ESTIMATING SEASONAL DRIVERS IN CHILDHOOD INFECTIOUS DISEASES WITH CONTINUOUS TIME MODELS

A Thesis

by

GEORGE HENRY ABBOTT III

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2010

Major Subject: Chemical Engineering

ESTIMATING SEASONAL DRIVERS IN CHILDHOOD INFECTIOUS DISEASES WITH CONTINUOUS TIME MODELS

A Thesis

by

GEORGE HENRY ABBOTT III

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Carl D. Laird
Committee Members,	Mahmoud El-Halwagi
	Juergen Hahn
	Michael Pilant
Head of Department,	Michael Pishko

May 2010

Major Subject: Chemical Engineering

ABSTRACT

Estimating Seasonal Drivers in Childhood Infectious Diseases
with Continuous Time Models. (May 2010)
George Henry Abbott III, B.S.; B.A., University of North Texas
Chair of Advisory Committee: Dr. Carl D. Laird

Many important factors affect the spread of childhood infectious disease. To understand better the fundamental drivers of infectious disease spread, several researchers have estimated seasonal transmission coefficients using discrete-time models. This research addresses several shortcomings of the discrete-time approaches, including removing the need for the reporting interval to match the serial interval of the disease using infectious disease data from three major cities: New York City, London, and Bangkok. Using a simultaneous approach for optimization of differential equation systems with a Radau collocation discretization scheme and total variation regularization for the transmission parameter profile, this research demonstrates that seasonal transmission parameters can be effectively estimated using continuous-time models. This research further correlates school holiday schedules with the transmission parameter for New York City and London where previous work has already been done, and demonstrates similar results for a relatively unstudied city in childhood infectious disease research, Bangkok, Thailand.

To Sreenath

ACKNOWLEDGMENTS

I extend appreciation to Dr. Carl D. Laird whose guidance proved invaluable in this endeavor. I thank Daniel Word for collaboration on our paper for the American Control Conference. I also acknowledge Sean Legg for his editing and corrections and German Oliveros, whose undergraduate research showed this to be a promisingly successful venture.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

A. Motivation

There are two major perspectives from which can be seen the desire and benefits of a reliable mechanistic model for the spread of infectious diseases. First, from a scientific and research perspective, having a reliable mechanistic model will provide improved understanding and identification of important factors affecting infectious disease dynamics. The dynamics of infectious disease spread are not only dependent on the characteristics of the disease itself, but also upon the social networks and geographical landscape within which the disease spreads. Second, from a public health perspective, having a long-term, quantitative, reliable, dynamic model could allow prediction of outbreaks, better response planning, and improvement in public health policy. Models such as these would ideally describe past system behaviour and be capable of effective extrapolation with adjustments to system inputs, the benefits of which will be felt the most within developing countries where infectious diseases continue to be a major cause of suffering and mortality [1]. Meeting both of these goals requires reliable estimation of disease model parameters and existing case data. The work in this thesis builds upon existing work in estimation of seasonal transmission parameters with discrete models to develop a sufficient framework for efficient estimation of parameters using continuous differential equation models.

The desire to transition from previous discrete disease model estimation methods to continuous-time disease model estimation methods primarily arises from trying to address a few drawbacks and inherent assumptions contained within the former

---

The journal model is *IEEE Transactions on Automatic Control.*

methods. These are several issues this work is concerned with:

- the creation of an estimation framework for continuous models with the flexibility to allow the use of different modelling structures,
- the actual infection process of measles and chickenpox is a continuous process,
- previous estimation procedures for discrete-time models relied on a log-transform that was specific to only that model [2],
- developing a estimation procedure that will handle data in its native format,
- developing an estimation procedure with the freedom to select a discretization strategy, allowing the use of higher order estimation methods.

Since disease modelling relies heavily on recorded real-world case data as input, a model which has the flexibility to use the data in its original form regardless of the properties of the serial interval of the disease and the reporting interval of the data itself is truly beneficial. Usually the recording interval is reduced to match the serial interval of the infection, in the case of measles this is on average one biweek or fourteen days and in the case of chickenpox the period averages approximately twenty days [3]. Another benefit is that the transmission parameter is discretized into the number of time points within the repeating seasonal time frame, in the case of measles this results in a yearly periodic function discretized into twenty six points. This proves to be very limiting if a finer discretization of the transmission parameter is desired, however with a continuous disease model estimation, matching the serial interval with the reporting interval is ideally unnecessary and allows for a greater and variable number of discretization points in the transmission parameter function.

B. Summary

Chapter I, Introduction, contains both the summary and motivation of this thesis.

Chapter II, Background, begins with an overview of the history of mathematical infectious disease modelling techniques before going into the most commonly used differential equation disease models, particularly compartment based disease models. The next section then covers various techniques that are used to estimate and solve dynamic differential equation systems.

Chapter III, Problem definition, explains the problem of formulating the continuous time disease model. The first section explains the problem of discretizing the differential equations with various methods. The second section explains the reason for needing and options for using regularization on the transmission parameter function and calculation of the regularization parameter by use of the l-curve method. The final section of this chapter covers the data sets used in the model estimations.

Chapter IV, Solution Approach, explains the formulation of the problem. It discusses the use of orthogonal collocation on finite elements with Radau points as a discretization technique. It further explains how and why total variation is used to regularize the transmission parameter function. In the last section, it explains the process of simulating artificially generated data sets first in order to arrive at the final model which will be ready to use the true data sets as input.

Chapter V includes the results of the L-curve calculation, the infectious data fit, and the resulting beta graphs are presented along with other important information.

Chapter VI presents the conclusions reached with developing the continuous-time disease model, compares the beta functions to school holiday patterns within the year, and proposes improvements and uses which may become the focus of future research.

CHAPTER II

BACKGROUND

The beginning of mathematical modelling of infectious diseases started in 1760 when Daniel Bernoulli created and analyzed a model for smallpox [4]. However this model was only capable of determining the variolation of healthy people with the smallpox virus and it was not until the 20th century that deterministic epidemiological models arose with Hamer [5]. Hamer formulated a discrete-time model which he used to understand the spread of measles, and most believe that this was the first use of the assumption that the new number of infectives per unit time (incidence) is directly related to the product of the densities of the susceptibles and infectives. Ross [6] used a differential equation model to understand the incidence and control of Malaria. Soon afterwards, several papers concerning different deterministic epidemiology models were published [7, 8, 9, 10]. While these models were not able to successfully predict when an outbreak might occur, they nonetheless were able to be used in analyzing important characteristics of these infections. In the late 1920s Kermack and McKendrick published an important result showing that an epidemic outbreak can occur only if a threshold density of susceptibles is exceeded. If the susceptible density is too low, then it is virtually impossible for the disease to develop into an epidemic within a population [11, 12]. Throughout the 20th century one can see from the many reviews of the literature how quickly the body of research concerning infectious disease modelling has grown [13, 14, 15].

From the middle of the twentieth century on, mathematical epidemiology has seemed to grow exponentially from its very basic beginnings in 1760. The publication of Bailey [7] is considered by many in the field to be a significant historical marker [1]. Many phenomena associated with different diseases have been incorporated into

models such as: passive immunity, gradual loss of vaccine induced immunity, disease acquired immunity, different stages of infection, age structure, social and sexual mixing groups, vaccination, quarantine, and chemotherapy [1].

A. Compartment-Based Disease Modeling

Compartment-based disease models separate the total population into the various compartments based on the different disease stages. The most commonly used models contain five types of compartments, but their inclusion and possible repetition within the model is determined by the behavior of the disease. The common compartments are M,S,E,I, and R. M is referred to as the maternal immunity compartment. This first compartment contains young children who are temporarily immune to the disease because of antibodies obtained from their mother. The second compartment listed is S and contains the susceptibles of the population. Susceptibles are individuals who have no passive maternal or genetic immunity to a disease and therefore are susceptible to it. E is the exposed individuals. The exposed individuals were previously in the susceptible category, but by some population mixing they came into contact with an infected individual and are now considered as infected but not infectious. That is, they contain the infection but are in the latent category where they cannot yet spread it to others. The next compartment is the I compartment. This is labeled I because it houses all of the infected individuals who can spread the infection to susceptible individuals. The final compartment is the R or recovered compartment. This compartment represents individuals who have gained infection-acquired immunity. However, it should be made clear that R is not necessarily a permanent classification, but is determined by the dynamics of the infection itself. In general the classification of the model is determined by the flow of the population from one

compartment to another and here are just a few to consider: SIR, SEIR, MSEIR, SIRS, SEIRS, SI, and SIS. [1]

These models can be used to provide crucial information about the diseases. One of the most important values that these models are associated with is the basic reproduction number, R_0. This value is defined as the average number of secondary infections when one infected individual is introduced into a purely susceptible population [16, 17]. Importantly, if $R_0 > 1$ then the infection has the potential to invade a population and, depending on other characteristics of the infection, may become an epidemic. Therefore, the basic reproduction number is also sometimes referred to as the threshold value of the infection, clearly indicating if the disease will die out or if it has the potential to invade and persist.

In this research, the above compartmental framework is used to model the childhood diseases, measles and varicella zoster or chickenpox. Prior to widespread vaccination, if someone became infected with measles or chickenpox and survived then they attained lifelong infection-acquired immunity [3]. Therefore, three compartments are necessary for the modelling of this infection: S, the susceptible, I, the people who acquire the infection and can transmit it to others, and R the people who are resistant either from vaccination or infection-acquired immunity. These three compartments have been shown to capture the behavior of measles dynamics [2]. Since chickenpox shares the same qualities of measles in being a childhood infectious disease where a survivor attains life-long acquired immunity, the same model is used for capturing chickenpox behavior as well. Equations 2.1 and 2.2 show this model.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{-\beta(t)S(t)I(t)}{N} + \mu(t)N \tag{2.1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta(t)S(t)I(t)}{N} - \gamma I(t) \tag{2.2}$$

Here in the above equations, S is the number of people in the susceptible compartment, and I is the number of people in the infected compartment. Notice that there is no equation to describe $\mathrm{d}R/\mathrm{d}t$. Since the entire population, N, is equal to the number of people in all of the compartments (or $N = S + I + R$) then the number of recovered individuals can always be expressed by $R = N - (S + I)$. Therefore, for simplification of the model and calculations, $\mathrm{d}R/\mathrm{d}t$ is omitted [1]. There are three main parameters in this model: $\beta(t), \gamma$, and $\mu(t)$. The time varying parameter, $\beta(t)$, is called the contact rate or transmission parameter and represents the number of contacts sufficient to spread the infection for a single person per unit time. The recovery parameter, γ, can be understood physically as the inverse of the average period of infection. The third and last parameter in this model is $\mu(t)$, the time varying birth rate. It can clearly be seen that this model uses a similar assumption as Hamer [5]. Here, instead of the product of the densities, $(I(t)/N)(S(t)/N)$, the numbers in each compartment divided by the total population is used, $I(t)S(t)/N$, a term formed from the use of the pseudo-mass action law [1]. This term is negative in (2.1) and positive in (2.2) showing that these individuals move from the susceptible compartment into the infective compartment. The second term in (2.1) explains the new entries into the susceptible compartment. This is represented by $\mu(t)N$, or birth rate times the total population. Since measles is considered a childhood disease with life-long acquired immunity, this is the only physical way to enter the susceptible compartment. The second term in (2.2) is the recovery term. This is essentially accounting for the infectives who pass through their infective period and recover from the disease. This model assumes a constant population of N with a continuous input of births into the S compartment. Since measles and chickenpox are rarely fatal and the average age of infection is quite young, ≈ 5 years old for measles and ≈ 6 years old for chickenpox [3], a death term is unnecessary for the S compartment. It is assumed that all deaths

occur from the R compartment. Constant population is a reasonable assumption with disease models when dealing with data that spans short time domains; however, this assumption can be problematic when dealing with longer time domains.

It has long been recognized that seasonal patterns of measles transmission appear to be correlated with school holiday schedules. Soper [18] estimates transmission rates from monthly incidence data for measles and asserts the possibility of a seasonal pattern related to school terms. Fine and Clarkson [19] extend this idea and estimate a time-varying profile for the transmission parameter using measles data for England and Wales from 1950 to 1965. Even though incidence follows a biennial pattern of alternating major and minor epidemics, their estimates of the time-varying transmission parameter are remarkably consistent from year to year. Furthermore, the yearly pattern of the transmission parameter is loosely correlated with the school holiday schedule.

The time-series SIR model developed in Finkenstädt and Grenfell [2] is a discrete time model discretized over 26 biweeks in each year (consistent with the serial interval for measles). This discrete time model allows for a temporally varying transmission parameter with a period of one year (given 26 unique parameter values for $\beta(\tau)$). This pioneering work makes no strong assumption about the functional form of the transmission parameter with the only restriction being yearly periodicity. Considering measles incidence data from England and Wales, their results further strengthen the evidence that measles transmission is indeed correlated with school terms.

Their discrete time model approach has several drawbacks. First, the model is discretized over the serial interval of the disease, and this interval must agree with the availability of the data. However, the data is often available at a different timescale than the serial interval. Furthermore, the discrete time-series SIR model has an additional parameter α in the incidence term (i.e. $\beta(\tau)I_i^{\alpha}S_i$), that is difficult

to interpret physically. This term has been spoken of in Xia et al. [20] as an adjusting term which accounts for the inherent errors of the discrete model, suggesting further there is no need of it in the continuous model.

The goal is to extend these results and estimate seasonal transmission parameters using a continuous time model by exploring different forms of regularization for the transmission parameter and performing the estimation using data sets where the reporting interval does not match the serial interval. The techniques necessary for accomplishing this requires formulation and estimation of dynamic systems.

B. State and Parameter Estimation of Dynamic Systems

The general form for a continuous system can be written as,

$$\dot{x} = f(x, u, w, p) \tag{2.3}$$

$$y = h(x, u, v, p). \tag{2.4}$$

Here x is the state vector, u is the input vector, w is model noise, p is the parameter vector, y is the output vector, and v is measurement noise. The two functions here, f and h, are either linear or nonlinear. In the case of infectious disease model estimation they are nonlinear.

1. Linear State Estimation

Linear State Estimation has two main categories which are described by how the data is processed. If the data is estimated one data point at a time then it is referred to as sequential state estimation, the most common technique being Kalman filtering [21, 22]. Kalman filtering is specifically interested in determining the states of the system over the entire time domain. It begins by taking an initial estimation and sys-

tematically calculating the current state estimate with respect to the corresponding measurement. Then through a series of calculations, the Kalman filter updates or improves that state estimate to better fit the measurement. Then the updated estimate is used to estimate the state for the next time value. Kalman filters can be formulated for either discrete or continuous time systems. Kalman filtering approaches have also been extended to nonlinear models.

The second category of linear state estimation is called Batch State Estimation [22]. Here, the technique differs from sequential state estimation because of the amount of information used to estimate the states at each iteration. Sequential state estimation considers the data one point at a time, while batch considers equal-sized subsets of the domain and optimizes the trajectory of the state for just that subset. The basic strategy of this estimation technique is to run two simultaneous Kalman filters across the subset. One runs forward in time and the other runs backward in time. The estimations are then corrected with respect to both filters. While these techniques are effective, it is difficult to handle constraints on the states.

2. Nonlinear State Estimation

A fundamental difference between linear and nonlinear problems is that unlike the linear case, the stability of the nonlinear problem depends on the particular solution trajectory considered [23]. One very popular method for estimation of nonlinear states is the extended Kalman filter, (EKF) [22, 24]. This method is based on linearization and the assumption that the estimated state values are sufficiently close to the true states [22]. The error dynamics are represented by a first-order Taylor series

expansion. Therefore for the general model:

$$\dot{x}(t) = f(x(t), u(t), t) + G(t)w(t) \tag{2.5}$$

$$\tilde{y}(t) = h(x(t), t) + v(t) \tag{2.6}$$

the functions f and h would be estimated as,

$$f(x(t), u(t), t) \cong f(\bar{x}(t), u(t), t) + \left[\frac{\partial f}{\partial x}\right]_{\bar{x}(t)} [x(t) - \bar{x}(t)] \tag{2.7}$$

$$h(x(t), t) \cong h(\bar{x}(t), t) + \left[\frac{\partial h}{\partial x}\right]_{\bar{x}(t)} [x(t) - \bar{x}(t)] \tag{2.8}$$

where $\bar{x}(t)$ is sufficiently close to $x(t)$.

