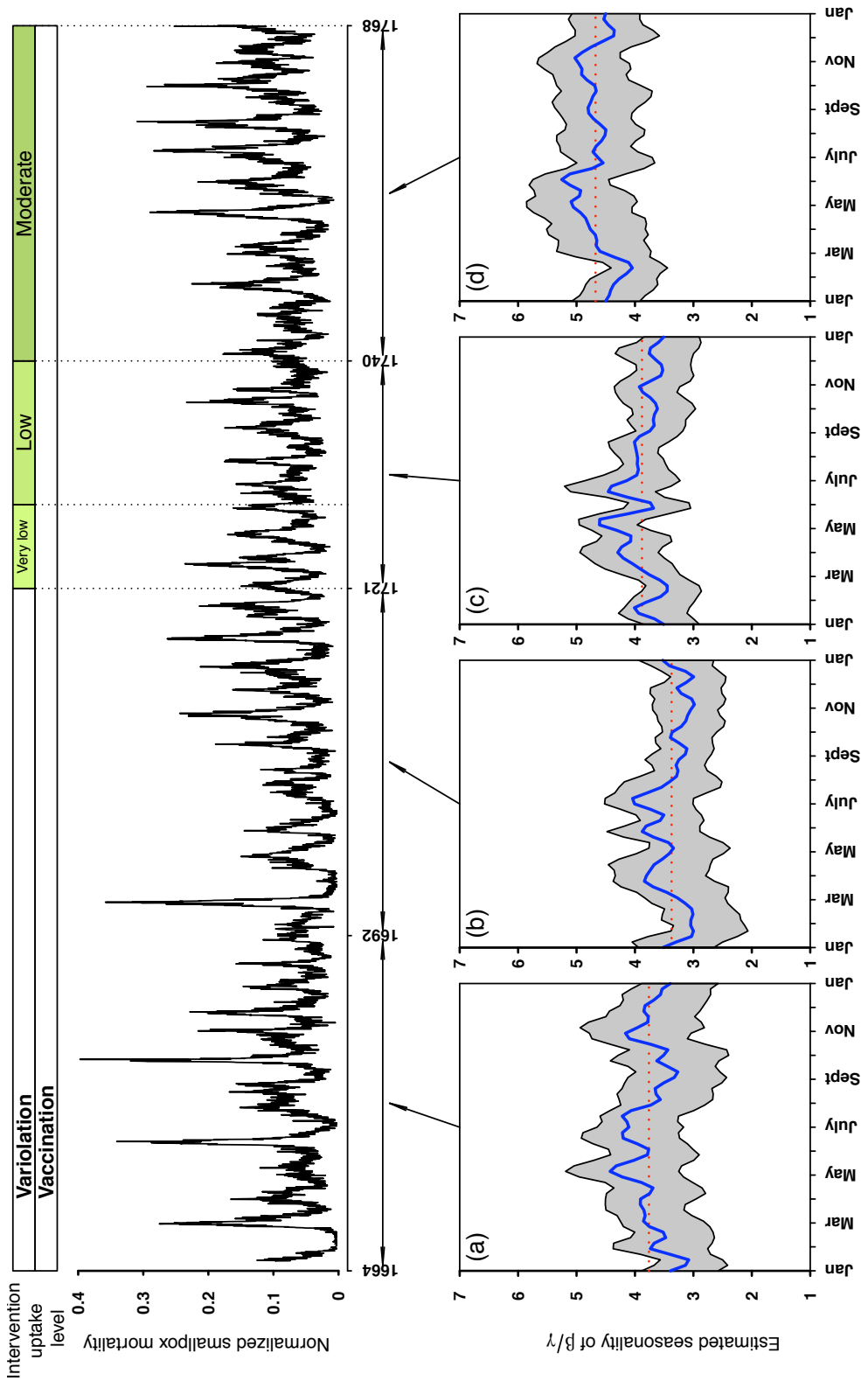


Figure 4.6: Estimated seasonal variation of smallpox transmission rate in historic London, England (1664–1768). **(continued on the next page)**

Figure 4.6: The **top panel** shows smallpox mortality data normalized by all-cause mortality (black curve) together with variolation and vaccination uptake levels (yellow green-dark olive colourbar for variolation with yellow green indicating the lowest level and dark olive the highest; and yellow-red colourbar for vaccination with light yellow indicating the lowest level and red the highest level). The **bottom panel** shows estimated seasonal variation of the transmission rate for specific time intervals. Solid blue curves represent the annual seasonal pattern of β/γ estimated by our method while the grey region shows the upper and lower quartiles. The red dotted line in each sub-panel shows the mean value of β/γ.

by school terms (**Figure 4.2**). The history of the school system in England shows that thousands of schools were opened after the Elementary Education Act of 1870, which introduced public schools in England (Middleton 2010). This could have had a tremendous effect on the contact pattern in children.

- The smallpox seasonal transmission pattern in panel (i) of **Figure 4.7** does not match the school-term dates as well as the seasonality of measles transmission (**Figure 4.3**). This can be explained by different starting dates of the school terms in 19th and 20th century England. While we did not find the exact days of the school terms in England in the 19th century to match with the seasonality of smallpox transmission, we did find that the schools were closed during the Christmas holidays (two weeks), Easter holidays (one week) and four or five weeks over the summer (Middleton 2010), which seems to coincide with the troughs in **Figure 4.10**.
- Intervention uptake levels did not seem to influence the seasonality of smallpox transmission. However, the mean of the transmission rate (the red dotted line in subfigure (a)-(i) **Figures 4.6–4.7**) declines drastically as variolation and vaccination uptake levels grow at the end of the 18th century (**Figure 4.7**). It appears that variolation on its own did not help to control smallpox transmission and mean β/γ was increasing. The introduction of vaccination seems to have played a major role in the reduction of the effective $\mathcal{R}_0$,

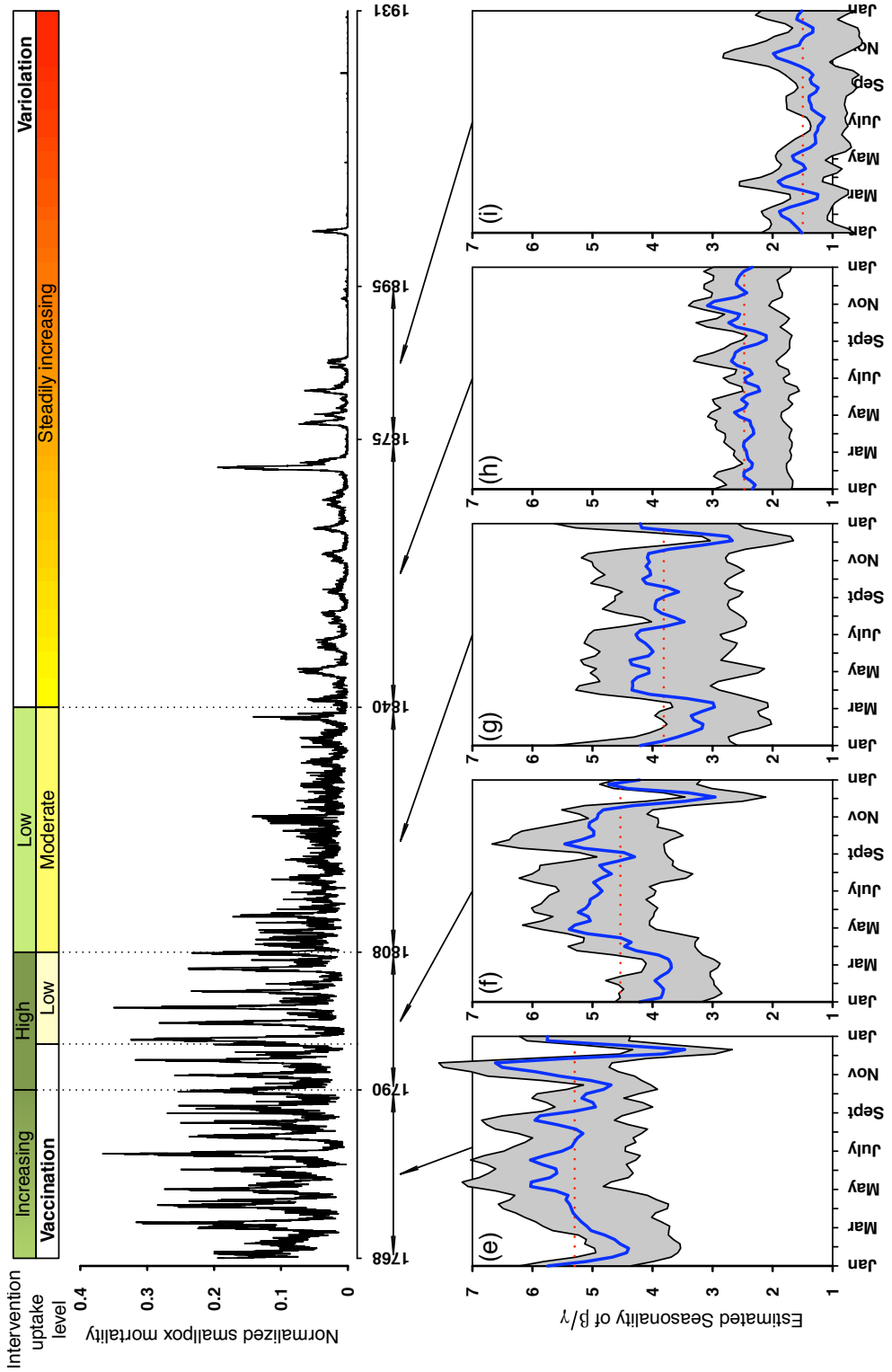


Figure 4.7: Estimated seasonal variation of smallpox transmission rate in historic London, England (1768-1930). Annotation is as in **Figure 4.6**.

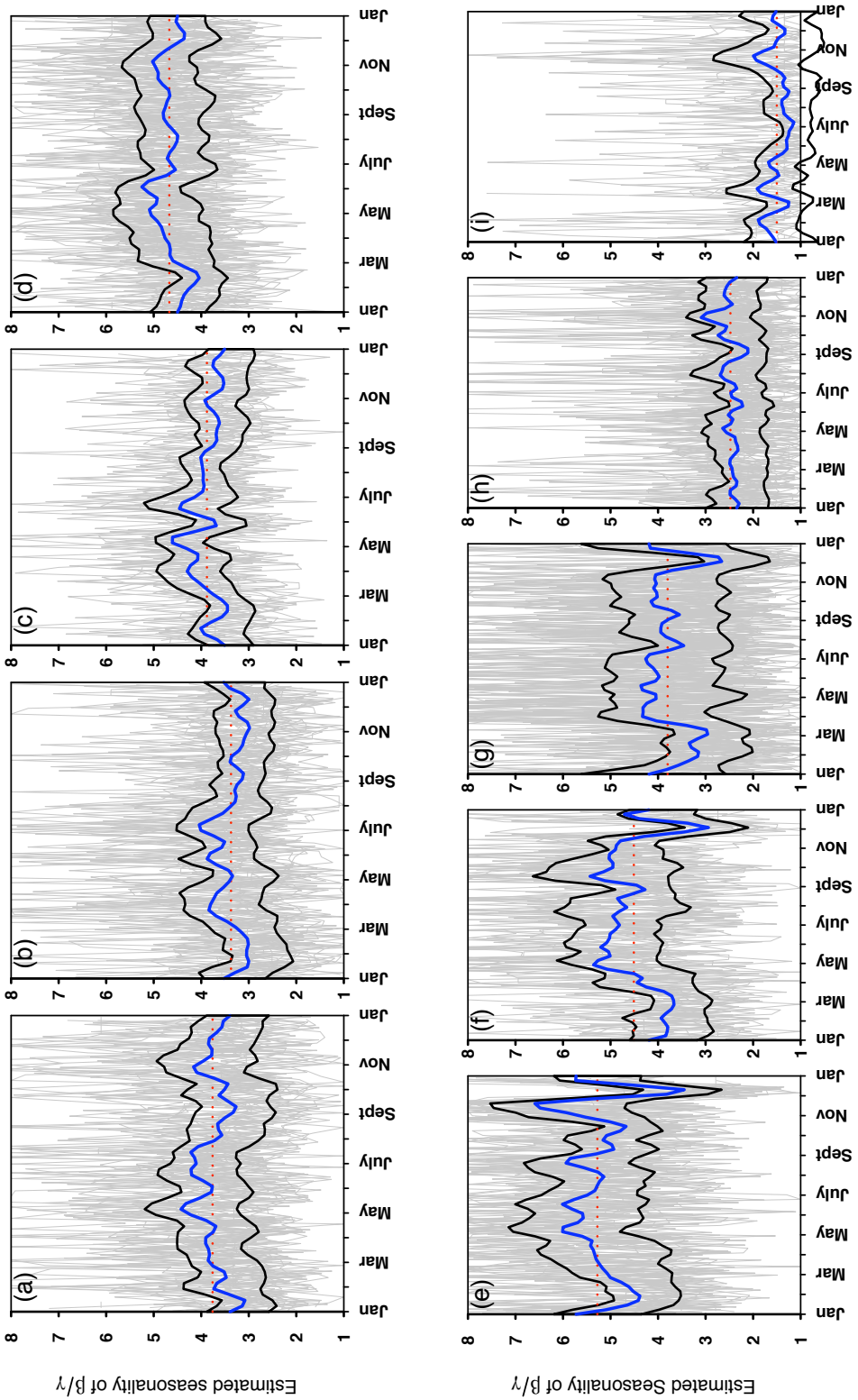


Figure 4.8: Estimated seasonal variation of smallpox transmission rate in historic London, England (1664-1930) Light grey curves are estimated β/γ for each year in the time interval according to **Figures 4.6–4.7**, so that subplot (a) shows estimated β/γ for each year from 1664 until 1692 together with their median (blue curve) and upper and lower quartiles (black curves). The red dotted line in each sub-panel shows the mean value of β/γ.