The extended Kalman filter structure for the state and output estimate is given by [22],

$$\begin{aligned} \dot{\hat{x}}(t) = f(\bar{x}(t), u(t), t) + F(\bar{x}(t), t)[\hat{x}(t) - \bar{x}(t)] \\ + K(t)\left[\tilde{y}(t) - h(\bar{x}(t), t) - H(\bar{x}(t), t)[\hat{x}(t) - \bar{x}(t)]\right] \end{aligned} \tag{2.9}$$

$$\hat{y}(t) = h(\bar{x}(t), t) + H(\bar{x}(t), t)[\hat{x}(t) - \bar{x}(t)] \tag{2.10}$$

where $\hat{x}(t)$ is the nominal state value (i.e. $E\{f(x(t), u(t), t)\} = f(\hat{x}(t), u(t), t)$), $\bar{x}(t)$ is the nominal a priori state vector, $F \equiv \partial f/\partial x$ from (2.7), $H \equiv \partial h/\partial x$ from (2.8), and $K(t) = P(t)H^T(\hat{x}(t), t)R^{-1}(t)$ is the Kalman gain ($P(t)$ is state-error covariance and $R(t)$ is measurement-noise covariance) [22]. This method can be reformulated for several different models, and has been improved by applying local iterations linearizing around the most recent estimates in the iterated extended Kalman filter. It has also been formulated to fit discrete models, continuous models, and continuous-discrete models where the model is continuous but the measurements are discrete [24]. The advantage with using this method is that it is a computationally efficient "one-step

recursion," however the disadvantages are that it is difficult to implement, tune, and it is only reliable for systems which are linear on the timescale [25]. For the disease estimation problem we have nonlinear-batch-constraints, the techniques listed so far do not typically handle constraints.

C. Constrained Estimation Approaches

There is a class of methods for constrained dynamic systems. Rawlings and Mayne[25] shows that when a dynamic system is modified to include constraints, it becomes a nonlinear optimization problem with differential algebraic equation constraints, or concisely DAE constraints. The estimation problem is explicitly formulated as an optimization problem. The necessary methods can be grouped into two general categories, the indirect and the direct [26]. The indirect methods, also called variational approach methods, are based on the solution of the first order necessary conditions for optimality that are obtained from Pontryagins Maximum Principle [27]. For problems without inequality constraints this give rise to a two-point boundary value problem (TPBVP). This technique is often labeled "optimize-then-discretize," since the optimality conditions are applied to the original optimization formulation and then the resulting system is discretized for solution. However, for more complicated problems with inequality constraints, the indirect method becomes increasingly difficult [28]. The estimation model used in this research has several inequality constraints, therefore this approach is unfavorable.

Direct methods are often labeled "discretize-then-optimize," since they first discretize the original DAE formulation and then apply the optimality conditions to the algebraic problem. The direct methods are further divided into two categories: sequential and simultaneous. In the sequential methods, only the control variables are

discretized. They are typically represented in the formulation as piecewise polynomials in what is known as the control variable parameterization methods. The goal of the optimization is to find the coefficients of these polynomials. The integration is solved in an inner loop while the control variable parameters are updated in an outer loop using the NLP solver. The advantages of this method are that they are relatively easy to construct and reliable DAE solvers are available. The disadvantages, however, become more apparent when one realizes that multiple numerical integrations of the DAE, along with the calculation of the derivatives, can become very time consuming for large problems [26].

The main methods that will be used with these models are the simultaneous, or direct transcription methods [29, 26]. These techniques completely discretize the state and control variables using a higher order fixed-step discretization strategy. The main feature of the simultaneous methods is that the solution of the DAE is converged simultaneously with the optimization problem, avoiding intermediate calculations and unnecessary computational effort. While these techniques are more difficult to implement, they can be very efficient. Furthermore, they allow direct enforcement of constraints on the state and control variables at the same level of discretization as the DAE system.

1. Discretization Methods

Using the direct simulation approach requires that the DAE be first discretized with a fixed timestep. There are several methods to consider. The simplest method of differential equation discretization is the explicit Euler or forward Euler method. The first step is to determine a mesh over the time period of the simulation,

$$a = t_0 < t_1 < ... < t_{n-1} < t_n = b \tag{2.11}$$

where a and b are the starting and ending points of the time domain respectively. The nth step size is denoted and defined as $h_n = t_n - t_{n-1}$. According to Ascher and Petzold [23], the next step is to construct the state approximations $x_0, x_1, ..., x_{n-1}, x_n$, with each x_n intended to be an approximation of $x(t_n)$. The concept of the explicit Euler method is to begin with the initial value, x_0, then proceed to integrate the model to each successive time step from the previous one (i.e. using x_{n-1} at t_{n-1} to roughly calculate x_n at t_n),

$$\frac{x_n - x_{n-1}}{h_n} = f(t_{n-1}, x_{n-1}). \tag{2.12}$$

It is important to note that the order of this method is highly dependent on the step size of each h_n, the smaller the step size the greater the accuracy. However, the forward Euler method is only accurate to order 1 (i.e. $error = O(h^1)$)[23].

The explicit Euler method may violate the absolute stability requirement when used to solve an IVP of a stiff equation, so the implicit Euler method was developed as an alternative method to avoid violating the requirement [23]. The implicit is derived the same way as the explicit but it is centered around the point t_n instead of t_{n-1}, giving

$$\frac{x_n - x_{n-1}}{h_n} = f(t_n, x_n). \tag{2.13}$$

This method is implicit since the unknowns appear on both sides of the equation. However, since the equations will be converged as part of the constraints in the optimization problem, this is not a serious complication.

While the implicit Euler strategy does provide more stability the numerical error, it attains can be improved upon by other techniques. The Runge-Kutta methods are ideal when a higher order of accuracy is desired without having to form the symbolic derivatives of the Taylor series method. If the explicit and implicit Euler methods

were to be reinterpreted as a quadrature, the methods would take on the form,

$$x_n - x_{n-1} = \int_{t_{n-1}}^{t_n} x'(\zeta) d\zeta \tag{2.14}$$

where the explicit method approximates the area under the curve based on $x'(t_{n-1})$ and the implicit method uses $x'(t_n)$, also called the left and right Riemann sum, respectively. For a better method, the midpoint of these two techniques may be calculated and used to obtain the implicit method,

$$x_n = x_{n-1} + hf\left(t_{n-1/2}, \frac{x_{n-1} - x_n}{2}\right) \tag{2.15}$$

To transform this technique into an explicit method, one may first use the explicit Euler method (2.16), to approximate the value $\hat{x}_{n-1/2}$ before then substituting into the implicit midpoint method (2.17), to arrive at something similar to the Runge-Kutta method.

$$\hat{x}_{n-1/2} = x_{n-1} + \frac{h}{2} f(t_{n-1}, x_{n-1}) \tag{2.16}$$

$$x_n = x_{n-1} + hf(t_{n-1/2}, \hat{x}_{n-1/2}) \tag{2.17}$$

This method has been shown to achieve an order of accuracy consistent with order 2 [23]. The classical fourth-order Runge-Kutta is an elaboration of (2.14), but utilizing the Simpson's quadrature estimation,

$$x_n - x_{n-1} \approx \frac{h}{6}(x'(t_{n-1}) + 4x'(t_{n-1/2}) + x'(t_n)) \tag{2.18}$$

then the explicit estimation is used to find the approximate value for $y'(t_{n-1/2})$. Giving

the commonly known,

$$\begin{aligned} k_1 &= x_{n-1}, \\ k_2 &= x_{n-1} + \frac{h}{2} f(t_{n-1}, k_1) \\ k_3 &= x_{n-1} + \frac{h}{2} f(t_{n-1/2}, k_2) \\ k_4 &= x_{n-1} + h f(t_{n-1/2}, k_3) \\ x_n &= x_{n-1} + \frac{h}{6}(f(t_{n-1}, k_1) + 2f(t_{n-1/2}, k_2) + 2f(t_{n-1/2}, k_3) + f(t_n, k_4)). \end{aligned} \tag{2.19}$$

While Runge-Kutta methods are highly stable and accurate, when using them with DAEs one is often faced with order reduction, the causes of which are closely related to the causes of order reduction for Runge-Kutta methods applied to stiff ODEs [23]. Therefore a method more compatible with DAEs, such as collocation on finite elements, should be explored.

Collocation methods can be shown to be equivalent to Runge-Kutta Methods [23], but collocation will be further elaborated later in this text.

CHAPTER III

PROBLEM DEFINITION

A. Model Formulation

The base estimation formulation used to motivate this research is given as,

$$\min \quad -\mathcal{L}(\cdot) \tag{3.1}$$

$$\text{s.t.} \quad \frac{\mathrm{d}S}{\mathrm{d}t} = \frac{-\beta(\tau)S(t)I(t)}{N} + \mu(t)N \tag{3.2}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta(\tau)S(t)I(t)}{N} - \gamma I(t) \tag{3.3}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I(t) - \mu(t)N \tag{3.4}$$

$$\tag{3.5}$$

where S is the number of susceptibles, I is the number of infectives, and R is the number of recovered individuals. The objective function,$-\mathcal{L}(\cdot)$, is some measure of fit for the data. The seasonal transmission parameter $\beta(\tau)$ is time varying, but will be restricted to have a yearly periodicity; therefore, $\tau = t \mod t_{year}$ with t_{year} being the amount of time within one year. The birth rate, $\mu(t)$, is a known time-varying input to the system. The population, N, and recovery rate, γ are known scalar inputs. When formulating the objective function (measure of fit), it is important to note the distinction between prevalence and incidence. In our case count data, the number recorded is the incidence, or number of new cases which have arisen within that specific recording period. However, the variable I in the model represents prevalence, not incidence. The discrete time models did not have this issue. Since they assumed that the recovery time was identical to the time discretization, they blur the distinction between incidence and prevalence. Over a particular reporting interval, the incidence is given by $\int_{t_{i-1}}^{t_i} \beta(\tau)I(t)S(t)\mathrm{d}t$. To include this feature in the

optimization formulation, a new variable $\phi(t)$ that represents the cumulative incidence at time t will be introduced. This can be easily implemented in the formulation by adding the differential equation $\mathrm{d}\phi/\mathrm{d}t = \beta(\tau)S(t)I(t)/N$. Assuming data is reported over intervals represented by a set of discretized points in time $\mathcal{R}$, and using the symbol $\Phi^*(t_i)$ to refer to the number of reported cases in the interval t_{i-1} to t_i, the objective function can then be written as,

$$\min \sum_{i \in \mathcal{R}: i>1} \left[(\phi(t_i) - \phi(t_{i-1})) - \Phi^*(t_i) \right]^2 . \tag{3.6}$$

where $\phi(t_i) - \phi(t_{i-1})$ is the number of new cases simulated for the time interval of (t_{i-1}, t_i).

The primary goal of this research is the estimation of the seasonal transmission parameter, $\beta(\tau)$. Since $\beta(\tau)$ should have a yearly periodicity, one approach would be to parameterize $\beta(\tau)$ using a trigonometric series. This could be problematic since we expect $\beta(\tau)$ to have sharp changes during the year corresponding to the on/off pattern of school term holidays. The plan of this research is to completely discretize $\beta(\tau)$ along the time domain and then require that $\beta(\tau)$ be periodic (e.g. forcing $\beta(\tau) = \beta(\tau - 1 \text{ year})$ in the formulation. Because of the large number of degrees of freedom that arise from discretizing $\beta(\tau)$, regularization may be necessary to produce a well-posed estimation problem. This will be achieved by adding another term to the objective function, $\frac{1}{\rho}R(\beta)$. The regularization parameter ρ will be determined by using the standard the L-curve method [30]. There are several different forms which we can use for the regularization function of $\beta(\tau)$, most notably: Tikhonov, Higher-Order Tikhonov, Total Variation, etc. [30]. The main focus will center on total variation since the seasonal transmission parameter is believed to be correlated with on/off holidays of the school year and total variation is known to allow for sharp changes in the profile. This is believed to give a better representation of $\beta(\tau)$ than

restricting it to a specific functional form like a polynomial or sine curve.

The final estimation formulation that will be used in this research is given as,

$$\min \sum_{i \in \mathcal{R}: i>1} \left[(\phi(t_i) - \phi(t_{i-1})) - \Phi^*(t_i)\right]^2 + \frac{1}{\rho}\mathrm{Reg}(\beta) \tag{3.7}$$

$$\text{s.t.} \quad \frac{\mathrm{d}S}{\mathrm{d}t} = \frac{-\beta(\tau)S(t)I(t)}{N} + \mu N \tag{3.8}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta(\tau)S(t)I(t)}{N} - \gamma I(t) \tag{3.9}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I(t) - \mu N \tag{3.10}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{\beta(\tau)S(t)I(t)}{N}. \tag{3.11}$$

B. Regularization

Our system is nonlinear; however, to simplify the discussion here we adopt the notation used in Aster et al. [30] for linear systems. Here, G represents the model, m the parameters, d the data, α the regularization parameter and $\|\cdot\|_2$ the euclidean norm. One of the more common regularization techniques is Thikhonov regularization. In Tikhonov regularization, the goal is to choose from the solutions of $\|Gm-d\|_2 \leq \delta$ the one which also minimizes $\|m\|_2$. The reasons for selecting the minimum norm from among those solutions can be adequately explained by the fact that any nonzero features that appear in the regularized solution increase the norm of m. Those nonzero features are necessary for the model to fit the data properly and by taking the minimum of $\|m\|_2$, the assurance of not having any unneeded features in the regularization is achieved. The following setup is the general form of the Tikhonov, or zeroth-order,

$$\begin{aligned} \min \quad & \|m\|_2 \\ & \|Gm - d\|_2 \ \leq \delta \end{aligned} \tag{3.12}$$

It should be noted, when the value of δ increases, the set of feasible solutions to the model expands, therefore causing the minimum of $\|m\|_2$ to decrease. If a plot is produced comparing these two values, minimum norm of m versus δ, the resulting shape would appear as the graph of the reciprocal function within the first quadrant. Inversely, this same plot can be produced by an alternate formulation of Tikhonov,

$$\begin{aligned} \min \|Gm - d\|_2 & \\ \|m\|_2 & \leq \epsilon. \end{aligned} \tag{3.13}$$

A third formulation involves taking both the residuals and the parameter norms within the objective function and applying a weight to the parameter norm called the regularization parameter, α.

$$\min \|Gm - d\|_2^2 + \alpha^2 \|m\|_2^2 \tag{3.14}$$

The euclidean norm has been modified to a least squares form and the parameter α causes a dampening in the problem. Hence, the name for this scheme is the dampened least squares problem. These three schemes can be shown to yield the same solution with a proper choice of δ, ϵ, and α [30]. When plotted with an appropriate scale, the curve of the optimum values of $\|Gm - d\|_2$ versus $\|m\|_2$ closely resembles the shape of the letter 'L.' This is caused by the fact that $\|Gm - d\|_2$ strictly increases as a function of the parameter α, and $\|m\|_2$ strictly decreases. Due to its characteristic shape this curve is called the L-curve [31].

1. Higher-order Tikhonov Regularization

With the basic Tikhonov technique, only the euclidean norm of the parameter is minimized. It is not always favorable to minimize the parameter directly and sometimes necessary to accomplish regularization through a different measure of m. With

higher-order Tikhonov, it is desirable to regularize the first or second derivative of m. Here the roughening matrix L is introduced. Within the model the time derivative of $m(t)$ is estimated by Lm. Typically Lm is defined to be a finite-difference derivative approximation. Thus, since it is minimized, the solution of m favored will be relatively flat in the case of regularizing the first derivative. Higher-order Tikhonov can be formulated as a dampened least squares problem,

$$\min \|Gm - d\|_2^2 + \alpha^2 \|Lm\|_2^2 \tag{3.15}$$

where in the case of first-order Tikhonov regularization, the roughening matrix L is defined as,

$$L = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \cdots & & \vdots \\ 0 & \cdots & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}. \tag{3.16}$$

With second-order Tikhonov regularization, Lm should appropriately estimate the second derivative of m. One possible L is defined as

$$L = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & & \cdots & & \vdots & \\ 0 & \cdots & 1 & -2 & 1 & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}, \tag{3.17}$$

where the solutions of m which are rough in the second derivative sense are penalized.

Note that here the roughening matrices are appropriate only for a one-dimensional model; however, if the model is two- or three-dimestional, these roughening matrices

are not appropriate. In those cases a finite difference to the Laplacian operator is used [30].

2. Other Regularization Techniques

Bound constraint methods are appropriate for situations where bounds exits on the maximum or minimum values of the model parameters. For example, in the case of the transmission parameter β someone who is infected cannot take away the infection from another individual, therefore it is a value that is inherently non-negative. The formulation of this sort of least squares problem can take the structure of,

$$\begin{aligned} \min \|Gm - d\|_2 & \\ m &\geq 0 \end{aligned} \tag{3.18}$$

where the bound makes sure that all values of m must remain nonnegative. This problem formulation is commonly known as the nonnegative least squares problem [32].

Another option for the bounded constraints is to constrain the m parameter in both directions by specifying both a set of upper bounds u and a set of lower bounds l. Again using the example of β this can be seen as a maximum on the number of people that the infected individual comes into contact within the designated time frame. He or she cannot infect more people than he or she comes into contact, therefore it is an upper bound. This problem formulation is commonly called the bounded variables least squares (BVLS) problem,

$$\begin{aligned} \min \|Gm - d\|_2 & \\ m &\leq u \\ m &\geq l \end{aligned} \tag{3.19}$$

There have been several algorithms for solving the BVLS problem which allow for Tikhonov regularization with bounds [32].