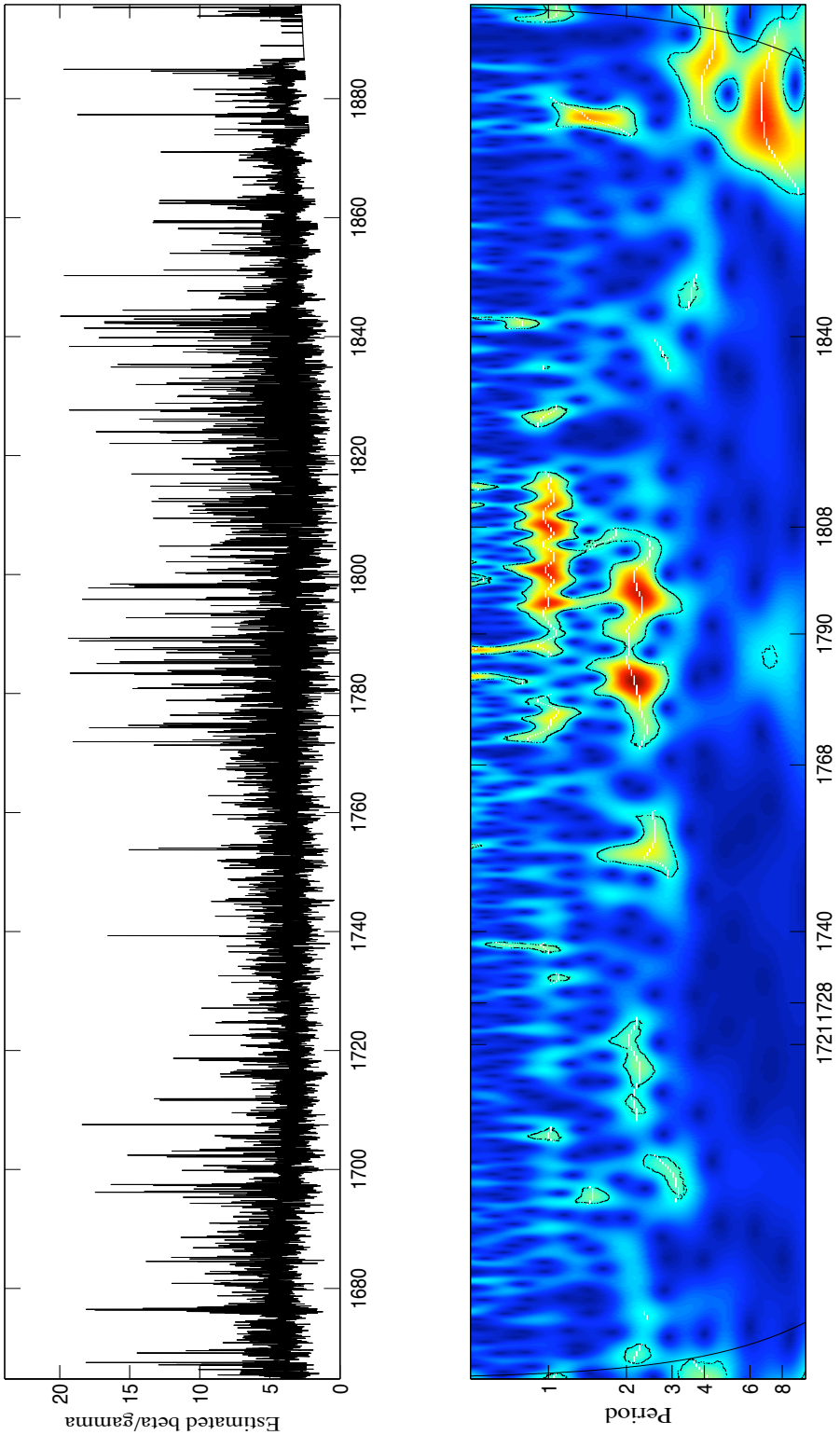


Figure 4.9: Wavelet transform of the reconstructed smallpox transmission rate, $\beta(t)$ (log-transformed and normalized to unit variance). Annotation is as in **Figure 3.10**.

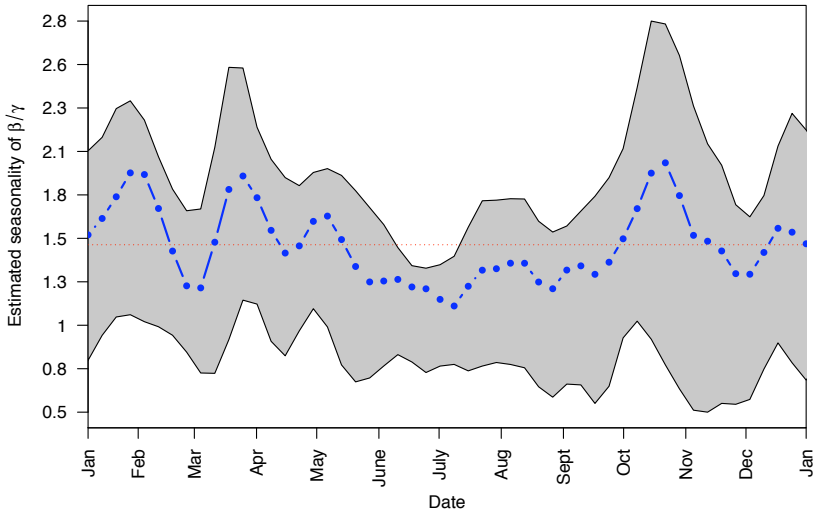


Figure 4.10: Estimated seasonality of β/γ for smallpox in London, England (1875-1895). The blue dots show the annual seasonal pattern of β/γ estimated by our method while the grey region shows the upper and lower quartiles. The red dotted line shows mean value of β/γ.

which resulted in a significant decrease of the severity of smallpox epidemics and their final elimination from London near the beginning of the 20th century (**Figure 4.7**).

4.4 Discussion

The main goal of our study was to estimate and analyze changes in the seasonality of smallpox transmission in London, England. We have presented a simple method based on the continuous time SEIR model, which allows one to estimate the disease transmission rate from large case notification or mortality data sets. Compared with state-of-the-art methods (He et al. 2010; Hooker et al. 2011; Ionides et al. 2006; King et al. 2008) it provides an algorithm that is easy to implement and is not computationally demanding. Although the method is naive and lacks a solid statistical foundation we found that it performed very well on simulated data and real data previously analyzed by other methods. Applying our method to the smallpox mortality records from the Weekly London Bills of Mortality, we were able to estimate

the seasonal pattern of smallpox transmission rate over the period 1664-1930.

We have shown that seasonal variations of the underlying transmission parameter change significantly over the centuries. The seasonal pattern of the transmission rate seems to be related to seasonality of population behaviour, such as migration of Londoners to the countryside during harvesting season, and (after the introduction of public schools in 1870) variation in contacts among children generated by school terms.

Variolation and vaccination uptake levels do not seem to influence seasonality of the transmission rate. These finding echo conclusions drawn previously for measles: seasonality of measles transmission had the same pattern for both pre- and post-imminization periods (Fine and Clarkson 1982; Finkenstadt and Grenfell 2000). However, it appears that vaccination played a major role in reducing the mean value of the smallpox transmission rate in the 19th century. This eventually reduced smallpox to the level where there were no longer recurrent epidemics and only occasional outbreaks seeded by immigration.

Seasonal variations in temperature, which significantly changed over these centuries, might also have had an impact on smallpox transmission. Future research may reveal if these environmental changes had any effect on smallpox seasonality.

Chapter 5

Modelling transitions in smallpox dynamics in London, England, 1664-1930

Abstract

Many infectious disease time series display changes in epidemiological pattern, such as transitions from an annual cycle to a biennial cycle, or to irregular dynamics. Such dynamical transitions have previously been associated with changes in the rate at which new susceptible individuals enter a population. Previous transition analysis has been restricted to epidemics of the 20th century over 50-70 year time intervals (Bauch and Earn 2003b). We investigate whether demographic and behavioural changes influenced transitions of smallpox dynamics in London, England over almost three centuries. We show that the periodic structure of the smallpox epidemics is associated with the period of damped oscillations onto the annual attractor of the SIR model for smallpox. Further, our analysis reveals that the changes in smallpox dynamics can be related to changes in birth rate, immigration to London, and in uptake of various control measures.

5.1 Introduction

The tragic events of September 11, 2001 generated great interest in modelling the potential release of smallpox virus and possible scenarios for public health responses to it (Aldis and Roberts 2005; Bauch et al. 2003; Burke et al. 2006; Eichner 2003; Ferguson et al. 2003; Glasser et al. 2008; House et al. 2010; Kaplan et al. 2002, 2003; Longini et al. 2007; Meltzer et al. 2001; Nishiura and Tang 2004; Riley and Ferguson 2006; Zenihana and Ishikawa 2010). Nevertheless, relatively little attention has been given to the analysis of historical epidemics of smallpox (Duncan and Gyongy 2006; Gani and Leach 2001; Nishiura 2007b; Nishiura and Kashiwagi 2009). One of the major reasons for the limited study of historical outbreaks is the lack of available data. Recently digitized weekly Bills of Mortality have allowed us to study smallpox epidemics in London, England over almost three centuries. This data set is of particular interest since it covers almost 100 epidemics triggered by *variola* virus from October 1664 until December 1930.

Statistical analysis of the weekly London Bills of Mortality, presented in **Chapter 3**, revealed significant variations in the amplitude and temporal frequency of smallpox epidemics over different periods of time. Analysis of the smallpox history in England, also described **Chapter 3**, led us to believe that these changes in the epidemiological pattern of smallpox epidemics were correlated with significant historical events related to the implementation of the control measures (i.e. variolation and vaccination), population movement, wars, *etc*. In this paper we confirm our observations using mechanistic mathematical modelling.

Previous theoretical work (Bauch and Earn 2003a,b; Earn 2009; Earn et al. 2000a; He and Earn 2007) established that transitions from oscillations of one periodic cycle to another, often observed in epidemiological time series, are determined not by chance but by the rate of recruitment of susceptible individuals into the population. Birth rate and vaccination levels are the two major factors associated with the input of new susceptibles into the population and, therefore, can be seen as

the main drivers of the changes in disease dynamics. In the current work we use the method of transition analysis explained in detail in **Chapter 2**, to investigate if indeed the observed demographic and behavioural changes triggered changes in smallpox dynamics. We determine the rate of input of new susceptible individuals into the population from birth and immigration and correlate changes in this rate with the changes in the temporal pattern of smallpox mortality. We also show that transient dynamics, and specifically damped oscillations onto the annual attractor, are the key elements determining the periodic structure of the London smallpox time series.

In the next section we describe the model we used to conduct transition analysis and its parameterization. In **Section 5.3** we apply transition analysis to smallpox dynamics in London, England (1664-1930). We summarize our results and highlight future research directions in **Section 5.4**.

5.2 The model

A variety of complicated multi-stage deterministic (Kaplan et al. 2002), stochastic (Bozzette et al. 2003b; Eichner 2003; Halloran et al. 2002; Longini et al. 2007; Meltzer et al. 2001), individual-based computational (Burke et al. 2006), integral-equation (Aldis and Roberts 2005), age-structured and socially/spatially structured (Glasser et al. 2008) models have been developed recently to study smallpox transmission. These models have been used to evaluate different scenarios for public health response in the event of a smallpox bioterrorism attack. Researchers modeled realistic features of the current social and spatial structure of the population as well as various scenarios for smallpox release, which resulted in extreme model complexity. Much simpler SEIR (*susceptible, exposed, infectious, recovered*)-type models (Anderson and May 1991) have been used to study historic smallpox mortality (Duncan et al. 1996; Duncan and Gyongy 2006; Gani and Leach 2001).

The standard SEIR model, which is based on the assumption of a large and homogeneously mixed population, appears to adequately approximate smallpox transmission dynamics and will be used to model smallpox dynamics in historic London. To make the model more realistic we consider a realistic distribution of the latent and infectious periods, which can be approximated by the Erlang distribution with the shape parameter $m = 40$ for the latent period and the shape parameter $n = 4$ for the infectious period (Wearing et al. 2005). This more realistic model would then include 45 disease stages and therefore 45 equations. Our work on the Erlang distributed models presented in Chapter 1 has shown that the dynamics of this complex system can be successfully captured by the classic SIR model (Anderson and May 1991; Kermack and McKendrick 1927) if the mean *serial interval*[1], T_{serial}, which is the time from initial infection of a primary case to initial infection of a secondary case (Fine 2003)), is the same for both models. Therefore in our analysis we use the SIR model (**equation** (5.1)) that the includes mean serial interval:

$$\frac{dS}{dt} = \nu(t)N_0 - \beta(t)S(t)I(t) - \mu S(t)\,, \tag{5.1a}$$

$$\frac{dI}{dt} = \beta(t)S(t)I(t) - \frac{1}{T_{\text{serial}}}I(t) - \mu I(t)\,, \tag{5.1b}$$

$$\frac{dR}{dt} = \frac{1}{T_{\text{serial}}}I(t) - \mu R(t)\,, \tag{5.1c}$$

where, S, I and R are the numbers of susceptible, infectious and recovered (immune) individuals in the population respectively. The parameters ν, μ and β are the rates of birth, *per capita* death and transmission respectively. T_{serial} is computed by the formula (Svensson 2007):

$$T_{\text{serial}} = T_{\text{latent}} + \left(\frac{n+1}{2n}\right) T_{\text{inf}}, \tag{5.2}$$

[1]The serial interval is also called the generation interval, the generation time, or the case-to-case interval.

where T_{latent} is the mean latent period, T_{inf} is the mean infectious period, and n is the shape parameter of the infectious period distribution.

The natural history of smallpox infection was described in **Chapter 3**. Smallpox has a mean latent period T_{latent} of 15 days and a mean infectious period (with multiple stages) of about 20 days. An infected person is extremely infectious during the first four days after the rash appearance, after which the infectiousness wains rapidly. We assume that the time during which the infectious individuals can transmit infection, T_{inf}, is about 7 days on average (Fenner et al. 1988). Substituting these estimates in **equation** (5.2) implies that $T_{\text{serial}} \approx 19$ days.

Statistical analysis preformed in **Chapter 3** showed the presence of seasonality in the observed smallpox data, which suggested that the transmission rate β was seasonally forced. In **Chapter 4**, we estimated seasonal variation in smallpox transmission for various time intervals. We will use those estimates in our model. All other parameter values used in the model are listed in **Table 5.1**.

Parameter	Description	Value	Source
T_{latent}	Mean latent period	15 days	Fenner *et. al.* (1988)
m	Shape parameter of the Erlang distributed latent period	40	Wearing *et. al.* (2005)
T_{inf}	Mean infectious period	7 days	Fenner *et. al.* (1988)
n	Shape parameter of the Erlang distributed infectious period	4	Wearing *et. al.* (2005)
T_{serial}	Mean serial interval	$\approx$ 19 days	Estimated (equation 5.2)
β	Seasonal transmission coefficient	Fig. 4.6-4.7	Estimated in Chapter 3
ν	Susceptible recruitment rate (for 1671–1684)	0.05	Estimated from our data
μ	Death rate	0.045	Estimated from our data

Table 5.1: Parameter notations and estimates.

5.3 Transition analysis

To identify transitions that occurred in smallpox dynamics we start with the description of the data and the analysis of its frequency structure. We estimate the suscep-

tible recruitment rate and infer the effective reproduction number, $\mathcal{R}_{0,\text{eff}}$ based on this estimate. Then we perform asymptotic and perturbation analysis of the SIR model with the mean serial interval estimated for smallpox. Finally, using the estimated value of $\mathcal{R}_{0,\text{eff}}$ and the results of the asymptotic and perturbation analysis we predict the changes in qualitative dynamical behaviour of smallpox incidence. Comparing our predictions based on the SIR model with the wavelet power spectrum of the smallpox mortality time series we try to establish whether demographic and behavioural changes triggered the transitions in smallpox dynamics observed in the data.

5.3.1 Description of the data

Reported mortality and inferred frequency structure

The time series of the weekly smallpox mortality in London (1664–1930) is shown in the upper panel of **Figure 5.1**. The data were normalized by the trend of the weekly all-cause mortality to account for the changes in the population size, city boundaries and data collection methods. **Figure 5.1** clearly shows the presence of different frequency components in the series. The smallpox pattern changes from frequent (2-3 year) periodic oscillations of large magnitude in the 18th century to periodic cycles of much longer (4-8 year) periods with dramatically reduced epidemic peaks right after the mid 19th century, except for the three large epidemics in 1871, 1876 and 1902.

The frequency structure of the analyzed time series is presented as a wavelet spectrogram in the lower panel of **Figure 5.1**. It shows how the periodicity of smallpox epidemics changed over the centuries (the white curves highlight the period with greatest power at each timepoint). From 1664 until 1700 the dominant period is 3–4 years. For most of the 18th century (1705–1808) a 2–3 year period dominates. However it shifts from being nearly 2-years in 1705–1728 to about 3 years in

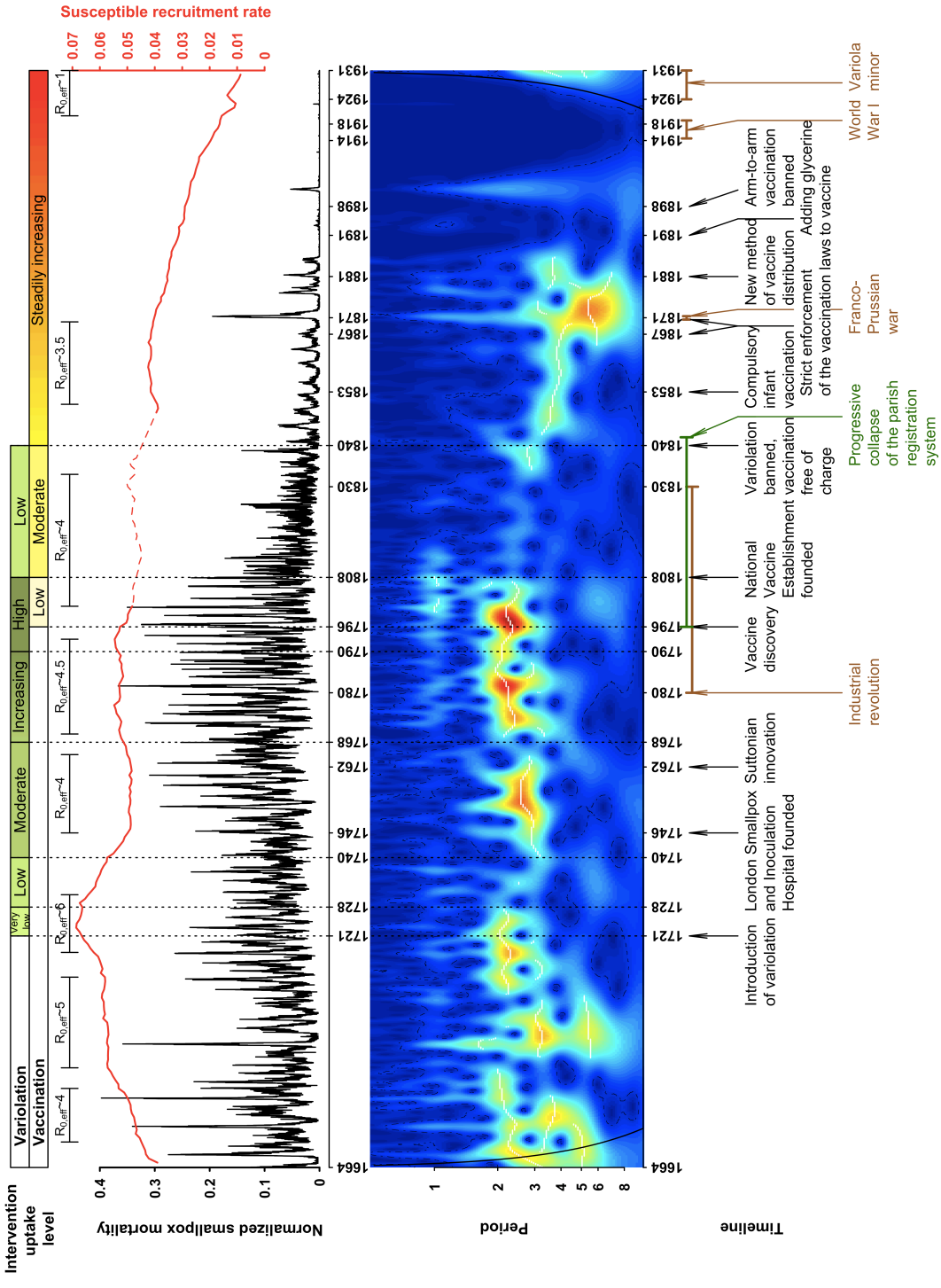


Figure 5.1: Smallpox mortality in London, England (1664-1930). **(continued on the next page)**

Figure 5.1: The **upper panel** shows weekly smallpox mortality normalized by the trend of weekly all-cause mortality (black). Annual susceptible recruitment relative to the population size in 1671-1684 is shown as a red curve. The period when data accuracy is reduced is shown in dotted line. We shift the recruitment curve forward by 5 years to account for the delay between birth and entering the well-mixed population. The line segments at the top of the upper panel highlight time intervals with distinct effective R_0, calculated with **equation** (5.10). Intervention uptake levels are shown as colour bars: yellow green-dark olive for variolation with yellow green indicating the lowest level and dark olive - the highest; and yellow-red for vaccination with light yellow indicating the lowest level and red - the highest level. The **lower panel** shows the wavelet power spectrum of the weekly smallpox mortality (square root-transformed and normalized to unit variance) for London, England (1664-1930) and its correlation with the historical timeline. The white curves show the local maxima of wavelet power (squared modulus of wavelet coefficients (Cazelles et al. 2008, p.291) at each time. The colours of the wavelet diagram vary from dark blue, which corresponds to low power, to dark red for high power. The dot-dashed curves indicate the 95% confidence region, estimated from 1000 bootstrapped time series generated by the method of (Cazelles et al. 2008, p.292-293). Below the cone of influence (Cazelles et al. 2008; Torrence and Compo 1998), the calculation of wavelet power is less accurate because it includes edges of the time series that have been zero-padded to make the length of the series a power of 2. The wavelet spectrum was computed using a MATLAB code kindly provided by Bernard Cazelles. In the timeline panel the text in black colour indicates events that influenced uptake of control measures; brown - events that influenced people behaviour; dark green shows the period when data accuracy is reduced.

1740–1762 and back to almost 2 years in 1768–1808. Between 1808 and 1842 the wavelet spectrogram shows no prominent period, suggesting irregular dynamics or simply low signal to noise ratio in the data. This time interval coincides with the progressive collapse of the parish registration system, triggered by the rapid growth of London's population and lack of expansion in Anglican parishes. Therefore the data during this period may not be very accurate. The introduction of the new registration system in 1837 significantly improved the accuracy of the data and we assume that the data obtained from this source (after 1842) is reliable. From 1842 onward the dominant period smoothly evolves to a longer cycle of 3–4 years and later to 4–8 years. Traces of an annual cycle are evident along the whole time series

with the most prominent spectral peak from 1768 to 1805.