The second major alternative to Tikhonov regularization is maximum entropy regularization. The maximum entropy regularization approach can be seen in its general form as,

$$\begin{aligned} \max \quad & -\sum_{i=1}^{n} m_i \ln(w_i m_i) \\ & \|Gm - d\|_2 \leq \delta \\ & m \geq 0 \end{aligned} \tag{3.20}$$

According to Aster et al. [30], the term "Maximum Entropy" comes from a Baysian approach to selecting a prior probability distribution subject to the constraint $\sum_{i=1}^{n} p_i = 1$ that maximizes the term $-\sum_{i=1}^{n} p_i \ln p_i$, which has the same form as entropy. The limitation of this formulation is that m_i must always be strictly positive for all i so that the logarithm is always defined. This method is widely applied in astronomical image processing [33].

The final alternative regularization technique discussed here is total variation. This functional form is most appropriate if the function to be regularized is believed to exhibit discontinuous jumps. The formulation is as follows,

$$\mathrm{TV}(m) = \sum_{i=1}^{n-1} |m_{i+1} - m_i| \tag{3.21}$$

$$= \|Lm\|_1 \tag{3.22}$$

where $\|\cdot\|_1$ is the 1-norm. Here, L is typically the same as that defined in (3.16). This differs from the basic Tikhonov techniques because where they penalize discontinuities and favor smoothness, the total variation penalizes all magnitude of change equally. The formulation term for total variation can be added into the objective function in

place of the other Tikhonov regularizations to arrive at the objective form,

$$\min \|Gm - d\|_2^2 + \alpha \|Lm\|_1 \tag{3.23}$$

Solution of this problem requires some manipulation since it is not in the typical form of a least squares problem. Furthermore, in our case we have a constrained nonlinear least squares problem. Reformulation of $\|Lm\|_1$ will be discussed later.

Motivation for choosing one regularization strategy over the other is mainly influenced by the desired or expected shape of the parameter profile. With childhood diseases, the transmission parameter is believed to have discontinuous jumps between high and low periods of transmission activity related to school term holidays. Therefore, the total variation regularization is chosen for this work.

C. Data

This research uses both measles and chickenpox case count data, as well as birth rate and population data from London, New York City, and Bangkok, Thailand. This data is formatted and converted into appropriate *.dat files usable in AMPL [34]. These data have been selected for the following reasons:

- London and New York City measles have been studied by other authors and students within our group, making them a good test bed for the new approach.
- These locations have different reporting intervals, and will provide insight as to the reliability of this method when the reporting interval is longer than the generation time.
- Bangkok has very different school holiday schedules from the other two cities. This will allow us to determine if there is a strong correlation between school terms and seasonal transmission as estimated from the continuous models.

Table I. Data used for New York City, London, and Bangkok

	Years	Reporting Interval	Tot. Time	Reporting Fraction
New York Measles	1947-1963	30 days	16 years	1/10
London Measles	1944-1958	14 days	14 years	1/2.3
Bangkok Measles	1975-1985	30 days	10 years	1/100 - 4/100
New York Chickenpox	1944-1964	30 days	20 years	6/100 - 4/100
Bangkok Chickenpox	1985-1995	30 days	10 years	2/100 - 1/100

Details of these data sets are shown in Table I. The Bangkok data set is missing records for the year 1979 so the objective is weighted to not include these points. Also, within these data sets there is severe under reporting. For example, in the case of New York City measles, approximately 1 in every 10 cases was reported.

In the discrete time models the data must be modified to fit the biweekly intervals required by the model. With regards to the continuous time model the data is unchanged and can be used in its original form with the reporting interval unaltered.

CHAPTER IV

SOLUTION APPROACH

The estimation formulation given in equations (3.7) to (3.11) is a nonlinear optimization problem with differential constraints. In Chapter II different techniques for solving this class of optimization problem was discussed. Here, the simultaneous approach is adopted for efficiency even though there is additional complexity. This section outlines the complete solution approach, beginning with the Radau collocation for discretization of the differential constraints. Next we show an effective reformulation of the non-differentiable regularization term, and present the full large-scale nonlinear programming formulation used in this study. As with all non-convex, nonlinear programming problems, an effective strategy for obtaining initial guesses is critical. This initialization strategy is discussed, followed by the interior-point solution approach used by IPOPT [35].

A. Problem Discretization and Formulation

1. Radau Collocation on Finite Elements

In order to transcribe the infinite dimensional dynamic problem to a finite dimensional problem, this approach uses orthogonal collocation to discretize the differential and algebraic equations. Radau collocation is used since it has almost the same order of accuracy as Gauss collocation, is also A-stable and has stiff decay [23]. It has been shown that this integration method is equivalent to a fully implicit Runge-Kutta scheme [36]. Consider the general differential equation,

$$\dot{x} = f(x, u), \qquad x(0) = x_0 \tag{4.1}$$

where x denotes the state variables, and u the control variables. The optimization horizon is divided into N_e finite elements which each contain N_c collocation points, $c_i \in [0, 1]$. Radau collocation requires that the final collocation point be located at the end of each finite element, $c_{N_c} = 1$. The time locations of each collocation point can be defined as $t_{ij} = t_i + h_i c_j$ where h_i is the length of each finite element and t_i is the beginning time for each finite element. In order to attain the collocation equations, introduction of the state variable and control variable estimation equations is necessary. The state variable estimation equation is discretized as,

$$x_{N_c+1}(t) = \sum_{k=0}^{N_c} x_{ik} L_k^{N_c+1}\left(\frac{t - t_{i-1}}{h_i}\right) \qquad t \in [t_{i-1}, t_i], \tag{4.2}$$

where $L_k^{N_c+1}$ denotes the Lagrange interpolation polynomials of order N_c and x_{ij} are parameters. The order of the Lagrange polynomials is increased one degree by adding the point $c_0 = 0$ to the set of collocation points within each finite element. The control variables are discretized using the approximation,

$$u_{N_c}(t) = \sum_{k=0}^{N_c} u_{ik} L_k^{N_c}\left(\frac{t - t_{i-1}}{h_i}\right) \qquad t \in [t_{i-1}, t_i], \tag{4.3}$$

where u_{ij} are parameters. It is now possible to write the collocation equations,

$$\sum_{k=0}^{N_c} x_{ik} \dot{L}_k^{(N_c+1)}(c_j) = h_i f(x_{ij}, u_{ij}), \tag{4.4}$$

defined $\forall : i \in \{1..N_e\}$ and $j \in \{1..N_c\}$. In order to enforce continuity between the finite elements the constraint,

$$x_{i-1,N_c} = x_{i,0} \tag{4.5}$$

is also included. The variables x and u and the original differential formulation, (4.1) are now replaced by the parameters x_{ij} and u_{ij} and the equality constraints (4.4) and (4.5).

This research employs a slightly different formulation for the Radau collocation equations as given by Bader and Ascher [37], that of the monomial-basis. The difference arises in the estimation of x. Instead of approximating x directly, $\dot{x}$ is approximated by the Lagrange polynomials resulting in the state variable being an approximation of the integral of the interpolation polynomials. However, despite this slight difference the monomial-basis representation has also been shown to correspond to a Runge-Kutta scheme [36]. Therefore the monomial-basis discretization scheme for the above dynamic state variable would be,

$$x(t) = x_{i,0} + h_i \sum_{k=0}^{N_c} L_k^{N_c} \left(\frac{t - t_{i-1}}{h_i} \right) \dot{x}_{ik} \tag{4.6}$$

remembering that $x_{i,0}$ is the equivalent to the calculated parameter x_{i-1,N_c} of the previous finite element, and $\dot{x}_{ik}$ is the value of the first derivative in element i at collocation point k. A similar scheme was used in the parameter estimation for low-density polyethylene tubular reactors and can be found in Zavala and Bielger [38].

For the model used (3.7), the discretized form of the differential equations are,

$$S_{i,j} = S_{i-1,3} + h_i \sum_{k=0}^{N_c} a_{k,j} \dot{S}_{i,k} \tag{4.7}$$

$$I_{i,j} = I_{i-1,3} + h_i \sum_{k=0}^{N_c} a_{k,j} \dot{I}_{i,k} \tag{4.8}$$

$$\phi_{i,j} = \phi_{i-1,3} + h_i \sum_{k=0}^{N_c} a_{k,j} \dot{\phi}_{i,k}, \tag{4.9}$$

subject to the equality constraints,

$$\dot{S}_{i,j} = -\left[\beta(\tau_{i-1}) + c_j(\beta(\tau_i) - \beta(\tau_{i-1}))\right] \frac{I_{i,j} S_{i,j}}{N} + \mu_{i,j} N \tag{4.10}$$

$$\dot{I}_{i,j} = \left[\beta(\tau_{i-1}) + c_j(\beta(\tau_i) - \beta(\tau_{i-1}))\right] \frac{I_{i,j} S_{i,j}}{N} - \gamma I_{i,j} \tag{4.11}$$

$$\dot{\phi}_{i,j} = \left[\beta(\tau_{i-1}) + c_j(\beta(\tau_i) - \beta(\tau_{i-1}))\right] \frac{I_{i,j} S_{i,j}}{N}. \tag{4.12}$$

The number of collocation points for each finite element has been chosen as $N_c = 3$, and the basis used for the calculation of the Lagrange interpolation polynomial is the monomial basis. The values of $\beta(\tau)$ are discretized only at the finite element boundaries and are assumed to be linear within each finite element. The subscripts are always in order of: finite element number (i), collocation point (j or k). The matrix $a_{k,j}$ in equation (4.13) is the evaluation of the Lagrange interpolation with the monomial basis for the set of collocation points, $c_j = \{\ 0.155051,\ 0.644949,\ 1\}$ on the normalized domain (0,1].

$$a_{k,j} = \begin{bmatrix} 1.968155E{-}01 & 3.944243E{-}01 & 3.764031E{-}01 \\ -6.553543E{-}02 & 2.920734E{-}01 & 5.124858E{-}01 \\ 2.377097E{-}02 & -4.154875E{-}02 & 1.111111E{-}01 \end{bmatrix} \tag{4.13}$$

2. Total Variation Regularization

The parameter function $\beta(\tau)$ has been shown previously in the literature to have a strong correlation with school holiday schedules [2]. Therefore the functional form of the estimated profile must be allowed to be discontinuous and a Tikhonov regularization technique is inappropriate. With this research a total variation approach is employed so as to try and capture the school holiday schedule more clearly.

The foundational form of the total variation regularization term will be,

$$\begin{aligned} \mathrm{TV}(\beta(\tau)) &= \sum_{i=1}^{n-1} |\beta(\tau_{i+1}) - \beta(\tau_i)| \\ &= \|\beta(\tau)\|_1 \end{aligned} \tag{4.14}$$

Unfortunately the absolute value function has a discontinuous first derivative at the value of zero, and there is a better way of formulating this regularization as a well-defined problem achieving the the same solution. If two new variables are introduced

into the problem, let us call them $\beta^{+\Delta}$ and $\beta^{-\Delta}$, then two equality constraints can be added into the model with additional non-negativity constraints,

$$\beta_i^{\Delta} = \beta(\tau_{i+1}) - \beta(\tau_i); \tag{4.15}$$

$$\beta_i^{\Delta} - \beta_i^{+\Delta} + \beta_i^{-\Delta} = 0; \tag{4.16}$$

$$\beta_i^{+\Delta}, \beta_i^{-\Delta} \geq 0; \tag{4.17}$$

then the objective function term $\|\beta(\tau)\|_1$ can be written as $\sum_{i=1}^{nfe}(\beta_i^{+\Delta} + \beta_i^{-\Delta})$. The non-differentiability of the absolute value function is avoided while still keeping the same solution method intact. β_i^{Δ} is unnecessary since it is only the definition of the difference between two consecutive $\beta(\tau)$ terms, but is used for clarity and ease of explanation. If $\beta(\tau_{i+1}) \geq \beta(\tau_i)$ then $\beta_i^{\Delta} \geq 0$, and similarly if $\beta(\tau_{i+1}) \leq \beta(\tau_i)$ then $\beta^{\Delta} \leq 0$. Therefore the difference variable, β_i^{Δ}, can be either positive or negative. The second constraint (4.16), forces all of the positive values of β_i^{Δ} to be assigned to $\beta_i^{+\Delta}$ and forces all of the negative values of β_i^{Δ} to be assigned to $\beta_i^{-\Delta}$ as a positive value, hence when they are summed together in the regularization term of the objective function the absolute value is preserved.

3. Full Estimation Formulation

Below is the full formulation of the nonlinear programming problem used for the estimation.

$$\min \sum_{i=1}^{\mathrm{N_{fe}}} \left[(\phi(t_i) - \phi(t_{i-1})) - \Phi^*(t_i)\right]^2 + \frac{1}{\rho} \sum_{i=1}^{\mathrm{N_{ye}}} (\beta_i^{+\Delta} + \beta_i^{-\Delta}) \tag{4.18}$$

$$\text{s.t.} \quad \dot{S}_{ij} = -[\beta(\tau_{i-1}) + c_j(\beta(\tau_i) - \beta(\tau_{i-1})]\frac{S_{ij}I_{ij}}{N} + \mu_i \tag{4.19}$$

$$\dot{I}_{ij} = [\beta(\tau_{i-1}) + c_j(\beta(\tau_i) - \beta(\tau_{i-1})]\frac{S_{ij}I_{ij}}{N} - \gamma I_{ij} \tag{4.20}$$

$$\dot{\phi}_{ij} = [\beta(\tau_{i-1}) + c_j(\beta(\tau_i) - \beta(\tau_{i-1})]\frac{S_{ij}I_{ij}}{N} \tag{4.21}$$

$$S_{ij} = S_{i-1,cp} + h\sum_{k=1}^{\mathrm{N_{cp}}} a_{kj}\dot{S}_{ik} \tag{4.22}$$

$$I_{ij} = I_{i-1,cp} + h\sum_{k=1}^{\mathrm{N_{cp}}} a_{kj}\dot{I}_{ik} \tag{4.23}$$

$$\phi_{ij} = \phi_{i-1,cp} + h\sum_{k=1}^{\mathrm{N_{cp}}} a_{kj}\dot{\phi}_{ik} \tag{4.24}$$

$$S_{0j} = S_0 + h\sum_{k=1}^{\mathrm{N_{cp}}} a_{kj}\dot{S}_{1k} \tag{4.25}$$

$$I_{0j} = I_0 + h\sum_{k=1}^{\mathrm{N_{cp}}} a_{kj}\dot{I}_{1k} \tag{4.26}$$

$$\phi_{0j} = h\sum_{k=1}^{\mathrm{N_{cp}}} a_{kj}\dot{\phi}_{1k} \tag{4.27}$$

$$\beta_i^{\Delta} = \beta(\tau_{i+1}) - \beta(\tau_i) \tag{4.28}$$

$$\beta_i^{\Delta} - \beta_i^{+\Delta} + \beta_i^{-\Delta} = 0 \tag{4.29}$$

$$\beta_i^{+\Delta}, \beta_i^{-\Delta} \geq 0 \tag{4.30}$$

$$S_{ij}, I_{ij}, \phi_{ij} \geq 0 \tag{4.31}$$

Here i is the index for each finite element and j is the index for each collocation point. The variable $\mathrm{N_{fe}}$ is the total number of finite elements used within the estimation, $\mathrm{N_{ye}}$ is the number of finite elements contained within one year, and $\mathrm{N_{cp}}$ is the number of collocation points used for each finite element. Equations (4.19)

through (4.21) are the derivative functions. Equations (4.22) through (4.24) are the collocation equations with a_{kj} being the various values of the collocation matrix and h being the timestep.

B. Initialization of the NLP

For any nonlinear estimation it is very important to initialize the problem appropriately. Initial conditions I_0 and S_0 are assumed along with reasonable values for the model parameters. An Explicit Euler integration is used to initialize the state profiles for I, S and ϕ. The mesh for the initialization is determined by the collocation points. There are three points per finite element n each notated as

$$\begin{aligned} t_{n,1} &= t_{n-1} + 0.155(t_n - t_{n-1}) \\ t_{n,2} &= t_{n-1} + 0.645(t_n - t_{n-1}) \\ t_{n,3} &= t_{n-1} + 1.000(t_n - t_{n-1}) = t_n \end{aligned} \tag{4.32}$$

with t_{n-1} denoting the beginning of the interval and t_n denoting the ending.

C. Primal-dual Interior-Point Algorithm for Large Scale NLP

The optimization package used to solve this model formulation, IPOPT [35], is an implementation of a primal-dual interior-point algorithm with a filter line-search method. An in depth description of the particulars of the algorithm along with comparisons against other interior-point codes for nonlinear programming can be found in [35]. Briefly, this is an interior-point or barrier method for nonlinear programming. The basic approach of a the method is explained with respect to the problem formulation below.

$$\min_{x \in R^n} \quad f(x) \tag{4.33}$$

$$\text{s.t.} \quad c(x) = 0 \tag{4.34}$$

$$x \geq 0$$

The variable bounds are moved to the objective in the form of a barrier function as shown in equation (4.35). This subproblem is approximately solved for each of a sequence of barrier parameters μ converging to zero.

$$\begin{aligned} \min_{x \in R^n} \quad \varphi_\mu(x) &= f(x) - \mu \sum_{i=1}^{n} \ln(x^{(i)}) \\ \text{s.t.} \quad c(x) &= 0 \end{aligned} \tag{4.35}$$

Equivalently, this method can be rewritten as a homotopy method to the primal-dual equations (4.36) with the homotopy parameter, μ being driven to zero. It is important to notice that when $\mu = 0$ this homotopy method, coupled with the non-negative constraints, becomes the Karush-Kuhn-Tucker conditions for the original problem (4.33).

$$\nabla f(x) + \nabla c(x)\lambda - z = 0 \tag{4.36}$$

$$c(x) = 0 \tag{4.37}$$

$$XZe - \mu e = 0 \tag{4.38}$$

Here, $\lambda \in R^m$ and $z \in R^n$ are the Lagrangian multipliers for the equality constraints (4.34) and the nonnegative bound constraints, $x \geq 0$, respectively. To see a more in depth discussion of the algorithm refer to Wächter and Biegler [35].