Estimated susceptible recruitment

The susceptible recruitment rate is the rate at which susceptible individuals are entering the population. We assume that recruitment of individuals susceptible to smallpox in London was associated with two major demographic processes, birth and immigration, and also with the smallpox control level (i.e. uptake of variolation and vaccination). We can estimate the annual rate of susceptible recruitment, S_{in}, by the following formula:

$$S_{\text{in}}(t) = B(t - \tau_S) \cdot \left[1 - p(t)\right] - \text{Imm}(t) \cdot \left[1 - \kappa(t)\right], \tag{5.3}$$

where $B(t)$ is the annual number of births, shifted by $\tau_S \approx 5$ years to account for the delay between the time of birth and the time of entering a well-mixed susceptible pool; $p(t)$ is the proportion of people immunized by vaccination or variolation before entering the susceptible pool; $\text{Imm}(t)$ is the the annual immigration rate and $\kappa(t)$ is the level of immunity in immigrants.

Based on the data available to us and analysis of the literature on the population of historical London we estimate the birth rate, variolation and vaccination uptake levels and immigration below.

Births

The London mortality records are an excellent source of information on the weekly number of births in London. However, before the introduction of the national registration system in 1842, the London Bills of Mortality recorded only the weekly number of infants baptized ($\text{Bap}(t)$) in Anglican churches 8–30 days after birth (Wrigley and Schofield 1981), and not the total number of infants born ($B(t)$) during that week. Therefore we must account for the number of infants who died between birth and baptism and infants who were not baptized due to

belonging to a different religion (Landers 1993; Wrigley and Schofield 1981). Landers (1993)[p.162–207] conducted a comprehensive analysis of the London Bills of Mortality and estimated the birth under-registration rate, which took into account the aforementioned factors. We use his estimates of the correction factor, ρ_{Bap}, shown in **Table 5.2**, to estimate the number of births from the weekly baptisms published in the bills $\big(B(t) = \rho_{\text{Bap}}(t)\text{Bap}(t)\big)$. We assume that during the period 1842–1930, after the introduction of the new registration system, reported births accurately reflect true births.

Year	Correction factor	
	birth, ρ_{Bap}	death, ρ_{ACM}
1680s	1.0276	1.0110
1690s	1.0321	1.0128
1700s	1.0350	1.0140
1710s	1.0382	1.0153
1720s	1.0397	1.0159
1730s	1.0409	1.0164
1740s	1.0459	1.0183
1750s	1.0511	1.0204
1760s	1.0567	1.0229
1770s	1.0620	1.0248
1780s	1.0980	1.0490
1790s	1.1266	1.1081
1800s	1.2172	1.2078
1810s	1.2870	1.3109
1820s	1.2262	1.1756

Table 5.2: Correction factors to account for birth and death under-registration, reproduced from Landers (1993) Table 5.3, p.166.

Intervention uptake level

In historic London the principal control measures related to smallpox prevention were variolation and vaccination. Variolation is considered to be the predecessor of vaccination and was in practice from 1721 until 1840. Vaccination was introduced in London in 1796. Data on the number of variolated or vaccinated

individuals are very limited and most sources provide only rough descriptive information (i.e. high or low level). Based on analysis of the literature on the population of historical London in **Chapter 3**, we have created a timeline of the significant historical events related to the implementation of various control measures (see lower panel of **Figure 5.1**). We estimate the variolation and vaccination proportions (p^{var} and p^{vacc} respectively) according to that timeline. Our estimates are presented in **Table 5.3**.

Year	p^{var}	p^{vacc}	$p = p^{\text{var}} + p^{\text{vacc}}$
1721	0.01	0	0.01
1728	0.03	0	0.03
1740	0.10	0	0.10
1768	0.15	0	0.15
1790	0.20	0	0.20
1797	0.25	0.05	0.30
1808	0.30	0.10	0.40
1835	0.13	0.40	0.53
1840	0.10	0.50	0.60
1841	0	0.60	0.60
1850	0	0.65	0.65
1870	0	0.75	0.75
1880	0	0.80	0.80
1930	0	0.80	0.80

Table 5.3: Variolation (p^{var}) and vaccination (p^{vacc}) uptake level for smallpox in London, England (1664-1930). The variolation and vaccination rates were estimated based on the historical analysis presented in Chapter 2. Linear interpolation was used to obtain estimates for the rest of the years.

Immigration

Births were not the only source of new susceptible individuals in London. Immigration to London also contributed significantly to the input of new susceptibles. Thousands of people were immigrating to London each year. Many of these immigrants were from rural towns, where smallpox outbreaks were relatively rare (happening only every 5–10 years (Duncan et al. (1994a); Fenner et al. (1988);

McNeill (1998))). Therefore young adults migrating from rural areas were at great risk of acquiring smallpox. Unfortunately there are no reliable data for tracking people migrating to and from London. We can only obtain a very crude estimate of the amount of immigration based on annual population growth:

$$N(t+1) - N(t) = B(t) - D(t) + \big(\text{Imm}(t) + \text{Emm}(t)\big)\,, \tag{5.4}$$

where $B(t)$ and $D(t)$ are the annual births and deaths already adjusted for under-reporting $\big(B(t) = \rho_{\text{Bap}}(t)\text{Bap}(t)$ and $D(t) = \rho_{\text{ACM}}(t)\text{ACM}(t)\big)$; $\text{Imm}(t)$ and $\text{Emm}(t)$ are the number of people immigrating to and emmigrating from London respectively. We assume $\text{Emm}(t)$ is much smaller than $\text{Imm}(t)$ $\big(\text{Emm}(t) \ll \text{Imm}(t)\big)$. $N(t)$ is London's population size estimated in **Chapter 3**. For simplicity we assume that $\text{Imm}(t)$ is roughly proportional to births,

$$\text{Imm(t)} \approx \eta(t)B(t), \tag{5.5}$$

where $\eta(t) \geq 0$. Using **equation** (5.5) and assuming that the proportion of immigrants who were immune $\kappa(t) \ll 1$ in **equation** (5.3) (i.e. the majority of immigrants to London were susceptible), **equation** (5.3) and **equation** (5.4) can be rewritten as:

$$S_{\text{in}}(t) = B(t - \tau_S) \cdot \big[1 - p(t)\big] + \eta(t)B(t)\,, \tag{5.6}$$

and

$$N(t+1) - N(t) = B(t) - D(t) + \eta(t)B(t)\,. \tag{5.7}$$

We estimate the trend of $\eta(t)$ from **equation** (5.7), and use in **equation** (5.6) to calculate the annual number of susceptible individuals entering the population.

Susceptible recruitment rate

The susceptible recruitment rate, $\nu(t)$ is calculated relative to the population

size at an "anchor time", t_0, as:

$$\nu(t) = \frac{S_{\text{in}}(t)}{N_0}. \tag{5.8}$$

The "anchor time" is the time for which we can obtain a reliable estimate of the basic reproduction number $\mathcal{R}_0$. An independent estimate of $\mathcal{R}_0$ (e.g. based on age-incidence or age-seroprevalence data) is always desirable (Bauch and Earn 2003a,b; Earn et al. 2000a). Unfortunately such data for London during 1664-1930 were not available to us. For our "anchor time" we choose the period 1671–1684, which is the first time interval where the birth rate was relatively constant and therefore we would assume that $\mathcal{R}_0$ would also be approximately constant. The value of $\mathcal{R}_0$ for smallpox, as previously estimated in the literature, varies between 3 and 10 (Ferguson et al. 2003; Gani and Leach 2001) with the most common estimate being within the 3–6 range (Anderson and May 1991). We assume $\mathcal{R}_0$ to be 4 at the "anchor time" 1671-1684. In the **Section 5.3.2** we explore the range $3 \leq \mathcal{R}_0 \leq 6$ and examine the sensitivity of our results to the choice of $\mathcal{R}_0$.

Equation (5.8) implicitly assumes that new susceptibles are dispersed uniformly within a region of a constant area. In London, the transition from one registration system to another resulted in inclusion of new areas in the reports. The data recorded in the parish registers (1664-1841) covered 15,000 acres (calculated from the geographic information system (GIS) shapefiles of the historic parishes of London), while the Registrar General (1842-1930) collected data from an area of 78,000 acres (as reported in the Registrar General's annual returns). If the population density were identical in the area covered by parish registers and the Registrar General we could simply adjust **equation** (5.8) by a factor $\frac{A(t)}{A(t_0)}$. However population density varied considerably, with the highest density in the original core of the city covered by the old registration system and a lower density in the areas added in the new registration system. Therefore we also need to account for non-uniform

density, which we do by introducing a factor ζ and rewriting **equation** (5.8) as

$$\nu(t) = \frac{S_{\text{in}}(t)}{N_0 \left(\zeta(t)\frac{A(t)}{A(t_0)}\right)}. \tag{5.9}$$

Since the area does not change for 1664–1841, $\zeta(t)$ for this period is equal to 1. For 1842-1931 we estimate $\zeta(t)$ based on the assumptions that it is constant, the $\mathcal{R}_0$ is at least 1 in 1931 when our time series ends, and the fact that any change in $\mathcal{R}_{0,\text{eff}}$ corresponds to a change in the susceptible recruitment rate (Bauch and Earn 2003a,b; Earn et al. 2000a):

$$\mathcal{R}_{0,\text{eff}} = \mathcal{R}_0 \frac{\nu(t)}{\nu(t_0)}\,. \tag{5.10}$$

Assuming $\mathcal{R}_{0,\text{eff}}(1931) \geq 1$ we can obtain the value of $\nu(1931)$ from **equation** (5.10):

$$\nu(1931) = \frac{\mathcal{R}_{0,\text{eff}}}{\mathcal{R}_0}\nu(t_0) \geq 0.0125. \tag{5.11}$$

We plugged this value into **equation** (5.9) and estimated $\zeta(1931) \leq \frac{2}{5}$. We make an assumption that $\zeta(1931) \approx \frac{2}{5}$ for 1842–1931.

The resulting annual susceptible recruitment rate (**equation** (5.9)) is shown in the top panel of **Figure 5.1** (red line).

Effective basic reproduction number, $\mathcal{R}_{0,\text{eff}}$

Based on our estimates of the susceptible recruitment rate we identify time intervals during which $\nu(t)$ is relatively constant, hence during which the dynamical features of the disease time series can be expected to be approximately stationary. The effective basic reproduction number, $\mathcal{R}_{0,\text{eff}}$, is then estimated for each of these intervals using **equation** (5.10). Periods when the susceptible recruitment rate was roughly constant are plotted in the top panel of **Figure 5.1** together with the corresponding $\mathcal{R}_{0,\text{eff}}$. It indicates that $\mathcal{R}_{0,\text{eff}}$ fluctuated between 4–6 from 1664 until the

1870s, after which it smoothly decreased.

We estimated the susceptible recruitment rate and the corresponding $\mathcal{R}_{0,\text{eff}}$ to the best of our ability. However, the accuracy of our estimates is compromised by uncertainty in variolation and vaccination levels and our crude estimates of the effects of immigration and population size. Asymptotic and perturbation analysis will reveal if we can correctly predict changes in the periodic dynamics of smallpox epidemics based on these rough estimates.

5.3.2 Asymptotic and perturbation analysis

The results of our asymptotic and perturbation analyses based on the SIR model with mean serial interval 19 days are presented in **Figures 5.2–5.3**. The top panel of each figure shows the wavelet spectrogram together with $\mathcal{R}_{0,\text{eff}}$ estimated based on the susceptible recruitment curve (**Figure 5.1**). The asymptotic and transient dynamics of the SIR model are shown in the bottom panel of each figure. Note that the seasonality of the transmission rate, estimated in **Chapter 4**, varied across the centuries. Since we included these estimates in our model, we have 9 diagrams that correspond to 9 time regions, where the seasonal pattern of smallpox transmission stayed roughly the same. We also matched the dynamics of the SIR model with the estimated seasonality to the sinusoidally forced SIR model by choosing an appropriate sinusoidal forcing amplitude α. The estimated amplitudes are shown in each corresponding sub-panel of **Figures 5.2–5.3**.

Our asymptotic analysis is presented as bifurcation diagrams in subplots (a)-(i) of **Figures 5.2–5.3**), which shows that for the biologically plausible values of $\mathcal{R}_0$ for smallpox ($3 \leq \mathcal{R}_0 \leq 6$) there is only one periodic attractor, an annual cycle. Therefore, asymptotic analysis correctly predicts the existence of the resonant period-one cycle observed in the wavelet spectrum. The period of damped oscillations onto the annual attractor (transient period) is used to explain the appearance of non-resonant periods in the wavelet spectrogram. The dashed red lines in each

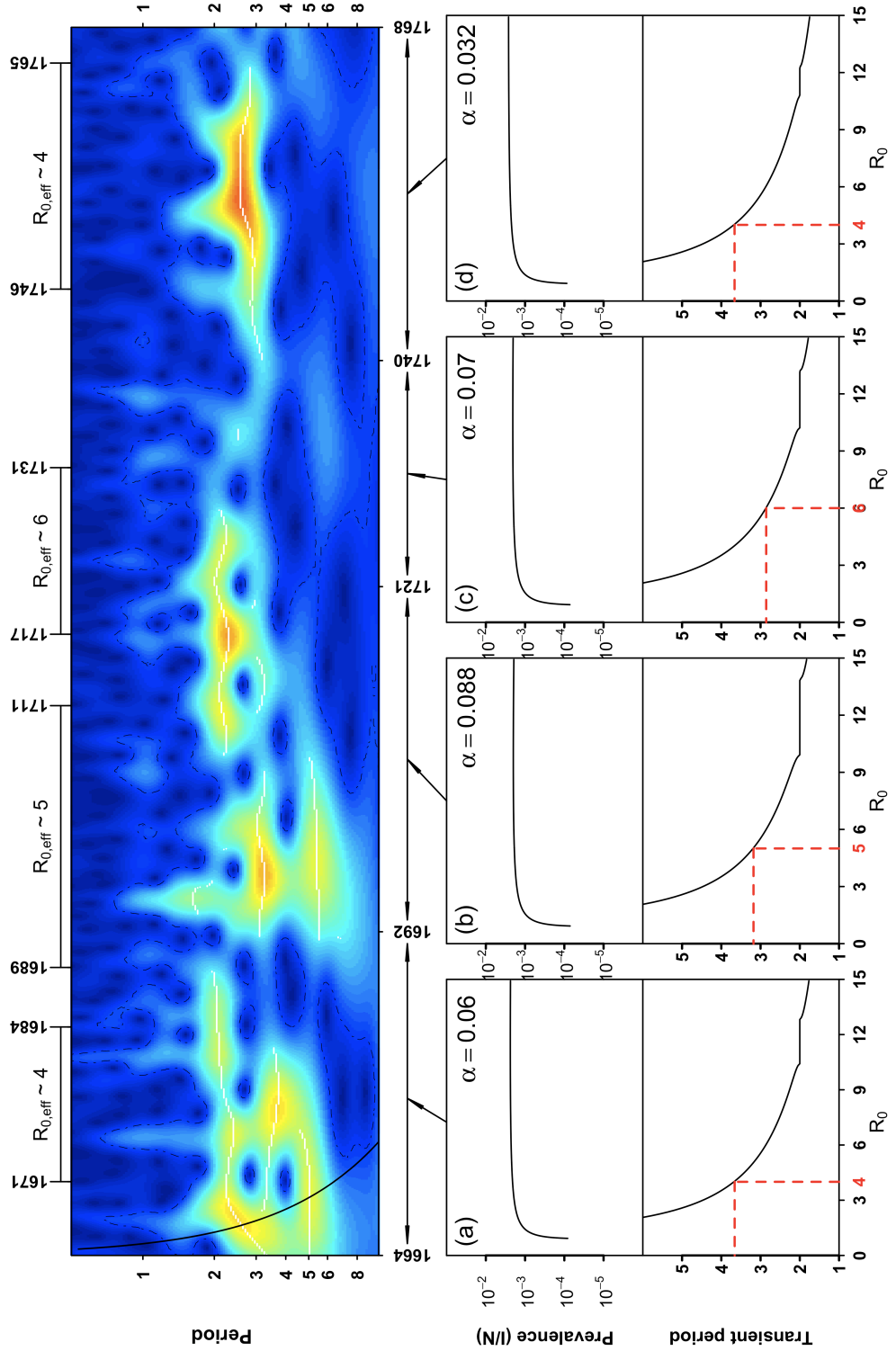


Figure 5.2: Transition analysis of the smallpox mortality time series for London, England (1664-1768). **(continued on the next page)**

Figure 5.2: The **top panel** shows wavelet power spectrum as described in **Figure 5.1** with corresponding values of $\mathcal{R}_{0,\text{eff}}$. Seasonal amplitude α of the corresponding sinusoidal forcing is stated in right top conner of each subplot. The **lower panel** shows a sequence of "transition diagrams" produced for every time interval when seasonality of the transmission rate β, determined in Chapter 3, was roughly the same. Each subplot shows the changes in the asymptotic dynamics (bifurcation diagram with stable cycle in black and unstable in grey) and transient dynamics (period of damped oscillations onto the annual attractor) as a function of $\mathcal{R}_0$. The red dashed line indicates predicted non-resonant period.

of the subplots of **Figures 5.2–5.3** show the period predicted by the model. Comparing these predicted non-resonant periods with the dominant mode of the wavelet power spectrum (the white line in the wavelet spectrogram) we observe that the transitions from one periodic cycle to another coincide with the changes in $\mathcal{R}_{0,\text{eff}}$. Thus the trend of the changes in the peak values of the wavelet power spectrum are captured by our model.

The predictions of the SIR model appear to be very accurate for 1664–1710. However for 1710–1840 the model predicts 3-4 year non-resonant periods while we observe 2-3 year periods in the wavelet spectrogram. This discrepancy may be related to the low accuracy of our estimated susceptible recruitment rate. It appears that for this period (1710–1840) we underestimated the susceptible recruitment and therefore $\mathcal{R}_{0,\text{eff}}$, and therefore predicted longer periods than were actually observed. Especially interesting dynamics occurred after 1850, when a long 3-4 and later 4-8 year period is observed in the wavelet spectrum. These changes are predicted by the decline in $\mathcal{R}_{0,\text{eff}}$ from 3.5 to 1, which implies a lengthening of the predicted non-resonant period from 4 to 6. The correlation between the predicted and observed periods are shown as different symbols for each time interval in **Figure 5.4**. Since we did not have a reliable estimate of $\mathcal{R}_0$ at the "anchor time" (initially we used $\mathcal{R}_0 = 4$), we repeated our analysis with $\mathcal{R}_0 = 3,\ 5,\ 6$. The grey bars represent the uncertainty in the intrinsic $\mathcal{R}_0$ (i.e. $\mathcal{R}_0$ at the "anchor time") with longer periods predicted with intrinsic $\mathcal{R}_0 = 3$ and shorter periods predicted with $\mathcal{R}_0 = 6$. The

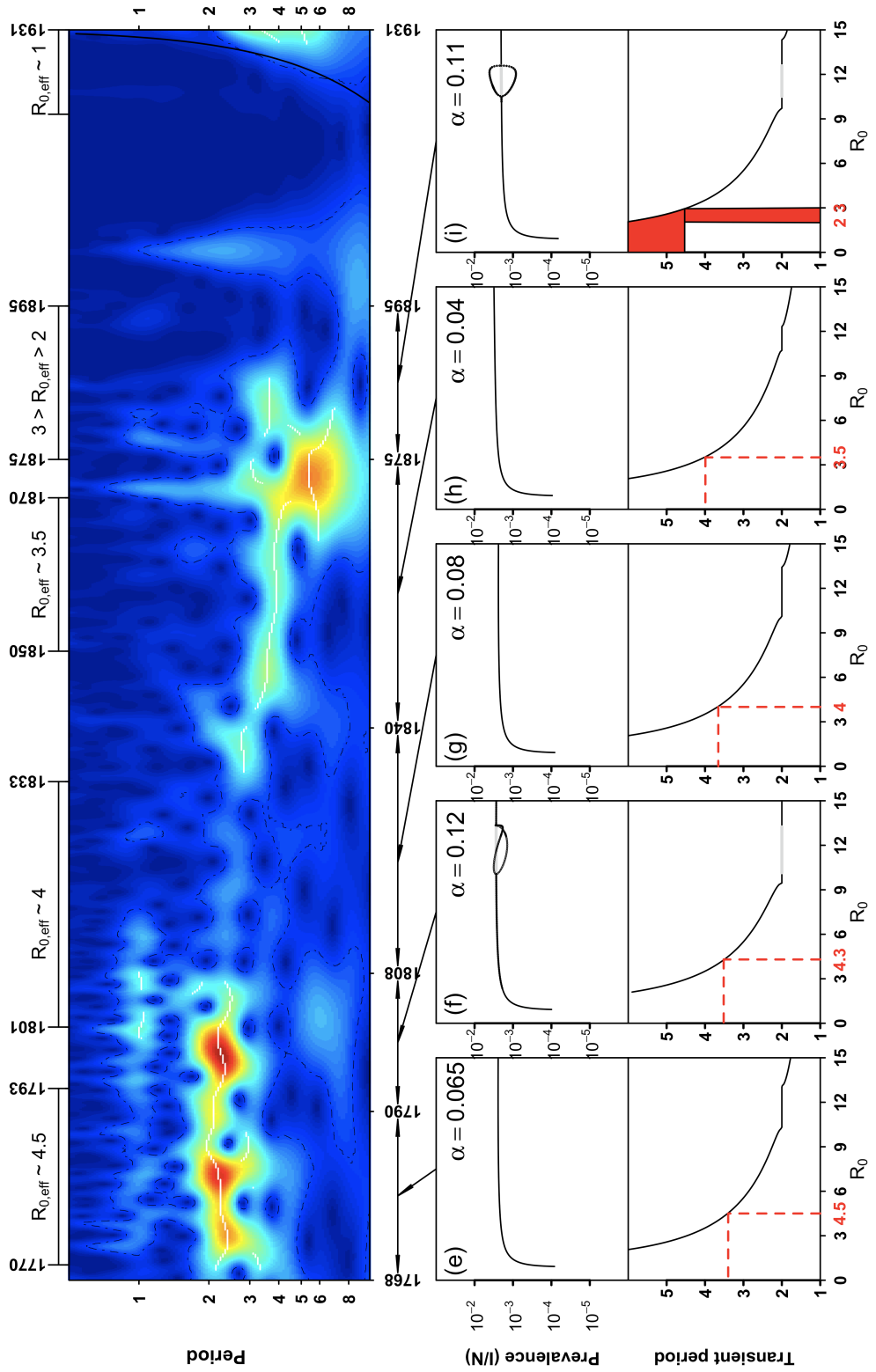


Figure 5.3: Transition analysis of the smallpox mortality time series for London, England (1768-1930). Annotation is as in **Figure 5.2**

best fit (defined by minimum sum of distances to the "predicted=observed" line) is obtained for $\mathcal{R}_0 = 4$.