CHAPTER V

RESULTS

Prior to performing the estimation using actual case data, numerous simulation studies were performed with varying patterns for β. Using simulated data aggregated over each finite element for Φ^*, the estimation formulation was solved using IPOPT. The formulation was able to correctly estimate both constant values of the transmission parameter as well as sharply changing seasonal profiles. In addition, the state variable profiles were also effectively reproduced. These studies were performed using different finite element sizes to ensure the formulation would handle different discretizations within the time domain. These simulation-estimation results verified the appropriateness of the entire procedure for use with actual reported case data.

A. Measles

The data for cases of measles was obtained and analysed for the cities of New York, London, and Bangkok.

Table II. Overall values for the measles model estimations of New York City, London, and Bangkok

	ρ	FE[a]	Rep. Fac.	Tot. Var.	Tot. Const.	Obj. Val.	CPU [b]
New York Measles	1	60	0.1083	8887	8764	1.817e+01	15.41
London Measles	2	78	0.43627	10147	9988	5.284e+01	2.52
Bangkok Measles	100	36	.012-.038	3391	3316	3.625e-01	1.32

[a] Finite elements per year.
[b] Times are reported in seconds.

1. New York Measles 1947-1963

The selected timespan under consideration for New York City measles is 1947-1963. This occurs during the pre-vaccination era for measles in this part of the United States. During this time, measles case counts are reported monthly. The reporting factor for New York during this time was assumed to be constant at a value of approximately 1 in 10 cases (see Table II). A total of 60 finite elements was selected per year, allowing 5 finite elements per reporting interval. Prior to presentation of the final results, the estimation was run multiple times with different values of the regularization parameter ρ. Figure 1 shows the resulting L-curve. This curve has a pronounced corner corresponding to a value of $\rho \approx 1$, and this value was used for all results presented in this section.

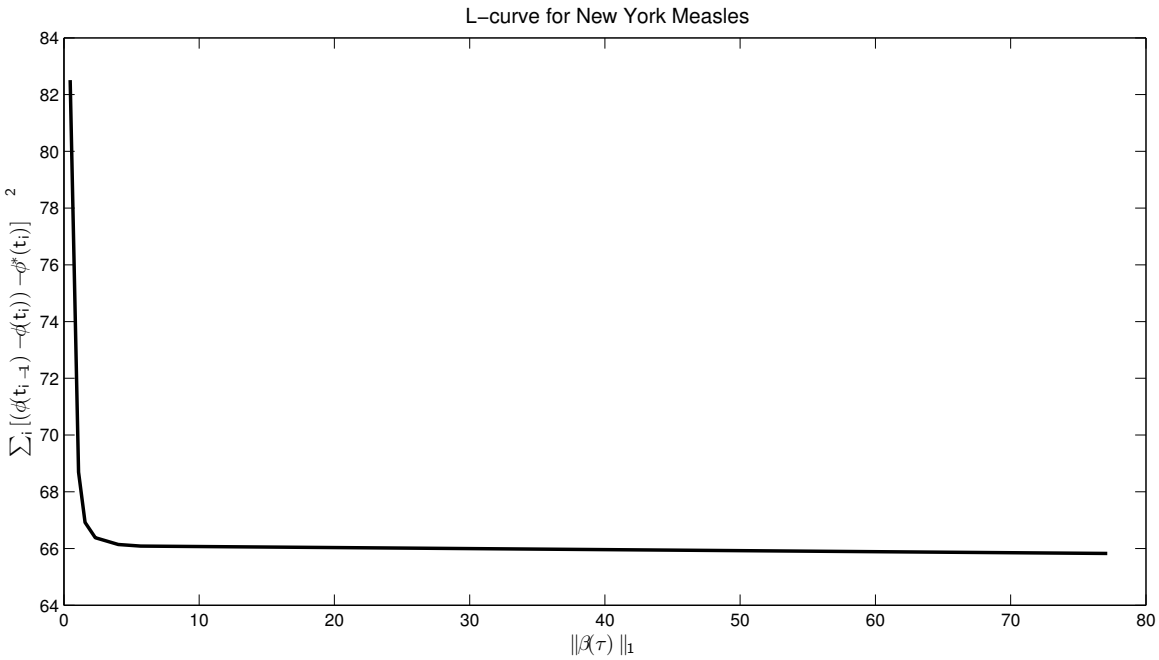

Fig. 1. The L-curve for New York measles used to determine $\rho^{\mathrm{M}}_{\mathrm{NYC}}$.

a. Seasonal $\beta(\tau)$ Infectious Cases Results

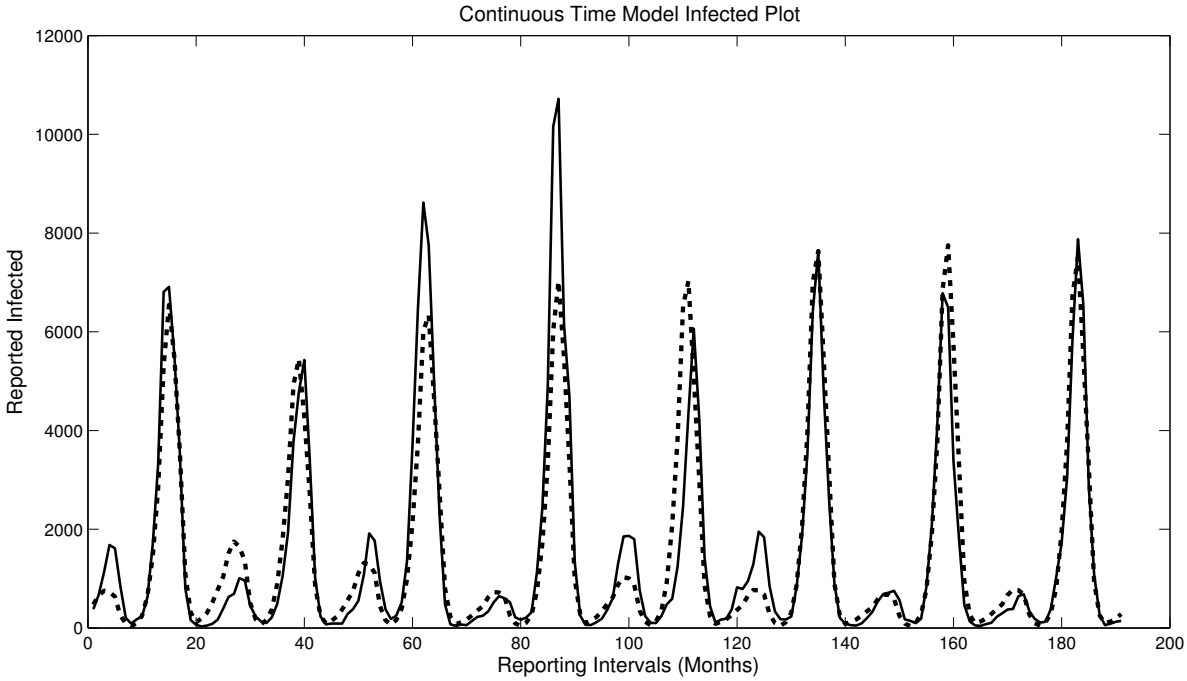

Fig. 2. Infectious cases results of the estimation for New York measles 1947-1963; (- -) represents the estimation and (–) represents the historical data.

Figure 2 shows the profiles for the actual and estimated reported cases. The estimated model is able to capture the biennial major-minor outbreak pattern that is characteristic of measles in this era. The estimated profile for the seasonal transmission parameter can be seen in Figure 3. A very pronounced on/off pattern is observed in the data that will be discussed further in the conclusions. The estimation was also performed using different numbers of finite elements. Table III shows a comparison of the solution times and objective values for different numbers of finite elements. Work has also been done to compare the continuous estimation results against previous results using a discrete-time model. These results can be found in Abbott III et al.

[39].

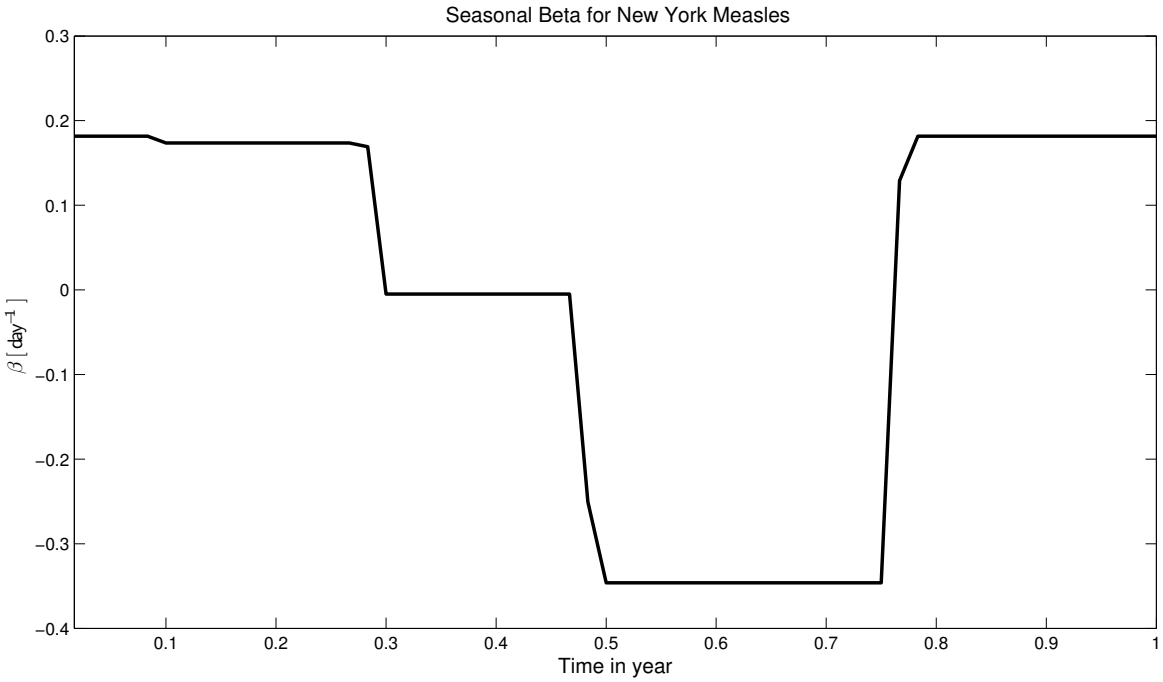

Fig. 3. $\beta(\tau)$ results for New York measles 1947-1963 (graph is normalized about its mean).

The nonlinear programming formulation with 60 finite elements has over 8000 variables and constraints, however solution is very efficient, and the estimation formulation itself only required 15.41 seconds to complete.

Table III. Number of finite elements per year versus CPU time for New York City measles model

FE per year	CPU time (sec)[a]	Obj. Val.
12	1.2	1.8242e+01
24	3.464	1.8186e+01
36	5.56	2.8035e+01
48	8.992	1.8172e+01
60	15.41	1.8170e+01

[a] All problems were solved on a 3.0 GHz Intel Xeon processor.

b. Estimation Without Periodicity Constraint

In order to further verify the seasonality in the transmission parameter, the estimation was solved over the entire time domain with no assumption about periodicity of β. Using 36 finite elements per year (chosen for a shorter calculation time) the formulation contained 7,132 total variables, and 6,481 equality constraints, and solved in a CPU time of 30.762 seconds. The resulting $\beta(t)$ function from the entire time domain was then plotted with each year superimposed, as can be seen in Figure 4. It is remarkable that such a consistent yearly pattern emerges with no constraint on the form of $\beta(\tau)$ other than the total variation regularization.

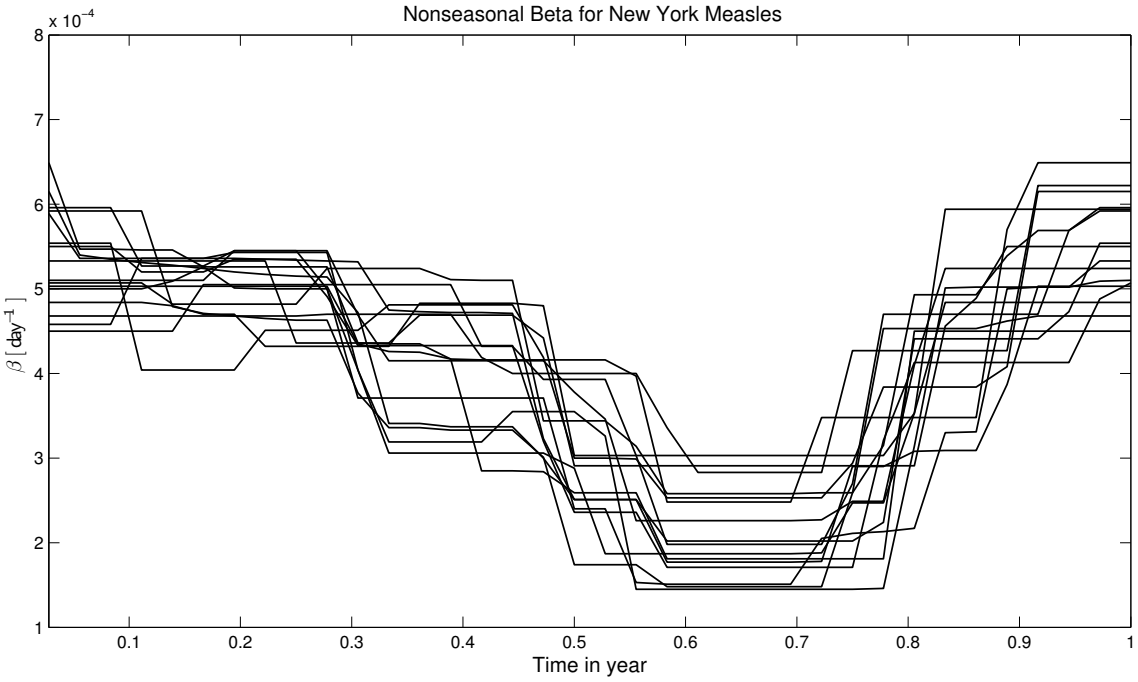

Fig. 4. New York measles 1947-1963 nonseasonal $\beta(t)$ estimation results (plotted as years superimposed over each other).

2. London Measles 1944-1958

London measles from the pre-vaccination era has been the subject of much research in the field of childhood infectious disease modelling [40, 19, 2, 18]. The time period of 1944-1958 was chosen to give a substantial time frame within the pre-vaccination era following the end of World War II.

The L-curve method was used to determine the regularization parameter, but unlike the L-curve for New York City the "corner" is not very pronounced (see Figure 5). Table IV has been included to show the calculated values that make up the curve of the graph. The value chosen for this estimation was $\rho = 5$, although it can be argued that the values 4 through 6 are also valid. Selecting 3 finite elements per reporting interval gives a total of 78 finite elements per year. The reporting interval

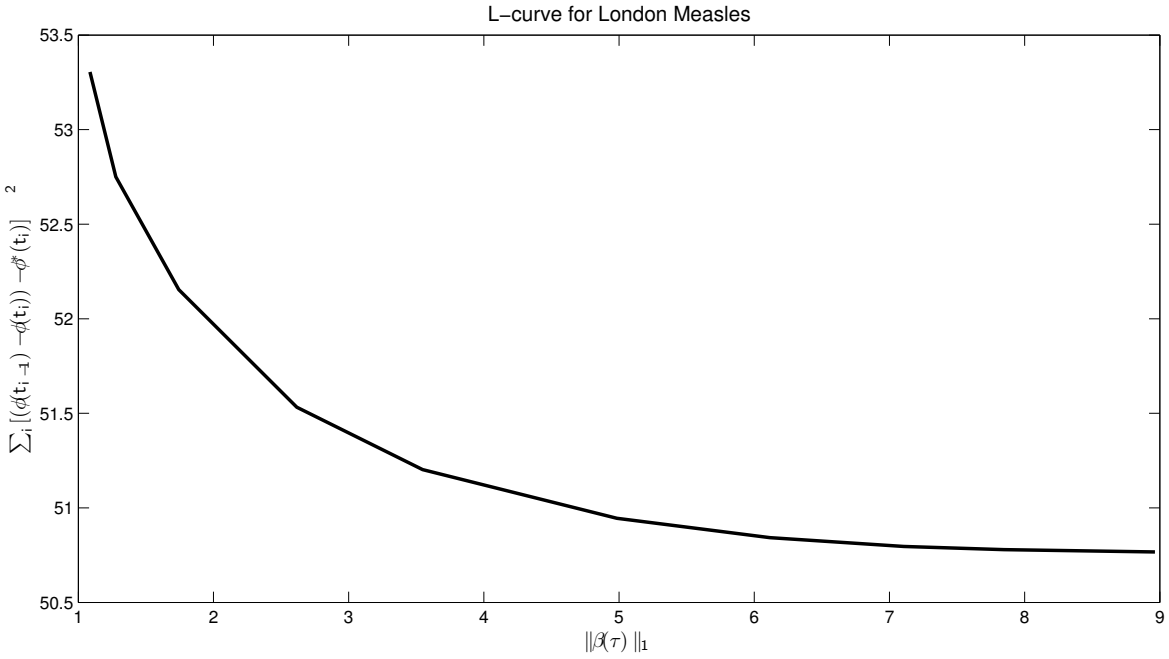

Fig. 5. The L-curve for London measles used to determine ρ_{Lon}.

for our London data during this time was two weeks. The biweek reporting interval is fortunate for previous discrete time methods, since it allows the reporting and the serial intervals to coincide. However, this is not necessary for our continuous formulation.

Table IV. Data calculated for London L-curve graph

ρ	16	8	4	2	1	0.5	0.25
$\|\beta(\tau)\|_1$	6.12	4.98	3.55	2.62	1.74	1.28	1.09
Obj. Val.	50.84	50.94	51.20	51.53	52.15	52.75	53.31

a. Seasonal $\beta(\tau)$ Infectious Cases Results

The estimated seasonal transmission parameter $\beta(\tau)$ for London can be seen in Figure 6. The data set is not as ideal as the New York data. Measles has a biennial major-minor pattern near the end of the data series. However, near the beginning the pattern appears annual. This is believed to be related to the birthrate. The reporting factor was held constant with the value of 0.43627. The estimation formulation contained a total of 10,147 variables and 9,988 equality constraints, and solved in only 2.52 seconds.

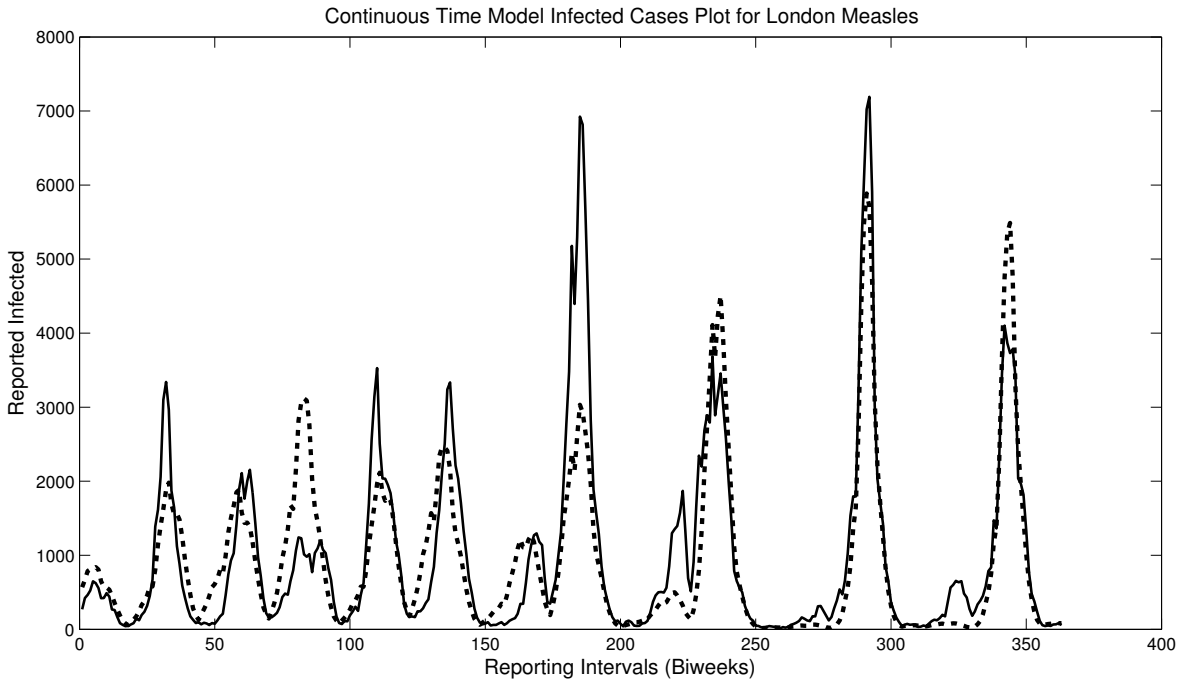

Fig. 6. Infectious cases results of the estimation for London measles 1944-1958; (- -) represents the estimation and (–) represents the historical data.

The estimated seasonal $\beta(\tau)$ profile can be seen in Figure 7. These results also show a prominent pattern that is further discussed in the conclusions.

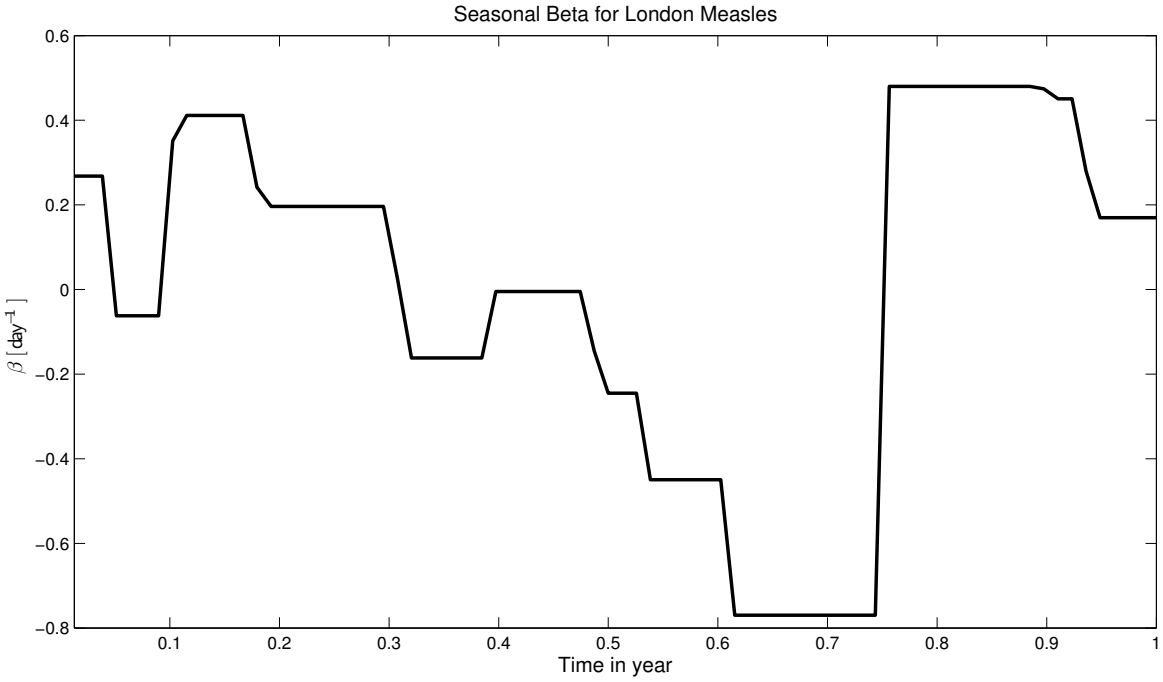

Fig. 7. $\beta(\tau)$ results for London measles 1944-1958.

b. Estimation Without Periodicity Constraint

Again the estimation was performed for London without the periodicity requirement (see Figure 8). The estimated transmission parameter profile has been plotted over the time of one year with each year being superimposed over the other for ease of comparison between years. The profiles for London are not as consistent as those for New York, however a definite trough can be seen right before the beginning of the fourth quarter of the year, close to the trough seen in the seasonal estimation of $\beta(\tau)$.

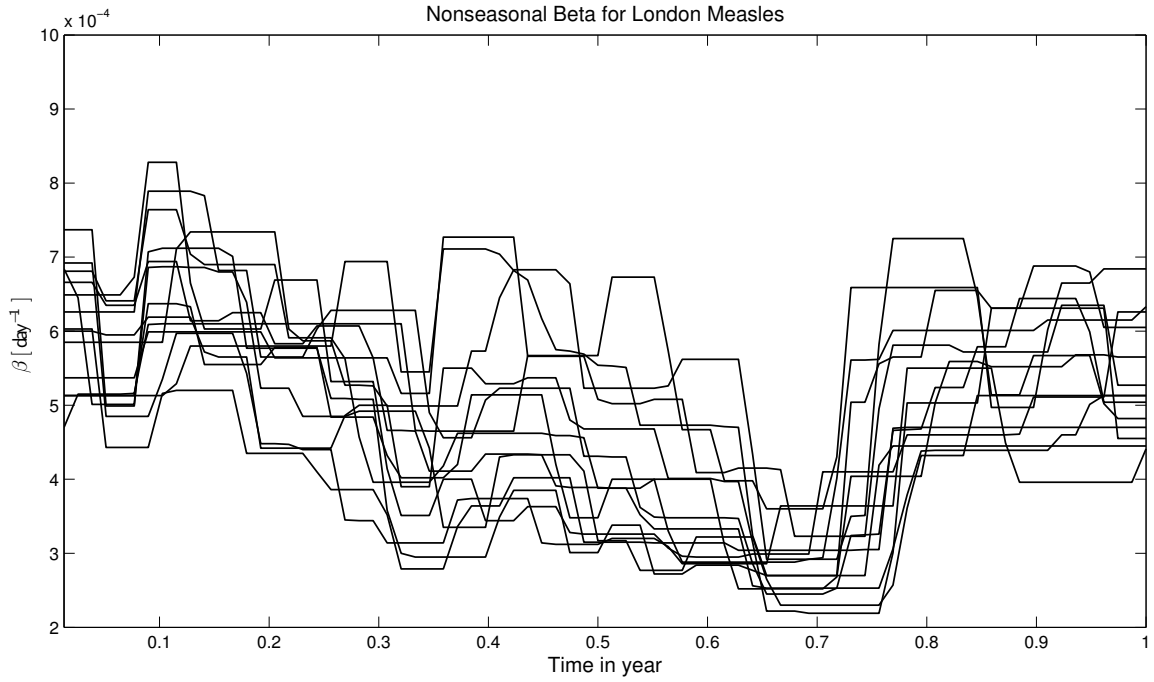

Fig. 8. London measles 1944-1958 nonseasonal $\beta(t)$ estimation results (plotted as years superimposed over each other).

3. Bangkok Measles 1975-1985

While there are numerous publications using the London and New York data sets, however, virtually no published research has been done using the Bangkok measles data studied here. This offers a very unique opportunity to see if measles seasonality is indeed dependent on school vacation periods.

The L-curve plot was generated for the Bangkok Measles recorded cases spanning the years 1975-1985 (Fig. 9). Although the "corner" of this graph is quite pronounced the values of ρ span a wide range of values from 1,000 to 100 (See Table V). The value of ρ selected for this estimation is 100.

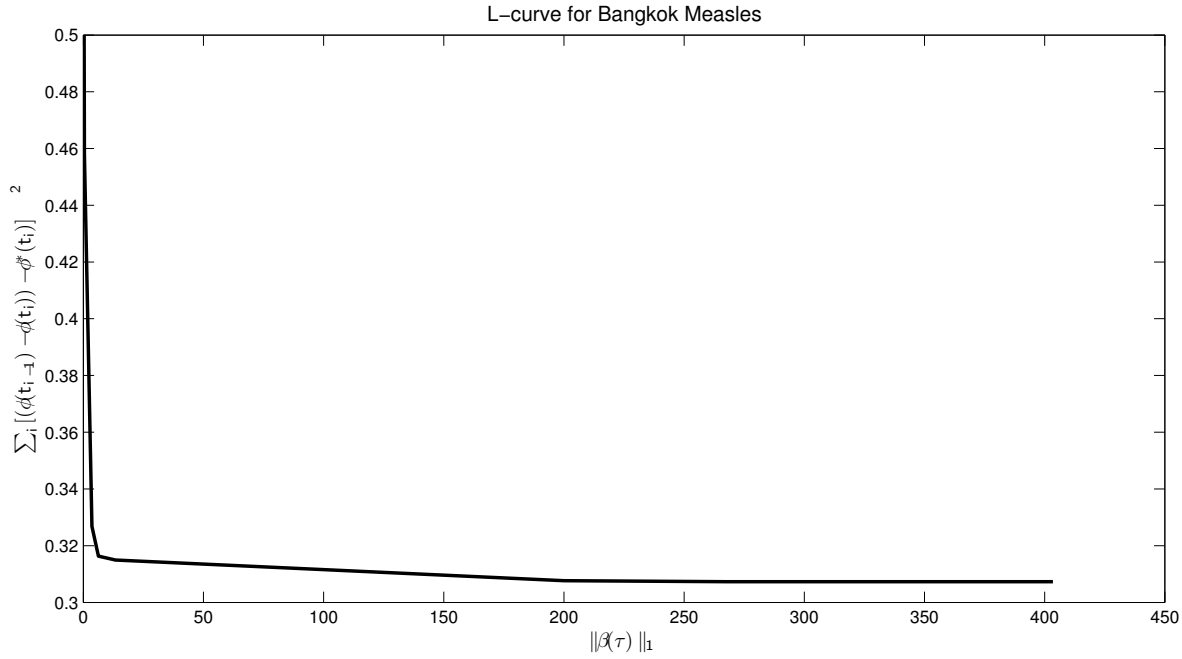

Fig. 9. The L-curve for Bangkok measles used to determine $\rho^{\mathrm{M}}_{\mathrm{Bkk}}$.

Table V. The "corner" data for the Bankok measles L-curve graph

ρ	1000	100	10	1
$\|\beta(\tau)\|_1$	6.30	3.57	0.42	0.00
Obj. Val.	0.32	0.33	0.46	0.64

a. Seasonal $\beta(\tau)$ Infectious Cases Results

As with New York City, the recorded case data has a reporting interval of one month, giving twelve data points for each year. Three finite elements were selected for each reporting interval giving a value of 36 finite elements per year which approximately

corresponds to 10.14 days per finite element. The estimation formulation had a total amount of 3,391 variables, 3316 equality constraints, and solved to an objective value of 3.625e-01 with a CPU time of only 1.32 seconds. The reporting factor was assumed to be linear over the interval with the values ranging 0.012-0.038. It is interesting to note is that even though the recorded cases do not have a definite major-minor biennial pattern, the estimated model does in fact appear to present those characteristics. The plot of the fit can be seen in Figure 10.

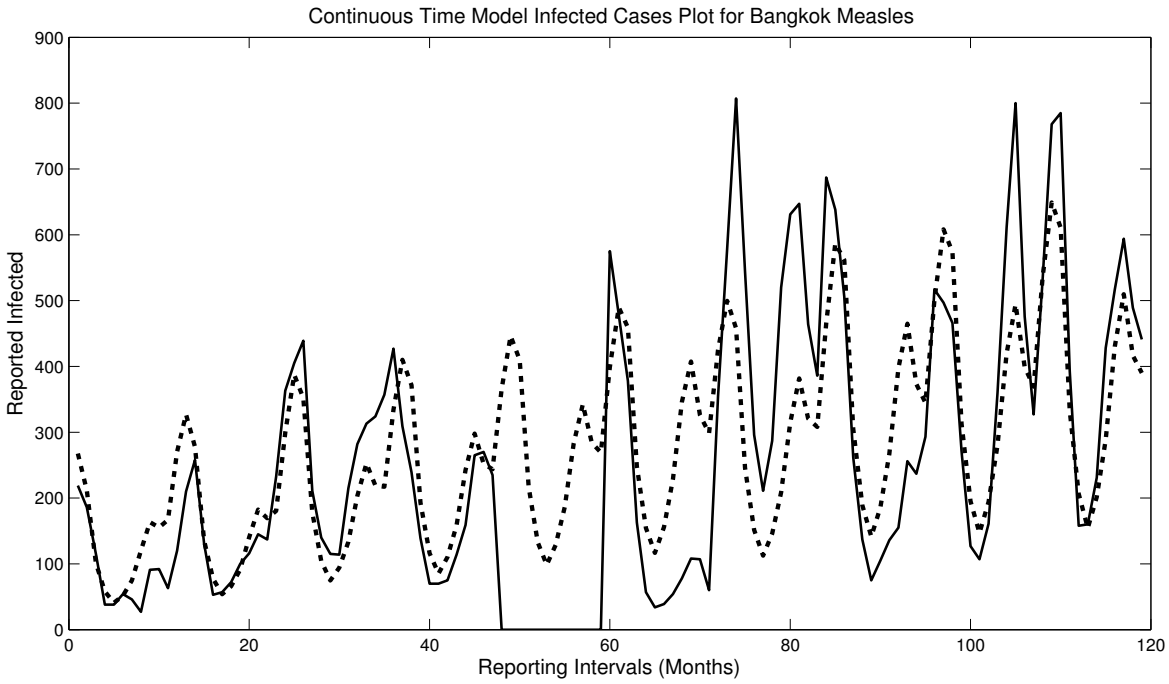

Fig. 10. Infectious cases results of the estimation for Bangkok measles 1975-1985; (- -) represents the estimation and (–) represents the historical data.