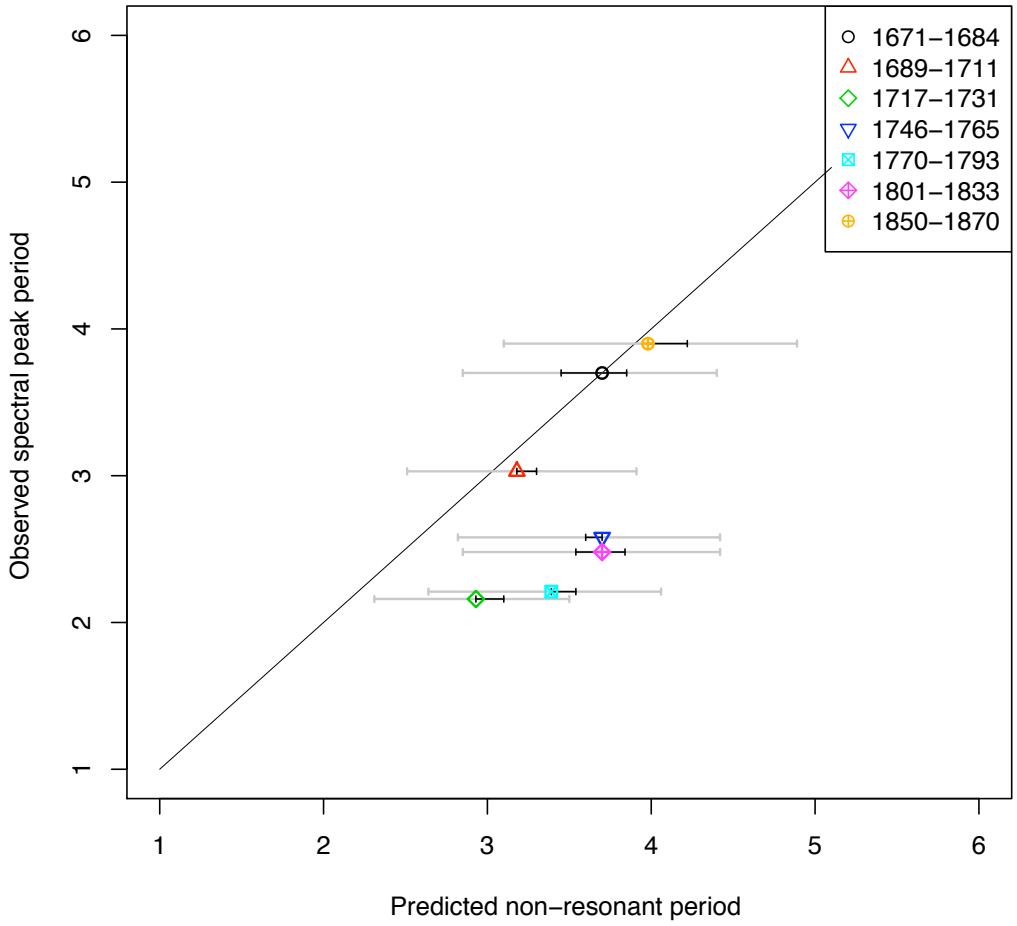


Figure 5.4: The correlation between observed and predicted periods of non-resonant peaks of the wavelet power spectrum. The line of slope 1 indicates where the points would lie if there is a perfect agreement between predicted and observed peaks. $\mathcal{R}_{0,\text{eff}}$ could not be determined precisely, but was given some small range. The black bars correspond to the min/max values of the predicted transient period for such range of $\mathcal{R}_0$, while points shows the period corresponding to the most frequent values of $\mathcal{R}_0$. The grey bars shows the predicted periods for the range of $3 \leq \mathcal{R}_0 \leq 6$ and represent uncertainty in the choice of $\mathcal{R}_0$ at the "anchor time"

5.4 Discussion

The goal of our current analysis was to determine if demographic changes induced by variations in birth rate and population movement, as well as behavioral trends that influenced the level of immunity in the population, can explain the changes in the temporal pattern of smallpox epidemics in London, England. We conducted a transition analysis based on the SIR model with the mean serial interval chosen for smallpox and seasonal forcing as estimated from the smallpox time series (see **Chapter 4**). Previous transition analyses were successfully used to explain a variety of outbreak patterns (Bauch and Earn 2003a,b; Earn et al. 2000a). When applied to the London smallpox mortality time series, transition analysis has also proven to be a powerful method that can correctly predict changes in the frequency structure of smallpox epidemics.

We began our analysis with a statistical description of the data. The periodic structure of smallpox epidemics in London showed very weak power at the only resonant period – one year. Much stronger signals in the wavelet spectrum were observed for non-resonant peaks (i.e. non-integer periods such as 2-3 and 3-4 years). These findings are consistent with conclusions, previously drawn by Bauch and Earn (2003a,b), that diseases with small basic reproduction number ($3 \leq \mathcal{R}_0 \leq 6$ for smallpox) can be expected to have more power in non-resonant peak.

We based our asymptotic and perturbation analyses on the seasonally forced SIR model. To understand how the amplitude of seasonal forcing changed over time we matched the dynamics of the model produced with the seasonal forcing as estimated in **Chapter 4** with the sinusoidally forced SIR model. **Figures 5.2–5.3** show that the amplitude changed significantly over time and fluctuated between 0.032 to 0.12. Previously, analysis based on time series spanning 30-50 years have shown that the amplitude of seasonal forcing, α, does not seem to vary significantly for a particular disease, even when compared between different places. Therefore it was concluded that α can be considered constant (Bauch and Earn 2003a,b).

Our results suggest that the amplitude of seasonal forcing can change significantly with time. Since this parameter influences the dynamics of the model, it should be carefully estimated from data if possible.

Asymptotic analysis has shown that for biologically plausible $\mathcal{R}_0$ the model has only one attractor (an annual cycle), which correctly predicts a resonant one-year period observed in the wavelet spectrogram. Perturbation analysis was less accurate: some of the periods were predicted very well while the others were poorly matched (**Figure 5.4**). However the trend of the highest peak power of the wavelet power spectrum was accurately predicted. We believe that inconsistency in the accuracy of our predictions is mainly caused by our very crude estimate of susceptible recruitment. Uncertainty in the population size, immigration and variolation and vaccination uptake, dramatically reduced the precision of our estimates of susceptible recruitment, which are vital for obtaining accurate predictions. Further analysis of the historical data as well as sensitivity analysis of our results to various parameter values, should help us to refine estimates of the recruitment rate and should be addressed in future research.

Behavioral and demographic changes induced by changes in intervention uptake level, births and immigration had affected the level of susceptibles in the population and therefore were major drivers of the transitions observed in the smallpox time series. The discovery of vaccination in 1796 and its continuously increasing uptake level helped drastically reduce the number of susceptibles in the second half of the 19th century. Introduction of variolation failed to produce a similar effect: when variolation was widely used (1768-1810), smallpox epidemics of high magnitude occurred every 2-3 years. Note that even at its highest level, variolation never exceeded 30% while vaccination was over 50% after 1840 (according to our estimates). Moreover variolated individuals were also a source of smallpox infection for a short period of time (Fenner et al. 1988) and therefore contributed to the spread of infection to the unprotected population. A more realistic model that cap-

tures this mechanism could also improve our predictions and should be considered in future research.

Chapter 6

Conclusions

Predicting the course of an epidemic and providing public health authorities with timely and accurate analysis during disease outbreaks have become principle objectives of mathematical epidemiology. These goals can be achieved only through understanding the mechanisms by which infectious diseases spread in populations. This thesis continues the work of Earn and colleagues (Bauch and Earn 2003a,b; Earn et al. 2000a) on the analysis of complex patterns of infectious diseases and the use of mathematical models to predict dynamical transitions induced by secular changes in demographic and behavioural parameters. We have investigated factors that influenced changes in the temporal structure of smallpox epidemics. Our analysis is based on historical data from the digitized weekly London Bills of Mortality and Registrar General's Returns – an extensive data set that covers a time period much longer than any other time series that has been analyzed to date.

Chapter 2 described the SIR and SEIR models with Gamma distributed disease stages (SI^nR and SE^mI^nR), which more accurately represent reality than the standard exponentially distributed models. We also revisited the method of transition analysis, which is used to predict changes in the periodic structure of epidemics. We investigated how the shape of the stage duration distributions influences the predictions of the transition analysis by systematically analyzing the

sequence of Gamma distributed SIR and SEIR models. As an illustrative application we used the well-known New York City weekly measles incidence data. We discovered that the key parameter that determines the predictions of the transition analysis is the mean serial interval, i.e. SE^mI^nR models exhibit similar behaviour for a given mean serial interval *for any shape parameter values* m and n. This discovery is extremely helpful, since it allows one to model disease dynamics using the simpler SIR model, if the mean serial interval is estimated correctly. Our analysis was based on the mean serial interval for measles. While we can surmise that the above conclusion holds true for the range of values of the mean serial interval, a detailed analysis should be done in the future to verify that the simpler SIR model can be used for other typical recurrent infectious diseases.

Chapters 3, 4, and **5** were focused on the study of the London weekly smallpox mortality data. In **Chapter 3** we presented a statistical description of the observed time series. We also established a timeline of important historical events related to changes in intervention uptake levels, population movement and wars, and correlated them with the transitions observed in smallpox dynamics. Historical sources allowed us to estimate the variolation and vaccination levels as well as to build hypotheses about what may have caused the observed changes in the magnitude and periodicity of smallpox epidemics. This study provided a necessary foundation for modeling work, which we presented in **Chapter 5**.

The statistical analysis in **Chapter 3** also explored the seasonal structure of the smallpox data. It revealed seasonality in the observed time series, suggesting that the transmission rate (β) between susceptible and infectious individuals was seasonally forced. Since seasonal variations in infectious disease transmission have been found to be a major determinant of epidemic dynamics, it was necessary to estimate the seasonality of β from the observed time series, which we did in **Chapter 4**. This task proved to be nontrivial, due to the long length of our data set. The majority of methods found in the literature aim to estimate the disease transmis-

sion rate for relatively short data sets (20-50 years) and if applied to a long data set would be very time consuming and require considerable computational resources. We have presented a simple, easy to implement, and not computationally demanding method based on the continuous-time SEIR model. We applied this method to the smallpox mortality records and estimated the seasonal pattern of smallpox transmission in London for 1664–1930.

Chapter 5 aimed to explain the changes observed in smallpox dynamics over the centuries. Using the method of transition analysis and the standard SIR model with mean serial interval for smallpox, we showed that changes in intervention uptake levels, birth rate and immigration appear to have caused transitions observed in the temporal smallpox pattern. We concluded that transition analysis can be successfully applied to long data sets and is able to predict changes in the smallpox temporal pattern. The major limitation in accuracy of the predictions arises from uncertainty in the estimated susceptible recruitment rate.

A major modelling tool not employed in this thesis is stochastic epidemic modelling and simulations. The inferences made in **Chapter 2** were based on the deterministic dynamics of the SI^nR and SE^mI^nR models. Further work should be done to investigate the stochastic versions of the SI^nR and SE^mI^nR models and determine whether the shapes of stage duration distributions affect the stochastic dynamics of those models. In addition, stochastic simulations of the SIR model for smallpox should be conducted to verify the robustness of our explanations of the observed time series, and examine finer details (Bauch and Earn 2003a,b; Earn 2009).

This thesis also includes a practical guide to constructing bifurcation diagrams with XPPAUT software (**Appendix 2.6**). We have used the seasonally forced SIR model as an illustrative example. We hope that this guide will be a useful learning tool for studying dynamcis of deterministic epidemic models. Another guide included in the thesis (**Appendix 3.5**) explains how to compute a trend in a nonlinear

and nonstationary time series using Empirical Mode Decomposition. This is a sophisticated method, which is becoming popular to analyze neural (Liang et al. 2005) and climatological (Lee and Ouarda 2011) data, and various other types of temporal signals (Bouzid and Ellouze 2004; Mutlu and Aviyente 2011). To our knowledge, we are the first to use this method for estimating trends of epidemiological time series.

In conclusion, transition analysis is a powerful mathematical tool that can be used to understand and predict changes in infectious disease dynamics, and has been particularly helpful in the analysis of the London smallpox mortality data. We hope that our findings and research results will inspire further mathematical analysis of historic time series of infectious disease data.

Bibliography

P. S. Addison. *The illustrated wavelet transform handbook*. Taylor & Francis, 2002.

G. K. Aldis and M. G. Roberts. An integral equation model for the control of a smallpox outbreak. *Mathematical Biosciences*, 195(1):1–22, 2005.

D. Alonso, A. J. McKane, and M. Pascual. Stochastic amplification in epidemics. *Journal of the Royal Society Interface*, 4(14):575–582, 2007.

S. Altizer, A. Dobson, P. Hosseini, P. Hudson, M. Pascual, and P. Rohani. Seasonality and the dynamics of infectious diseases. *Ecology Letters*, 9(4):467–484, 2006.

D. Anderson and R. Watson. On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika*, 67(1):191–198, 1980.

R. M. Anderson and R. M. May. *Infectious diseases of humans: Dynamics and Control*. Oxford University Press, 1991.

J. L. Aron and I. B. Schwartz. Seasonality and periodoubling bifurcations in an epidemic model. *Journal of Theoretical Biology*, 110:665–679, 1984.

N. T. J. Bailey. On estimating of latent and infectious periods of measles: I. Families with two susceptibles only. *Biometrika*, 43(1/2):15–22, 1956a.

N. T. J. Bailey. On estimating of latent and infectious periods of measles: II. Families with three or more susceptibles. *Biometrika*, 43(3/4):322–331, 1956b.