The estimated seasonal $\beta(\tau)$ for Bangkok measles can be found in Figure 11. It is important to note the characteristics of this graph compared to the previous two. This profile does not have the characteristic single trough around the third and fourth quarter of the year as in previous results. Instead, this graph has two very

pronounced troughs, one at the beginning of the second quarter of the year and the second appearing in the later part of the fourth quarter. This implies that there are two seasons where the infection has the opportunity to spread readily among the population and two distinct times when the transmission of infection is slowed.

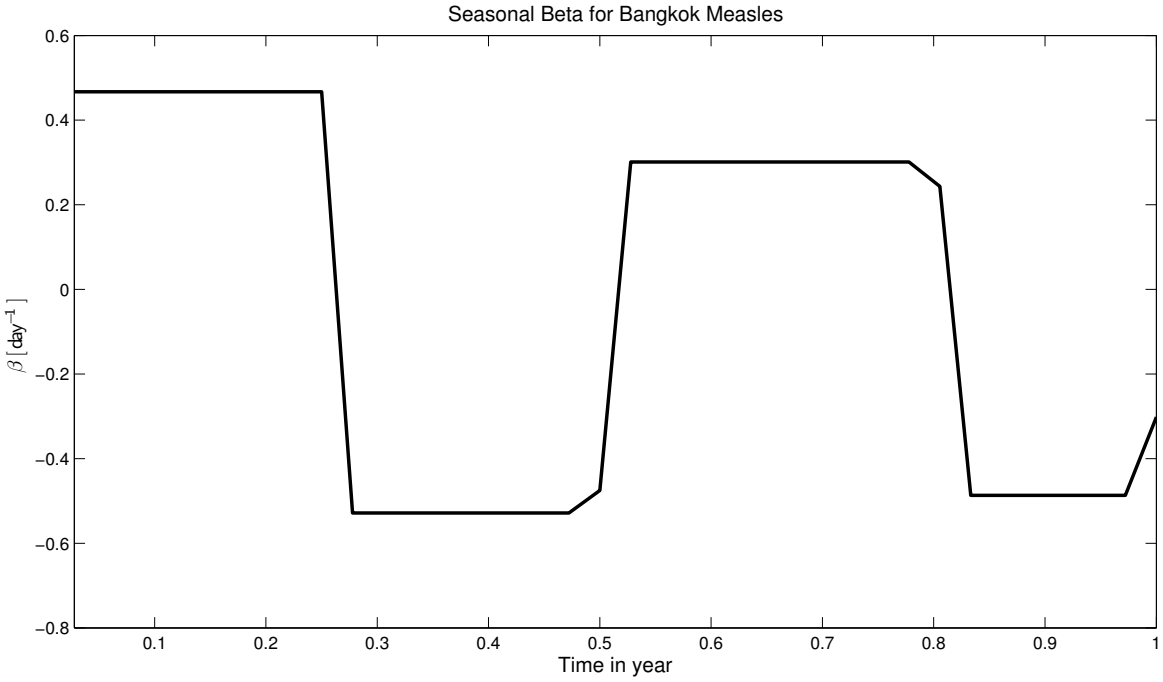

Fig. 11. $\beta(\tau)$ results for Bangkok measles 1975-1985.

b. Estimation Without Periodicity Constraint

Again, the estimation was solved witout the yearly periodicity constraint. Figure 12 shows the transmission profile for each year superimposed on eachother. Even with the missing data, the transmission parameter profiles are remarkably consistent.

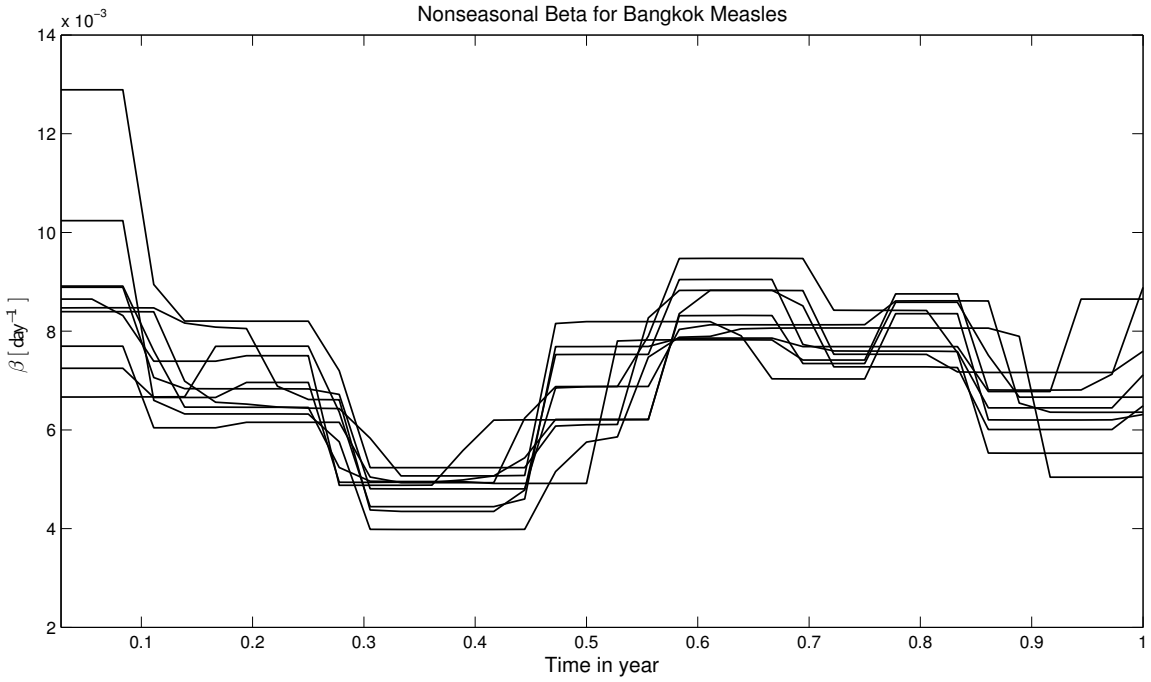

Fig. 12. Bangkok measles 1975-1985 nonseasonal $\beta(t)$ estimation results (plotted as years superimposed over each other).

B. Chickenpox

The data for cases of chickenpox was obtained and analysed for the cities of New York and Bangkok.

Table VI. Overall values for the chickenpox model estimations of New York City and Bangkok

	ρ	FE[a]	Rep. Fac.	Tot. Var.	Tot. Const.	Obj. Val.	CPU [b]
New York Chickenpox	3	36	0.067-0.043	6631	6556	2.022e+00	1.16
Bangkok Chickenpox	16	36	.02-.01	3391	3316	4.850e-01	0.712

[a] Finite elements per year.
[b] Times are reported in seconds.

1. New York Chickenpox 1944-1964

The time frame for our chickenpox data in New York City spans 20 years with a monthly reporting interval. Again 3 finite elements were selected per reporting period resulting in 36 finite elements per year. The L-curve is relatively clear and the value of ρ corresponding to the "corner" is approximately 3, (see Figure 13). The reporting factor was assumed to be linear beginning at 0.067 and decreasing down to 0.043 (see Table VI). The data set used is within the era of pre-vaccination since a vaccine for chickenpox was not developed until the mid 1990s.

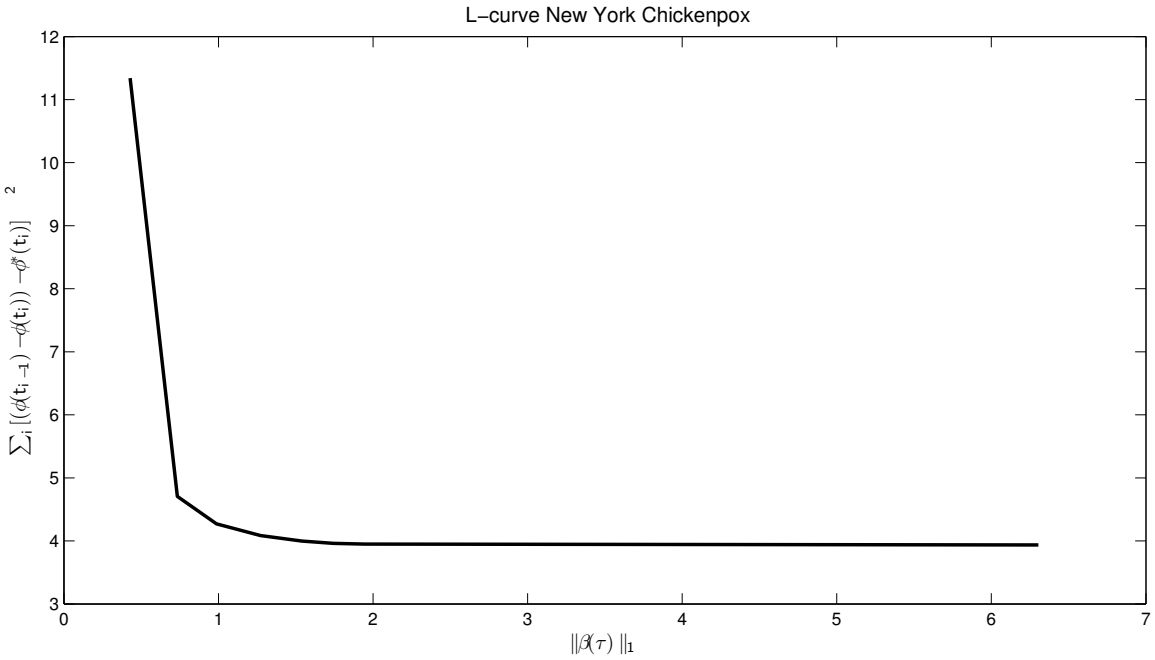

Fig. 13. The L-curve for New York chickenpox used to determine $\rho^{\mathrm{C}}_{\mathrm{NYC}}$.

a. Seasonal $\beta(\tau)$ Infectious Cases Results

This formulation contains 6,631 total variables and 6,556 equality constraints, and solved to an objective value of 2.022e+00 in a CPU time of 1.16 seconds. Chickenpox, unlike measles, has yearly outbreaks of relatively the same magnitude with major-minor neighboring outbreaks being the exception rather than the rule. The plot of the fit can be seen in Figure 14. The fit follows the general trends of the outbreaks. With the exception of the first outbreak, the estimated model accurately captures when there is an increase or a decrease in each outbreak peak.

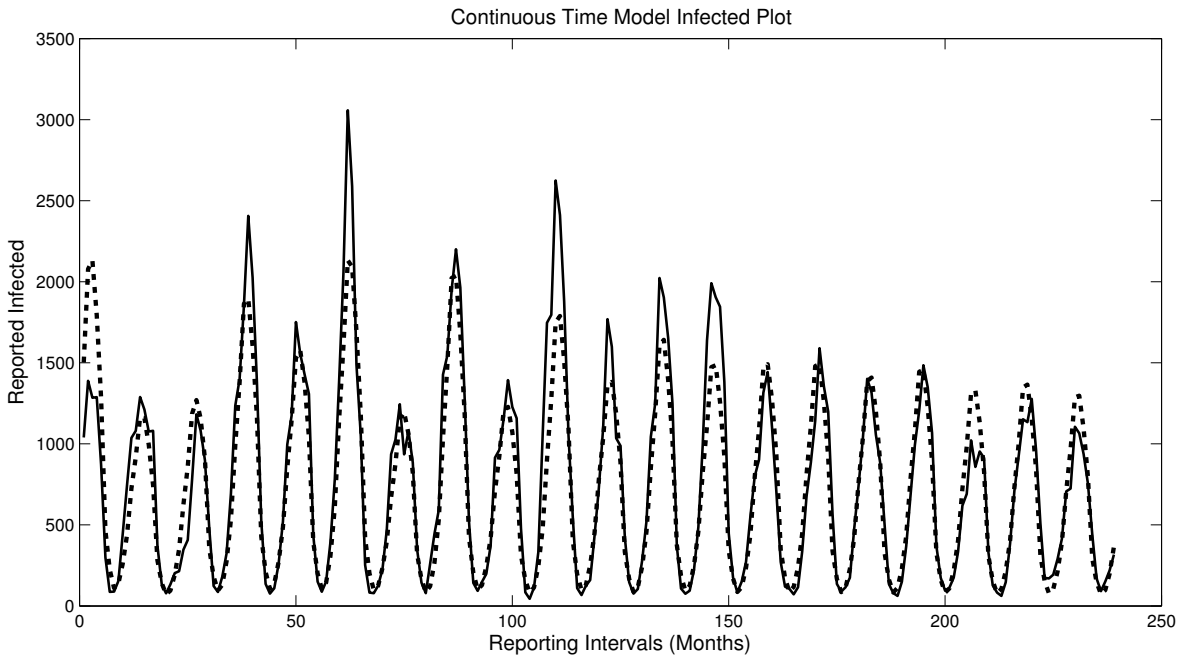

Fig. 14. Infectious cases results of the estimation for New York chickenpox 1944-1964; (- -) represents the estimation and (–) represents the historical data.

The seasonal $\beta(\tau)$ estimated for New York chickenpox can be seen in Figure 15. The shape of this $\beta(\tau)$ is very reminiscent of the seasonal $\beta(\tau)$ for measles in both New York and London. The only major difference is the length of the trough for this

parameter lasting further into the fourth quarter of the year. This is quite interesting because it could imply that the same factors that drive the transmission of measles also drive the spread of chickenpox, at least with respect to the cities of New York and London.

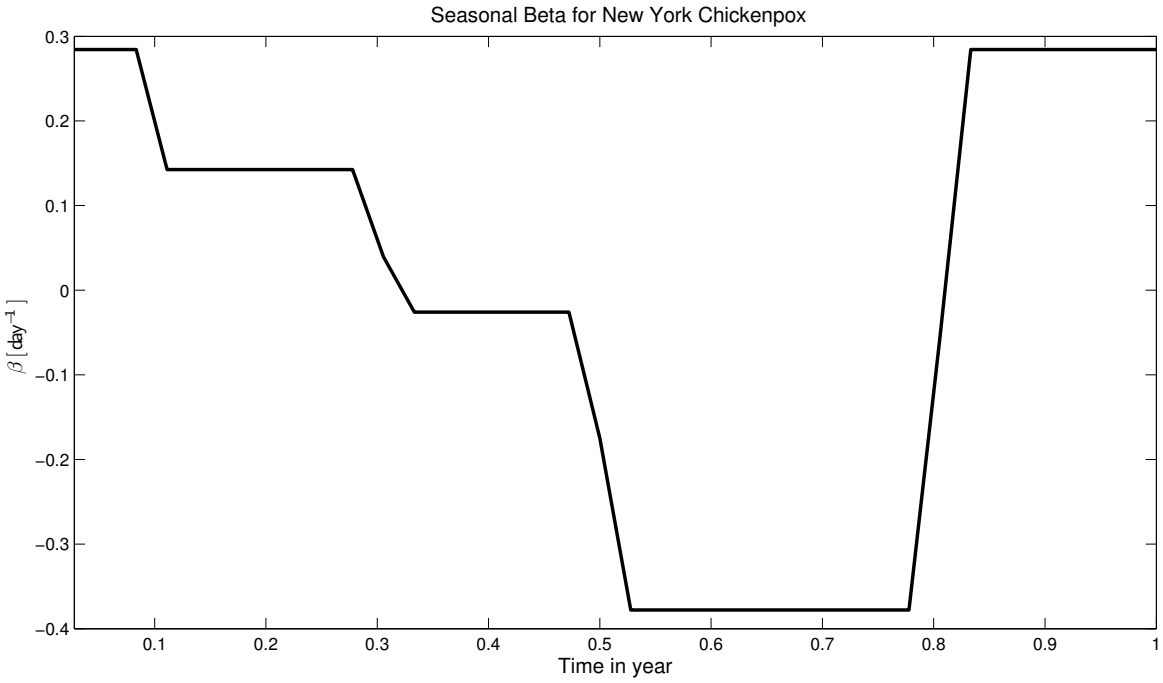

Fig. 15. $\beta(\tau)$ results for New York chickenpox 1944-1964.

b. Estimation Without Periodicity Constraint

When the periodicity constraint on $\beta(\tau)$ was removed, the resulting estimates again gave quite consistent results from year to year as seen in Figure 16. The shape of the graphs is defined, and the trough of this graph seems to match up with the seasonal transmission parameters for New York and London much more accurately, beginning during the third quarter and ending before the fourth quarter. The formulation

properties for this estimation are: 9,368 total variables, 7,924 constraints, an objective value of 1.8785e-02, and a CPU time of 15.93 seconds.

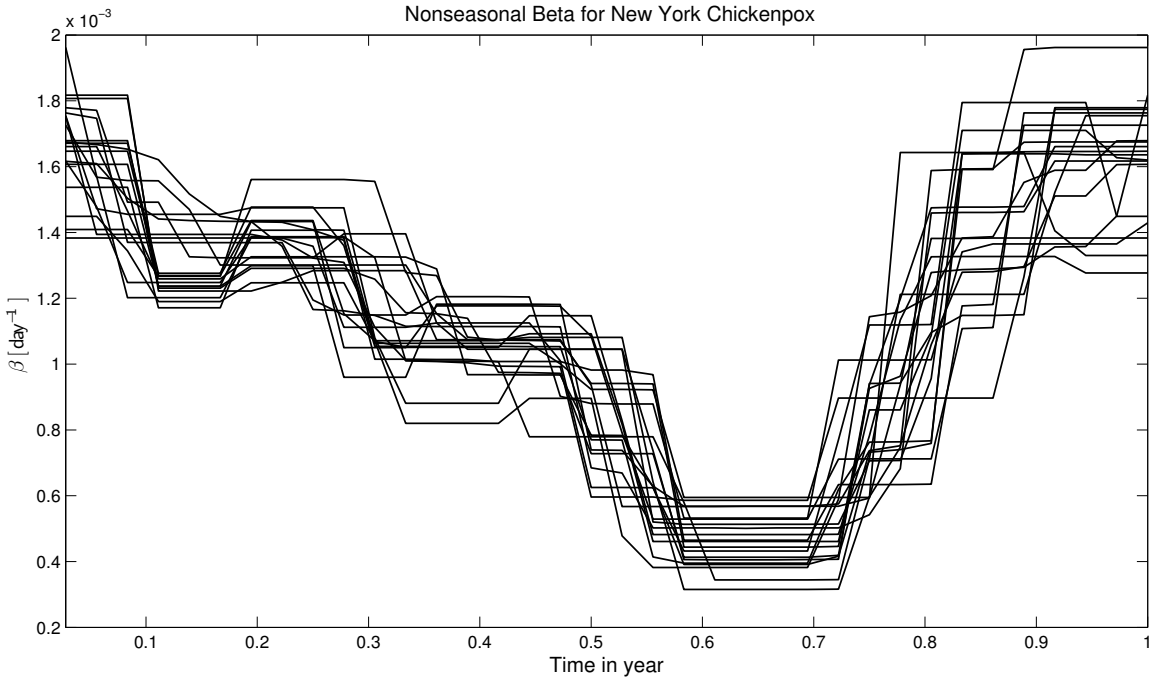

Fig. 16. New York chickenpox 1944-1964 nonseasonal $\beta(t)$ estimation results (plotted as years superimposed over each other).

2. Bangkok Chickenpox 1985-1995

Our available data for chickenpox in Bangkok spans ten years from 1985-1995 with monthly case reporting. Again, 3 finite elements per reporting interval were selected giving a total of 36 finite elements per year. The L-curve plot does not have a clear value for ρ since there is not a clear "corner" (see Figure 17). The value judged to be the best ρ for this estimation was 16. This value corresponds to the point (0.5936, 0.4479) on the plot, which is the point right as the curve starts to flatten out. The motivation behind the choice of this point is to allow as much "fit" of the cases as possible while also trying to get above the minimum regularization.

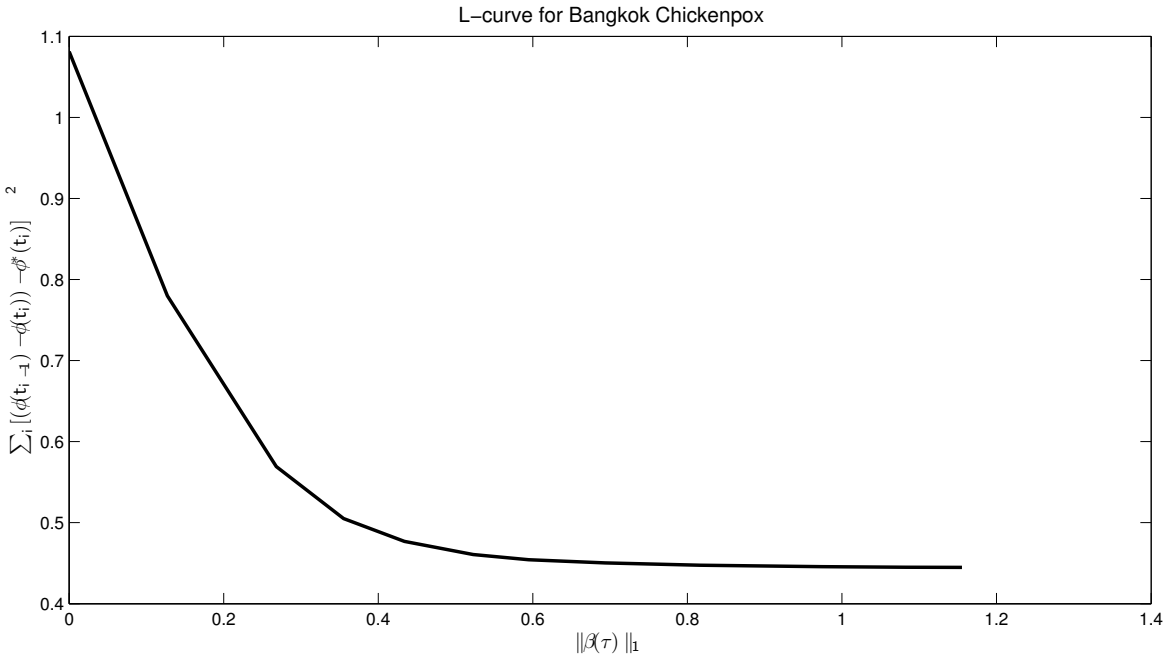

Fig. 17. The L-curve for Bangkok chickenpox used to determine $\rho^{\mathrm{C}}_{\mathrm{Bkk}}$.

a. Seasonal $\beta(\tau)$ Infectious Cases Results

Outlier elimination was performed on the two uncharacteristic points located around reporting intervals 41 and 55, in order to have a more consistent data set for the estimation. The reporting factor is assumed to be linear, decreasing from 0.02001 to 0.01011. The properties of the estimation formulation are as follows: 3,391 total variables, 3,316 constraints, an objective value of 4.850e-01, and a CPU time of 0.712 seconds. The fit can be seen in Figure 18.

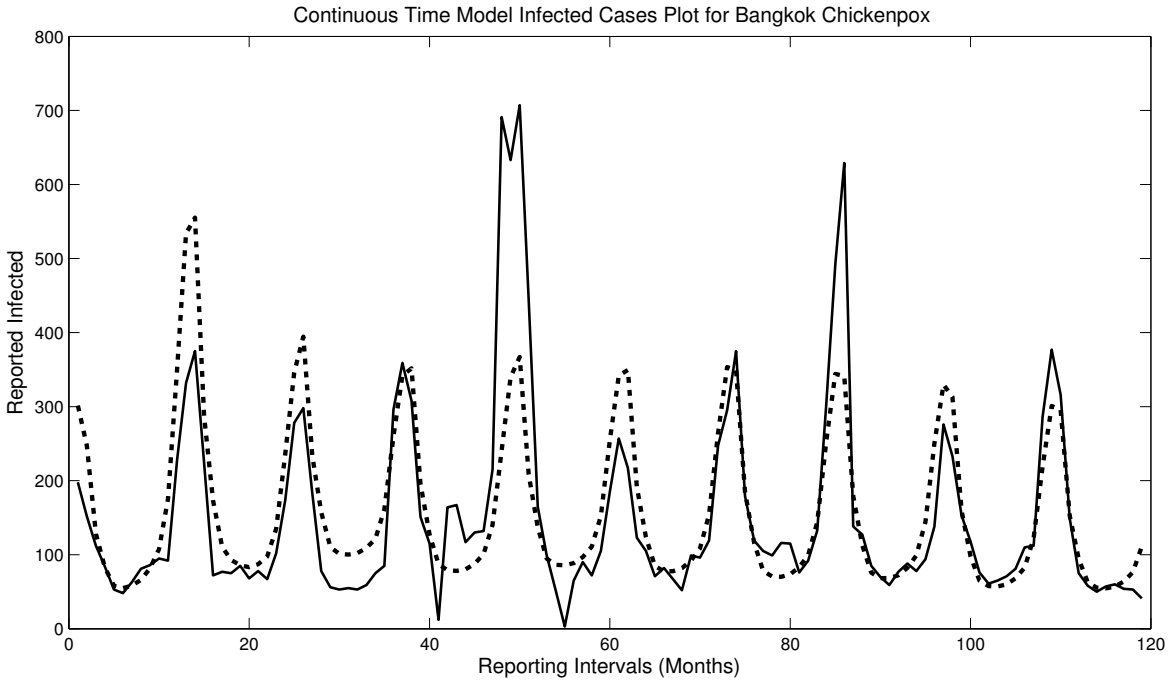

Fig. 18. Infectious cases results of the estimation for Bangkok chickenpox 1985-1995; (- -) represents the estimation and (–) represents the historical data.

The estimated seasonal $\beta(\tau)$ profile can be seen in Figure 19. This seasonal $\beta(\tau)$ does not resemble the Bangkok seasonal parameter for measles. However, the trough which appears in this function corresponds closely to the first trough of the seasonal

measles transmission parameter for Bangkok.

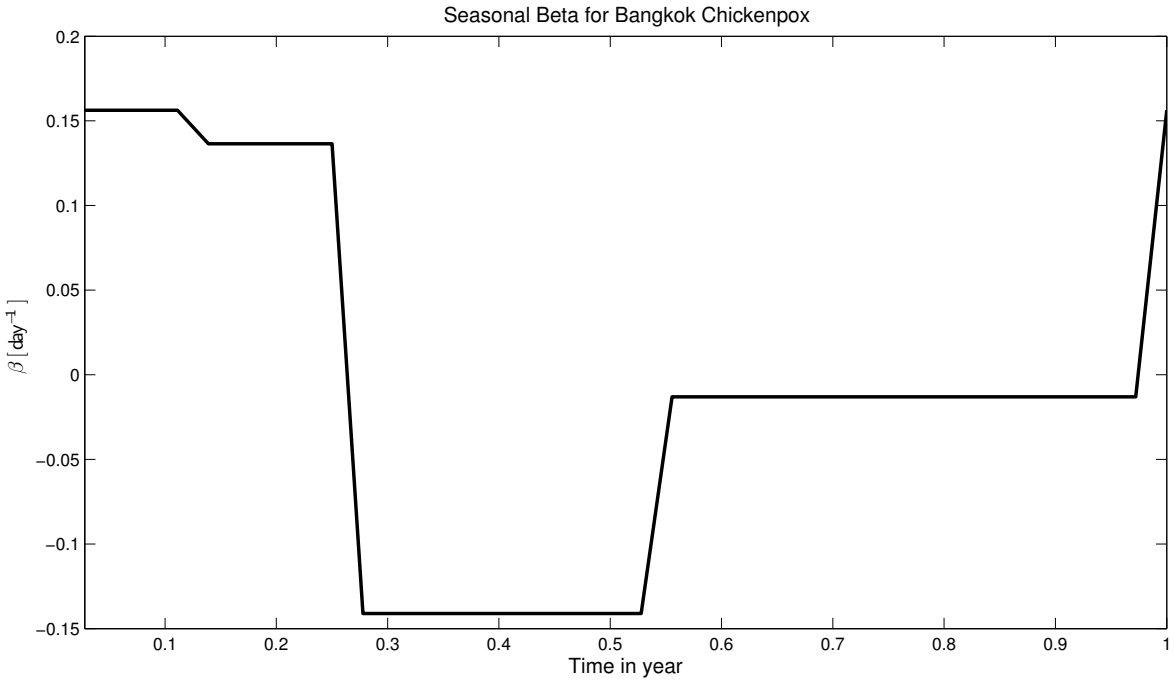

Fig. 19. $\beta(\tau)$ results for Bangkok chickenpox 1985-1995.

b. Estimation Without Periodicity Constraint

Again, the estimation was performed without the periodicity constraint on $\beta(\tau)$. Looking at the graph in Figure 20, it can be easily seen that the general shape is evident, and there are only a couple of years that do not match the trend. This could be attributed to several different factors including noise in the model, inconsistent measurements within the outlier intervals, or temporary changes in reporting procedures. However, the general trend of this graph does correspond with the shape of the seasonal transmission parameter giving greater confidence that the estimations are reasonable despite the fact they do not match the measles transmission parameter.

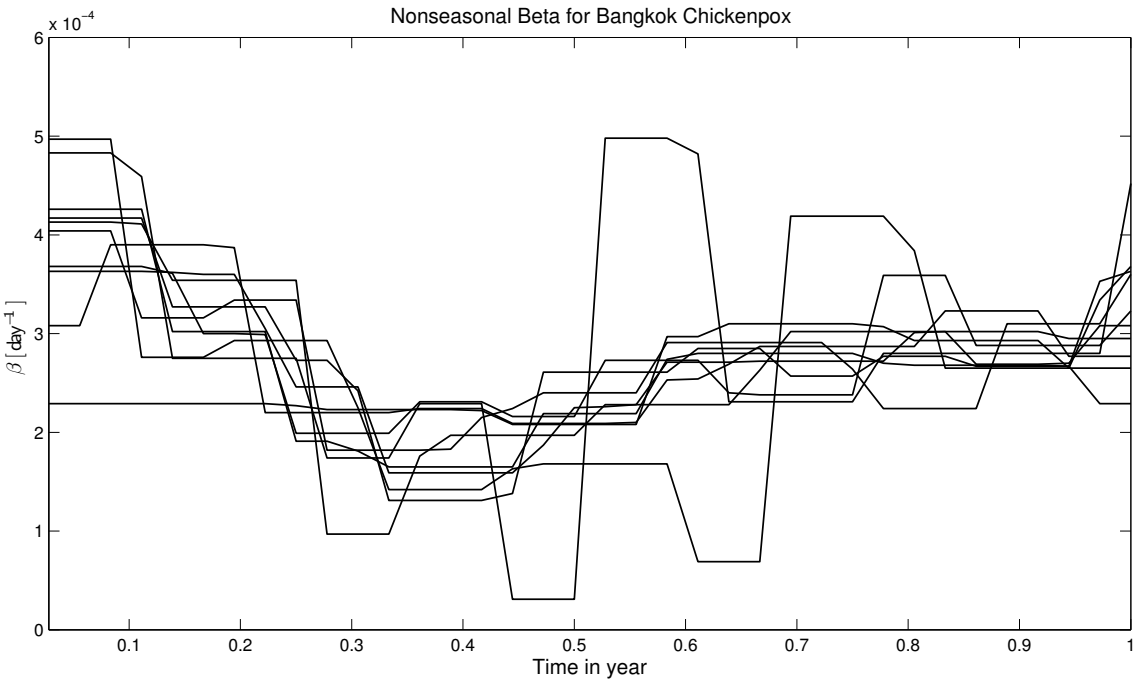

Fig. 20. Bangkok chickenpox 1985-1995 nonseasonal $\beta(t)$ estimation results (plotted as years superimposed over each other).

CHAPTER VI

CONCLUSIONS

A. New York and London Seasonality

Previous literature concerning the seasonality of London measles has demonstrated a correlation between the seasonal transmission parameter and school holiday schedules [2]. The results of our continuous-time estimation also show this correlation not only with London measles but also with New York measles and chickenpox.

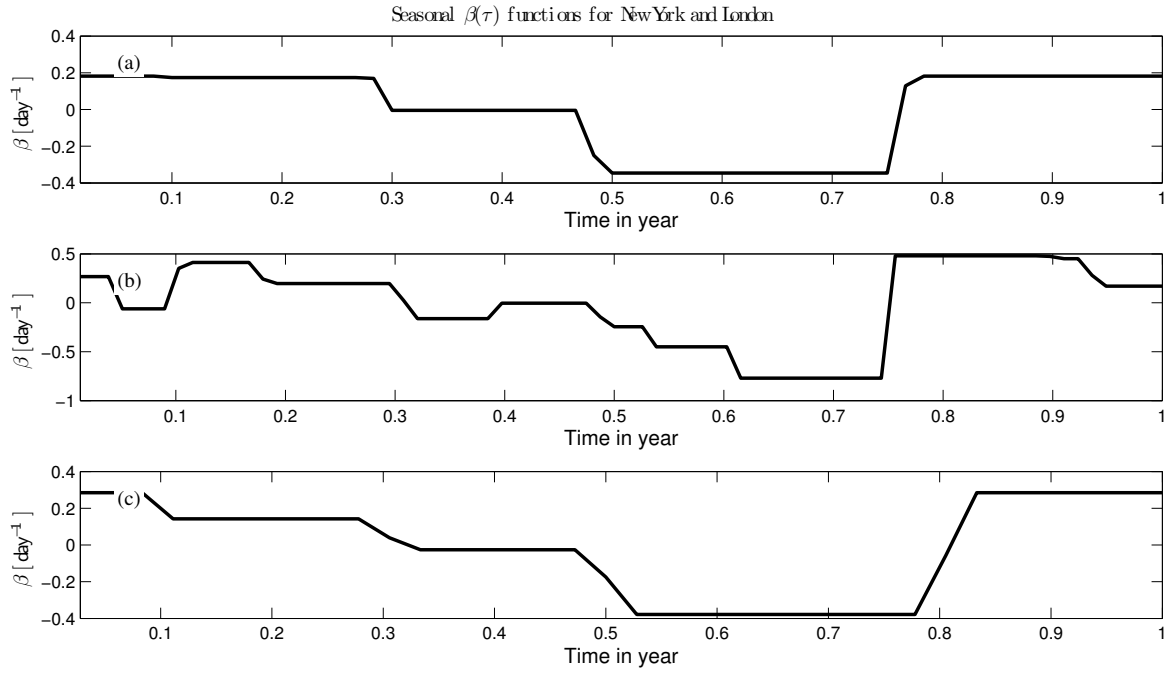

Fig. 21. Comparison of the graphs of $\beta(\tau)$ for (a) New York measles 1947-1963, (b) London measles 1944-1958, (c) New York chickenpox 1944-1964.