N. T. J. Bailey. Some stochastic models for small epidemics in large populations. *Applied Statistics*, 13(1):9–19, 1964.

N. T. J. Bailey. *The mathematical theory of infectious diseases and its application*. Griffin, London, 2nd edition, 1975.

J. E. Banatvala and D. W. G. Brown. Rubella. *LANCET*, 363(9415):1127–1137, 2004.

M. S. Bartlett. Measles periodicity and community size. *Journal of the Royal Statistical Society, Series A*, 120:48–70, 1957a.

M. S. Bartlett. *Stochastic Population Models in Ecology and Epidemiology*. Methuen, London, 1960.

M. S. Bartlett. Measles periodicity and community size. *Journal of the Royal Statistical Society, Series A*, 120:48–70, 1957b.

C. T. Bauch. The role of mathematical models in explaining recurrent outbreaks of infectious childhood diseases. In F. Brauer, P. van den Driessche, and J. Wu, editors, *Lecture Notes in Mathematical Epidemiology*, volume 1945 of *Lecture Notes in Mathematics*, pages 297–319. Springer, 2008.

C. T. Bauch and D. J. D. Earn. Interepidemic intervals in forced and unforced seir models. *Dynamical Systems and Their Applications in Biology. Fields Institute Communications*, 36:33–44, 2003a.

C. T. Bauch and D. J. D. Earn. Transients and attractors in epidemics. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 270(1524):1573–1578, 2003b.

C. T. Bauch, A. P. Galvani, and D. J. D. Earn. Group interest versus self-interest in smallpox vaccination policy. *Proceedings of the National Academy of Sciences of the United States of America*, 100(18):10564–10567, 2003.

D. Bernoulli. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole. *Mémoires de Mathématique et de Physique L'Imprimerie Royale, Paris*, 1766.

A. J. Black and A. J. McKane. Stochasticity in staged models of epidemics: quantifying the dynamics of whooping cough. *Journal of the Royal Society Interface*, 7(49):1219–1227, 2010a.

A. J. Black and A. J. McKane. Stochastic amplification in an epidemic model with seasonal forcing. *Journal of Theoretical Biology*, 267(1):85–94, 2010b.

S. Blower and D. Bernoulli. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. 1766. *Reviews in Medical Virology*, 14:275–288, 2004.

B. M. Bolker and B. T. Grenfell. Chaos and biological complexity in measles dynamics. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 251:75–81, 1993.

A. Bouzid and N. Ellouze. Empirical mode decomposition of voiced speech signal. *Control, Communications and Signal Processing, 2004. First International Symposium on*, pages 603–606, 2004.

S. A. Bozzette, R. Boer, V. Bhatnagar, J. L. Brower, E. B. Keeler, S. C. Morton, and M. A. Stoto. A model for a smallpox-vaccination policy. *New England Journal of Medicine*, 348(5):416–425, 2003a.

S. A. Bozzette, R. Boer, V. Bhatnagar, J. L. Brower, E. B. Keeler, S. C. Morton, and M. A. Stoto. A model for a smallpox-vaccination policy. *New England Journal of Medicine*, 348(5):416–425, 2003b.

F. Brauer. *Compartmental models in epidemiology*, volume 1945 of *Lecture Notes in Mathematics*, pages 19–79. Springer-Verlag Berlin, Berlin, 2008.

D. S. Burke, J. M. Epstein, D. A. T. Cummings, J. I. Parker, K. C. Cline, R. M. Singa, and S. Chakravarty. Individual-based computational modeling of smallpox epidemic control strategies. *Academic Emergency Medicine*, 13(11):1142–1149, 2006.

S. Cauchemez and N.M. Ferguson. Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in london. *Journal of the Royal Society Interface*, 5(25):885–897, 2008.

B. Cazelles, M. Chavez, G. C. de Magny, J. F. Guegan, and S. Hales. Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society Interface*, 4:625–636, 2007.

B. Cazelles, M. Chavez, D. Berteaux, F. Menard, J. O. Vik, S. Jenouvrier, and N. C. Stenseth. Wavelet analysis of ecological time series. *Oecologia*, 156(2):287–304, 2008.

CDC. Centers for Disease Control and Prevention: History of the MMR vaccine. `http://www.immunizationinfo.org/vaccines/measles`, 2010.

CDC, 2004. Centers for Disease Control and Prevention: Smallpox disease overview, 2004. URL `http://www.bt.cdc.gov/agent/smallpox/overview/disease-facts.asp`.

CDC, 2009. Centers for Disease Control and Prevention: Questions and answers about smallpox disease, 2009a. URL `http://emergency.cdc.gov/agent/smallpox/faq/smallpox_disease.asp`.

CDC, 2009. Centers for Disease Control and Prevention: Smallpox, 2009b. URL `http://emergency.cdc.gov/agent/smallpox/`.

Census. London Census Data: 1801-1931, 2011. URL `http://www.demographia.com/dm-lon31.htm`.

Center for Global Development. Case 1: Eradicating smallpox, 2011. URL `http://www.cgdev.org/section/initiatives/_active/millionssaved/studies/case_1`.

C. Chatfield. *The analysis of time series: an introduction*. Chapman and Hall, 1989.

P. Clark. Migration in England during the Late Seventeenth and Early Eighteenth Centuries. *Past and present*, 83:57–90, 1979.

A. J. K. Conlan, P. Rohani, A. L. Lloyd, M. Keeling, and B. T. Grenfell. Resolving the impact of waiting time distributions on the persistence of measles. *Journal of the Royal Society Interface*, 7, 2010.

A. J. K. Conlan, K. T. D. Eames, J. A. Gage, J. C. von Kirchbach, J. V. Ross, R. A. Saenz, and J. R. Gog. Measuring social networks in British primary schools through scientic engagement. *Proceeding of the Royal Society*, 278:1467–1475, 2011.

J. L. Conrad, R. Wallace, and J. J. Witte. The epidemiologic rationale for the failure to eradicate measles in the United States. *American Journal of Public Health*, 61 (11):2304–2310, 1971.

B. Cooper. Poxy models and rash decisions. *Proceedings of the National Academy of Sciences of the USA*, 103(33), 2006.

C. Creighton. *A history of epidemics in Britain*, volume 2. Frank Cass & Co. Ltd, second edition, 1965.

K. Dietz and J. A. P. Heesterbeek. Bernoulli was ahead of modern epidemiology. *Nature*, 408(6812):513–514, 2000.

K. Dietz and J. A. P. Heesterbeek. Daniel Bernoulli's epidemiological model revisited. *Mathematical Biosciences*, 180:1–21, 2002.

E. Doedel. AUTO: software for continuation and bifurcation problems in ordinary differential equations, 2007. URL `http://indy.cs.concordia.ca/auto/`.

H. B. Dull and J. J. Witte. Progress of measles eradication in the united states. *Public Health Rep*, 83(3):245–248, 1968.

C. J. Duncan, S. R. Duncan, and S. Scott. Oscillatory dynamics of smallpox and the impact of vaccination. *Journal of Theoretical Biology*, 183(4):447–454, 1996.

S. R. Duncan and M. Gyongy. Using the em algorithm to estimate the disease parameters for smallpox in 17th century london. *Proceedings of 2006 IEEE International Conference on Control Applications*, 1-4:2129–2134, 2006.

S. R. Duncan, S. Scott, and C. J. Duncan. Modeling the different smallpox epidemics in England. *Philos Trans R Soc Lond B Biol Sci*, 346(1318):407–419, 1994a.

S. R. Duncan, S. Scott, and C. J. Duncan. Smallpox epidemics in cities in Britain. *Journal of Interdisciplinary History*, pages 255–271, 1994b.

K. T. D. Eames, N. L. Tilston, and W. J. Edmunds. The impact of school holidays on the social mixing patterns of school children. *Elsevier*, 3(2):103–108, 2011.

D. J. D. Earn. Mathematical epidemiology of infectious diseases. In M. A. Lewis, M. A. J. Chaplain, J. P. Keener, and P. K. Maini, editors, *Mathematical Biology*, volume 14 of *IAS/ Park City Mathematics Series*, pages 151–186. American Mathematical Society, 2009.

D. J. D. Earn, P. Rohani, and B. T. Grenfell. Persistence, chaos and synchrony in ecology and epidemiology. *Proceedings of the Royal Society B-Biological Sciences*, 265:7–10, 1998.

D. J. D. Earn, P. Rohani, B. M. Bolker, and B. T. Grenfell. A simple model for complex dynamical transitions in epidemics. *Science*, 287(5453):667–670, 2000a.

D.J.D. Earn, P. Rohani, B.M. Bolker, and B.T. Grenfell. A simple model for complex dynamical transitions in epidemics. *Science*, 287, 2000b.

M. Eichner. Case isolation and contact tracing can prevent the spread of smallpox. *American Journal of Epidemiology*, 158(2):118–128, 2003. Times Cited: 41.

M. Eichner and K. Dietz. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *American Journal of Epidemiologyl*, 158, 2003.

EMD. Emd package in r, 2011. URL `http://cran.r-project.org/web/packages/EMD/index.html`.

B. Ermentrout. *Simulating, analyzing, and animating dynamical systems: a guide to XXPAUT for researchers and students*. Society for Industrial and Applied Mathematics: Philadelphia, 2002.

Z. Feng and H. R. Thieme. Endemic models with arbitrarily distributed periods of infection I: fundamental properties of the model. *SIAM Journal of Applied Mathematics*, 61(3):803–833, 2000a.

Z. Feng and H. R. Thieme. Endemic models with arbitrarily distributed periods of infection II: fundamental properties of the model. *SIAM Journal of Applied Mathematics*, 61(3):983–1012, 2000b.

Z. L. Feng, D. S. Xu, and H. Y. Zhao. Epidemiological models with non-exponentially distributed disease stages and applications to disease control. *Bulletin of Mathematical Biology*, 69(5):1511–1536, 2007.

F. Fenner, D. A. Henderson, I. Arita, Z. Jezek, and I.D. Ladnyi. *Smallpox and its eradication*. Geneva: World Health Organization, 1988.

N. M. Ferguson, M. J. Keeling, W. J. Edmunds, R. Gani, B. T. Grenfell, R. M. Anderson, and S. Leach. Planning for smallpox outbreaks. *Nature*, 425(6959): 681–685, 2003.

P. E. M Fine. The interval between successive cases of an infectious disease. *American Journal of Epidemiology*, 158:1039–1047, 2003.

P. E. M. Fine and J. A. Clarkson. Measles in england and wales .1. an analysis of factors underlying seasonal patterns. *International Journal of Epidemiology*, 11 (1):5–14, 1982.

B. F. Finkenstadt and B. T. Grenfell. Time series modelling of childhood diseases: a dynamical systems approach. *Applied Statistics*, 49(2):187–205, 2000.

R. Finlay. *Population and metropolis: the demography of London 1580-1650*. Cambridge University Press, 1981.

R. Finlay and B. Shearer. *Population growth and suburban expansion*. Longman Group Limited, 1986.

R. Gani and S. Leach. Transmission potential of smallpox in contemporary populations. *Nature*, 414(6865):748–751, 2001.

N. M. Gantz, R. B. Brown, S. L. Berk, and J. W. Myers. *Manual of clinical problems in infectious disease*. Lippincott Williams & Wilkins, fifth edition, 2005.

A. M. Geddes. The history of smallpox. *Clinics in dermatology*, 25(3):152–157, 2006.

J. W. Glasser, S. O. Foster, J. D. Millar, and J. M. Lane. Evaluating public health responses to reintroduced smallpox via dynamic, socially structured, and spatially distributed metapopulation models. *Clinical Infectious Diseases*, 46:S182–S194, 2008.

K. J. Gough. The estimation of latent and infectious periods. *Biometrika*, 64(3): 559–565, 1977.

B. T. Grenfell, O. N. Bjornstad, and J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, 2001.

Z. Grossman. Oscillatory phenomenon in a model of infectious diseases. *Theoretical Population Biology*, 18:204–243, 1980.

I. Grundy. *Lady Mary Wortley Montagu: Comet of the Enlightenment*. Oxford University Press, Oxford, UK, 2001.

M.E. Halloran, i I.M. Longin, A. Nizam, and Y. Yang. Containing bioterrorist smallpox. *Science*, 298, 2002.

W. H. Hamer. The Milroy Lectures on epidemic disease in Englandthe evidence of variability and of persistency of type. *The Lancet*, 167(4305):569–574, 1906a.

W. H. Hamer. The Milroy Lectures on epidemic disease in Englandthe evidence of variability and of persistency of type. *The Lancet*, 167(4306):655–662, 1906b.

W. H. Hamer. The Milroy Lectures on epidemic disease in Englandthe evidence of variability and of persistency of type. *The Lancet*, 167(4307):733–739, 1906c.

A. Hardy. Smallpox in london: factors in the decline of the disease in the nineteenth century. *Medical History*, 27, 1983.

D. He and D. J. D. Earn. Epidemiological effects of seasonal oscillations in birth rates. *Theoretical Population Biology*, 72:274–291, 2007.

D. H. He, E. L. Ionides, and A. A. King. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society Interface*, 7(43):271–283, 2010.

J.A.P Heesterbeek and J.A.J Metz. *The saturating contact rate in epidemic models*, pages 308–310. Cambridge University Press, Cambridge, 1996.

E. P. Hennock. Vaccination policy against smallpox, 1835-1914: a comparison of england with prussia and imperial germany. *The Society for the Social History of Medicine*, 11(1), 1998.

H. W. Hethcote. The mathematics of infectious diseases. *Society for Industrial and Applied Mathematics Review*, 42(4):599–653, 2000.

H. W. Hethcote and D. W. Tudor. Integral equation models for endemic infectious diseases. *Journal of Mathematical Biology*, 9:37–47, 1980.

G. Hooker, S. P. Ellner, L. De Vargas Roditi, and D.J.D. Earn. Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario. *Journal of the Royal Society Interface*, 2011.

R. E. Hope-Simpson. Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps). *Lancet*, 2(6734):549–54, 1952.

D. R. Hopkins. *The greatest killer*. The University of Chicago press, 1983.

T. House, I. Hall, L. Danon, and M. J. Keeling. Contingency planning for a deliberate release of smallpox in great britain-the role of geographical scale and contact structure. *Bmc Infectious Diseases*, 10, 2010.

N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N. Yen, C.C. Tung, and H.H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London*, 454, 1998.

E. L. Ionides, C. Breto, and A. A. King. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49):18438–18443, 2006.

E. Jenner. *The origin of the vaccine inoculation*. London, Shury, 1801.

E. H. Kaplan, D. L. Craft, and L. M. Wein. Emergency response to a smallpox attack: The case for mass vaccination. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10935–10940, 2002.

E. H. Kaplan, D. L. Craft, and L. M. Wein. Analyzing bioterror response logistics: the case of smallpox. *Mathematical Biosciences*, 185(1):33–72, 2003.

M. J. Keeling and B. T. Grenfell. Understanding the persistence of measles: reconciling theory, simulation and observation. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 269(1489):335–343, 2002.

C. Kemp. Bioterrorism: Introduction and major agents. *Journal of the American Academy of Nurse Practitioners*, 13(11):483–491, 2005.

W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A*, 115:700–721, 1927.

D. Kim and H. Oh. Emd: A package for empirical mode decomposition and hilbert spectrum. *The R Journal*, 1(1):40–46, 2009.

A. A. King, E. L. Ionides, and M. J. Pascual, M.and Bouma. Inapparent infections and cholera dynamics. *Nature*, 454:877 880, 2008.

Y.A. Kuznetsov. *Elements of applied bifurcation theory*, volume 112 of *Applied Mathematics Science*. Springer, New York, 1995.

J. Landers. *Death and the metropolis: studies in the demographic history of London*. Cambridge University Press, 1993.

P. J. Landrigan and J. L. Conrad. Current status of measles in the United States. *The Journal of Infectious Diseases*, 124(6):620–622, 1971.

T. Lee and T. B. M. J. Ouarda. Prediction of climate nonstationary oscillation processes with empirical mode decomposition. *Journal of Geophysical Research*, 116:15, 2011.

H. Liang, S. L. Bressler, R. Desimone, and P. Fries. Empirical mode decomposition: a method for analyzing neural data. *Neurocomputing*, 65-66:801–807, 2005.

M. Lima. A link between the north atlantic oscillation and measles dynamics during the vaccination period in england and wales. *Ecology Letters*, 12(4):302–314, 2009.

C. C. Jr. Linnemann. Measles vaccine: immunity, reinfection and revaccination. *American Journal of Epidemiolgy*, 97(6):365–371, 1973.

A. L. Lloyd. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings of the Royal Society of London, Series B*, 268:985–993, 2001a.

A. L. Lloyd. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theoretical Population Biology*, 60:59–71, 2001b.

W. P. London and J. A. Yorke. Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *American Journal of Epidemiology*, 98(6):453–68, 1973. 4767622.

I. M. Longini, M. E. Halloran, A. Nizam, Y. Yang, S. F. Xu, D. S. Burke, D. A. T. Cummings, and J. M. Epstein. Containing a large bioterrorist smallpox attack: a computer simulation approach. *International Journal of Infectious Diseases*, 11 (2):98–108, 2007.

J. Ma and D. J. D. Earn. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology*, 68:679–702, 2006.

T. B. Macaulay. *The history of England from the accession of James II*. New-York: Hurd and Houghton, 1866.

G. Magney, J. Guegan, M. Petit, and B. Cazelles. Regional-scale climate-variability synchrony of cholera epidemics in west africa. *BMC Infectious Diseases*, 7, 2007.

V. Maystruk. *XPPAUT to produce bifurcation diagrams as demonstrated with seasonally forced SIR and SEIR models. A supplementary material to the Bachelor of Arts and Science degree thesis: "Malaria: Transmission Dynamics Modeling the Effect of Seasonal Gametocyte Prevalence"*. McMaster University, Hamilton, 2006.

W. H. McNeill. *Plagues and peoples*. New York : Anchor Books Doubledays, 1998.

M. I. Meltzer, I. Damon, J. W. LeDuc, and J. D. Millar. Modeling potential responses to smallpox as a bioterrorist weapon. *Emerging Infectious Diseases*, 7 (6):959–969, 2001.

A. Mercer. *Disease, mortality and population in transition*. Leicester University Press, 1990.

C. J. E. Metcalf, O. N. Bjornstad, B. T. Grenfell, and V. Andreasen. Seasonality and comparative dynamics of six childhood infections in pre-vaccination copenhagen. *Proceeding of the Royal Society B*, 276:4111 4118, 2009.

J. Middleton. Busting a holiday myth. `http://www.bbc.co.uk/news/magazine-11536678`, 2010.

G. Mooney. "a tissue of the most flagrant anomalies": Smallpox vaccination and the centralization of sanitary administration in nineteenth-century london. *Medical History*, 41, 1997.

A. Y. Mutlu and S. Aviyente. Multivariate empirical mode decomposition for quantifying multivariate phase synchronization. *EURASIP Journal on Advances in Signal Processing*, page 13, 2011.

R. J. Nelson, G. E. Demas, S. L. Klein, and L. J. Kriegsfeld. *Seasonal patterns of stress, immune function, and disease*. Cambridge University Press, Cambridge, 2002.

H. T. H. Nguyen and P. Rohani. Noise, nonlinearity and seasonality: the epidemics of whooping cough revisited. *Journal of the Royal Society Interface*, 5:403–413, 2008.

H. Nishiura. Early efforts in modeling the incubation period of infectious diseases with an acute course of illness. *Emerging Themes in Epidemiology*, 4(2), 2007a.

H. Nishiura. Determination of the appropriate quarantine period following smallpox exposure: an objective approach using the incubation period distribution. *International Journal of Hygiene and Environmental Health*, 212(1):97=104, 2007b.

H. Nishiura and M. Eichner. Infectiousness of smallpox relative to disease age: estimates based on transmission network and incubation period. *Epidemiology and Infection*, 135(7):1145–1150, 2007.

H. Nishiura and T. Kashiwagi. Smallpox and season: Reanalysis of historical data. *Interdisciplinary Perspectives on Infectious Diseases*, 2009, 2009.

H. Nishiura and I. M. Tang. Modeling for a smallpox-vaccination policy against possible bioterrorism in japan: The impact of long-lasting vaccinal immunity. *Journal of Epidemiology*, 14(2):41–50, 2004. Times Cited: 1.

L. F. Olsen and W. M. Schaffer. Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. *Science*, 249:499–504, 1990.

D. A. Rand and H. B. Wilson. Chaotic stochasticity - a ubiquitous source of unpredictability in epidemics. *Proceedings of the Royal Society B-Biological Sciences*, 246(1316):179–184, 1991.

P. Razzell. *The conquest of smallpox: The impact of inoculation on smallpox mortality in eighteenth century Britain*. Caliban Books, 1977.

S. Riley and N. M. Ferguson. Smallpox transmission and control: spatial dynamics in Great Britain. *Proceedings of the National Academy of Sciences of the United States of America*, 103(33):12637–12642, 2006.

G. Rilling, P. Flandrin, and P. Goncalves. On empirical mode decomposition and its algorithms. *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, NSIP-03, 2003.

R. Ross. *The Prevention of Malaria*. Murray, London, 2nd edition, 1911.

K. J. Ryan and C. G. Ray. *Sherris medical microbiology*. McGraw-Hill, 2004.

D. Schenzle. An age-structured model of pre- and post-vaccination measles transmission. *IMA Journal of Mathematics Applied in Medicine and Biology*, 1:169–191, 1984.

R. Schmidt. A glossary of archaic medical terms, diseases and causes of death, 2011. URL `http://www.antiquusmorbus.com/`.

I. B. Schwartz and H. L. Smith. Infinite subharmonic bifurcation in an seir epidemic model. *Journal of Mathematical Biology*, 18(3):233–253, 1983.

R.H. Shumway and D.S. Stoffer. *Time series analysis and its applications*. Springer, 2006.

H. Soper. The interpretation of periodicity in disease prevalence. *Journal of the Royal Statistical Society*, 92:34–73, 1929.

A. Svensson. A note on generation times in epidemic models. *Mathematical Biosciences*, 208:300–311, 2007.

C. Therrien and M. Tummala. *Probability for Electrical and Computer Engineers*. Taylor and Francis, 2 edition, 2011.

C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78, 1998.

USE. US Department of State: Structure of U.S. Education. `http://infousa.reingoldweb.com/education/overview/edlite-structure-us.html`, 2011.

V. Vatche and y R. Sharple. Decomposition of functions into pairs of intrinsic mode functions. *Proceedings of the Royal Society*, 464(2097), 2008.

H. J. Wearing, P. Rohani, and Keeling M. J. Appropriate models for the management of infectious diseases. *PLOS Medicine*, 2(7):621–627, 2005.

WHO, 1996. World Health Organization: The World Health Report, 1996. URL `http://www.who.int/whr/1996`.

WHO, 2004. World Health Organization: The global burden of disease: 2004 update, 2004. URL `http://www.who.int/healthinfo/global_burden_disease`.

WHO, 2011. World Health Organization: Smallpox fact sheet, 2011a. URL `http://www.who.int/mediacentre/factsheets/smallpox`.

WHO, 2011. World Health Organization: Archives of the smallpox eradication programme, 2011b. URL `http://www.who.int/archives/fonds_collections/bytitle/fonds_6`.

S. Wiggins. *Introduction to applied nonlinear dynamical systems and chaos*. Springer-Verlag: New York, 2003.

J. J. Witte and N. W. Axnick. The benefits from 10 years of measles immunization in the united states. *Public Health Rep*, 90(3):205–7, 1975. 807933.

E. Wrigley and R. Schofield. *The Population History of England*. Arnold, London, UK, 1981.

Z. Wu, N.E. Huang, S.R. Long, and C. Pend. On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences*, 104(38), 2007.

Y. Xia, J. R. Gog, and B. T. Grenfell. Semiparametric estimation of the duration of immunity from infectious disease time series: inuenza as a case-study. *Journal of the Royal Statistic Society C*, 54:659 672, 2005.

J. A. Yorke and W. P. London. Recurrent outbreaks of measles, chickenpox and mumps .2. systematic differences in contact rates and stochastic effects. *American Journal of Epidemiology*, 98(6):469–482, 1973.

T. Zenihana and H. Ishikawa. Effectiveness assessment of countermeasures against bioterrorist smallpox attacks in japan using an individual-based model. *Environmental Health and Preventive Medicine*, 15(2):84–93, 2010.

F. Zhang, Z. Li, and F. Zhang. Global stability of an sir epidemic model with constant infectious period. *Applied Mathematics and Computation*, 199:285–291, 2008.