1. New York Measles 1947-1963

Traditionally, the American summer holiday lasts roughly from June to September, or 0.42-0.67 in the normalized year scale. Looking at Figure 21 (a), one can see that the summer break is reflected in the seasonal transmission parameter. During the summer break, when school is not in session social interaction is decreased. The seasonal $\beta(\tau)$ estimated is further justified by the results in Figure 22 which show the mean (solid line) and standard deviation (dashed line) for the nonseasonal $\beta(t)$ estimation without the periodicity constraint.

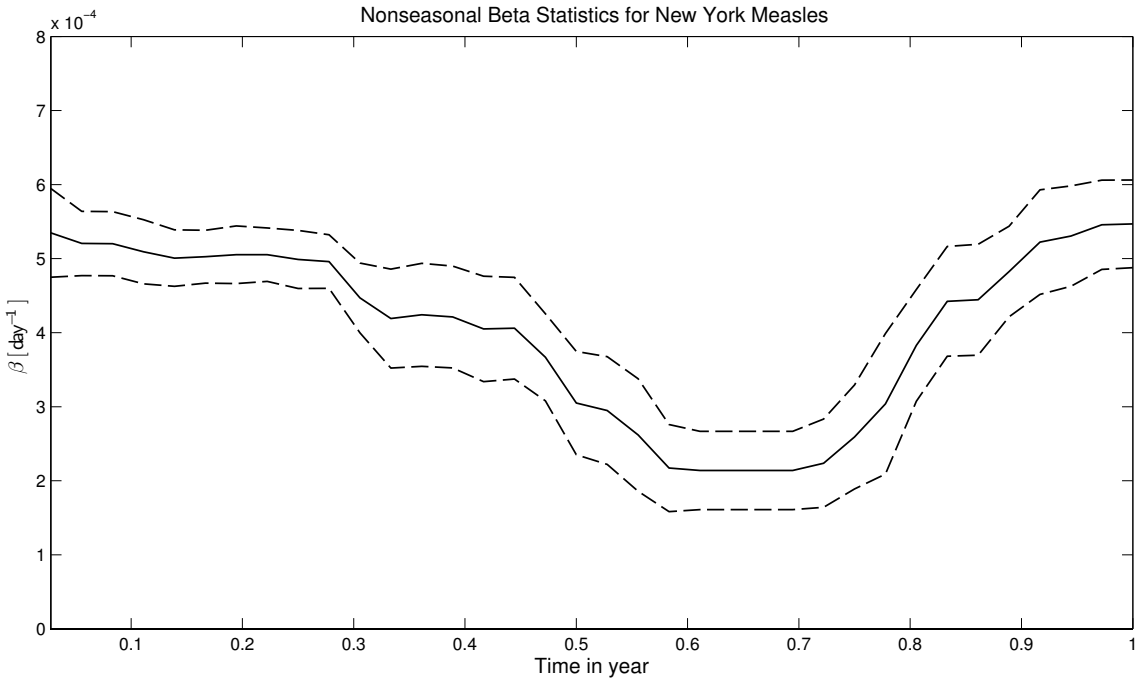

Fig. 22. New York measles 1947-1963 estimated $\beta(t)$ without periodicity constraint; (–) is the mean and (- -) is one standard deviation from the mean.

2. London Measles 1944-1958

The vacations for London roughly occur, utilizing the normalized scale of the graphs, at: 0.00-0.07 (continuation of winter break), 0.31-0.34 (two-week break), 0.57-0.69 (major summer holiday), 0.88-0.92 (two-week break), and 0.96-1.00 (beginning of winter break) [2]. The major summer holiday of London also roughly corresponds to the American summer holiday. In Figure 21 (b) and Figure 23, it can be seen that the transmission parameter drops for each of these holidays.

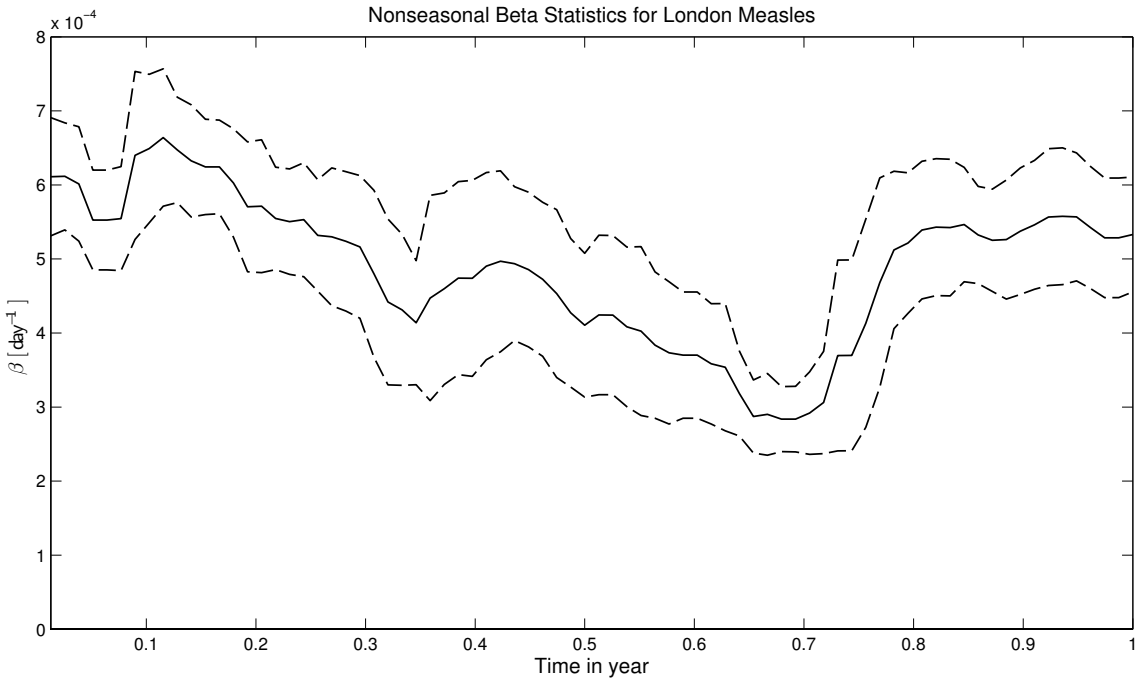

Fig. 23. London measles 1944-1958 estimated $\beta(t)$ without periodicity constraint; (–) is the mean and (- -) is one standard deviation from the mean.

3. New York Chickenpox 1944-1964

Observing Figure 21 (c) for New York City chickenpox, this too captures the summer school holiday, however it lingers longer into the fall school term, which may be

attributed to the longer serial interval of chickenpox being approximately 21 days as opposed to the serial interval of measles which is approximately 14 days. These results are further shown in Figure 24.

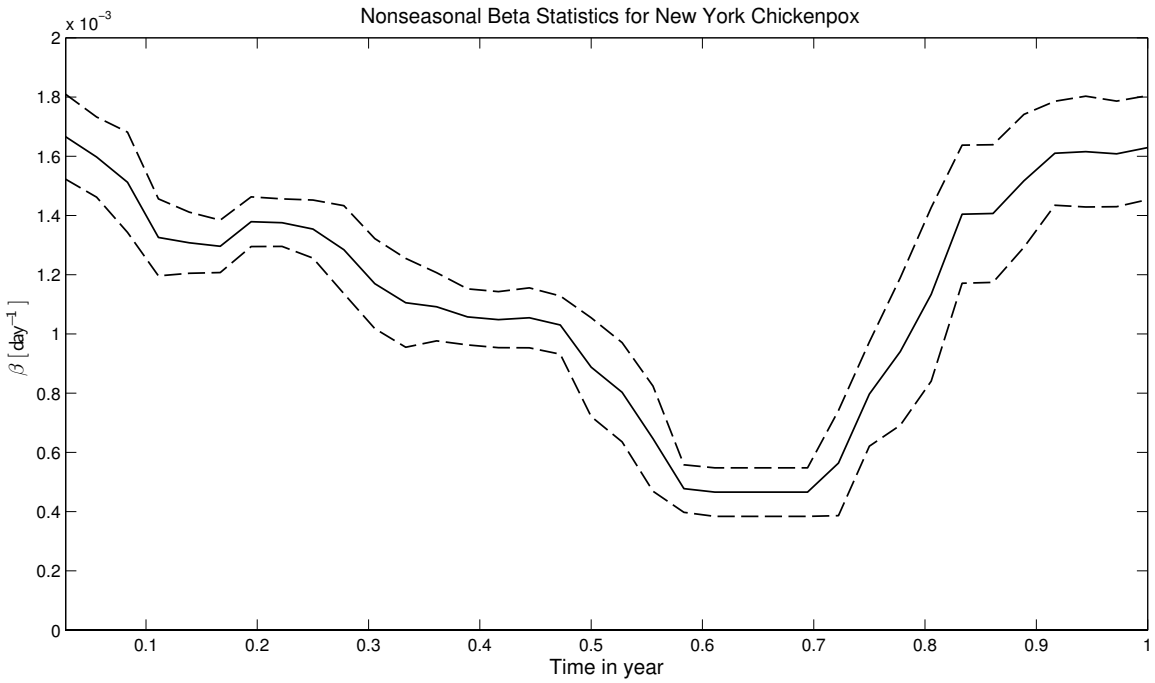

Fig. 24. New York chickenpox 1944-1964 estimated $\beta(t)$ without periodicity constraint; (–) is the mean and (- -) is one standard deviation from the mean.

B. Bangkok Seasonality

The data from Bangkok allows us to further verify that transmission parameters in childhood infectious diseases are correlated with school holiday schedules since Bangkok has a vastly different holiday schedule. The schools in Bangkok have two major holidays. The first holiday occurs in March and lasts through April, (0.16-0.33), and the second holiday lasts all of October, (0.75-0.92).

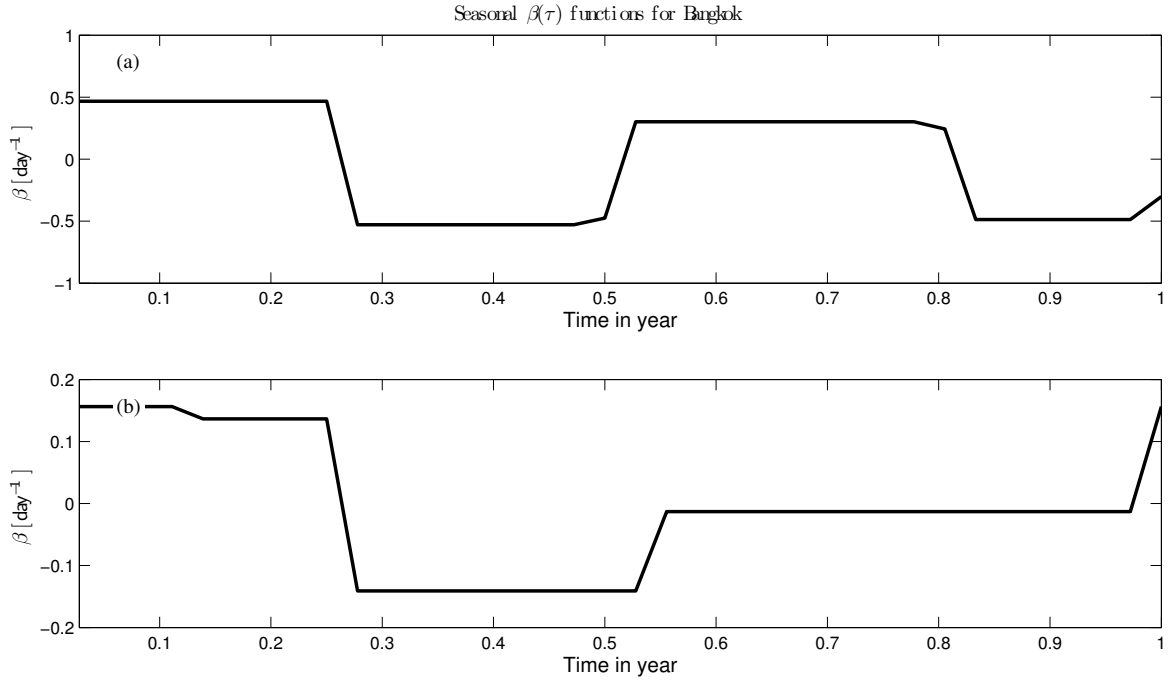

Fig. 25. Comparison of the graphs of $\beta(\tau)$ for (a) Bangkok measles 1975-1985, (b) Bangkok chickenpox 1985-1995.

1. Bangkok Measles 1975-1985

In Figure 25 (a) we see that the trough lingers past the end of each holiday before returning to the high value. There are several possible reasons for this. First, reporting in Bangkok may be delayed compared with the other locations studied. Second, this may be an inherent flaw in the model. Since the recovery rate is included in the differential equation as a constant, $-\gamma I$, it assumes an exponential distribution of the recovery time which is not representative of the true disease dynamics within the body. To remedy this one might incorporate a time-delay term instead of a constant rate, or a more accurate distribution function. Despite this, the estimation still exhibits strong correlation with the holiday schedule. The conclusions are further justified by the results in Figure 26.

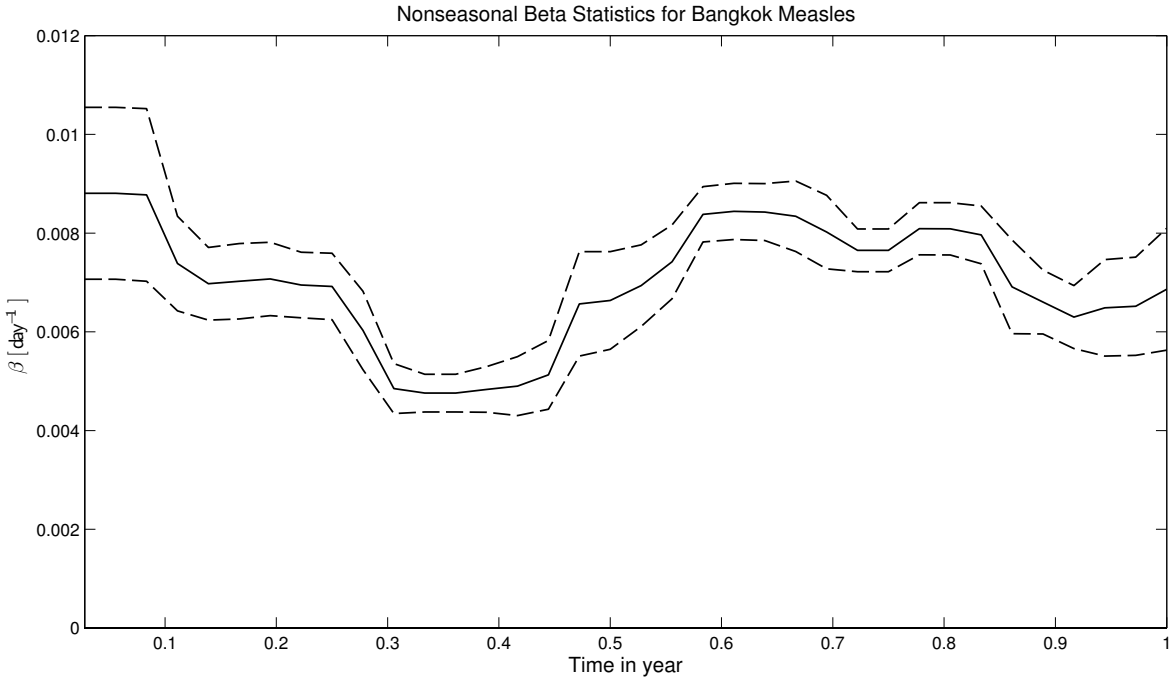

Fig. 26. Bangkok measles 1975-1985 estimated $\beta(t)$ without periodicity constraint; (–) is the mean and (- -) is one standard deviation from the mean.

2. Bangkok Chickenpox 1985-1995

The seasonal transmission parameter for Bangkok chickenpox can be seen in Figure 25 (b). This is not what one would expect to see. The first holiday can clearly be seen, but the second holiday is not apparent at all. Several factors may be contributing to this. This deficiency could be the exponentially distributed recovery rate, the longer serial interval of chickenpox, the quality of the data, or an invalid modeling assumption. These factors might also attribute to the apparent averaging of the high and low periods spanning 0.55-0.98. Similar results can be seen in Figure 27.

These results show that the school holiday schedule is strongly correlated with the estimated transmission parameter profile for all of the data sets studied. This

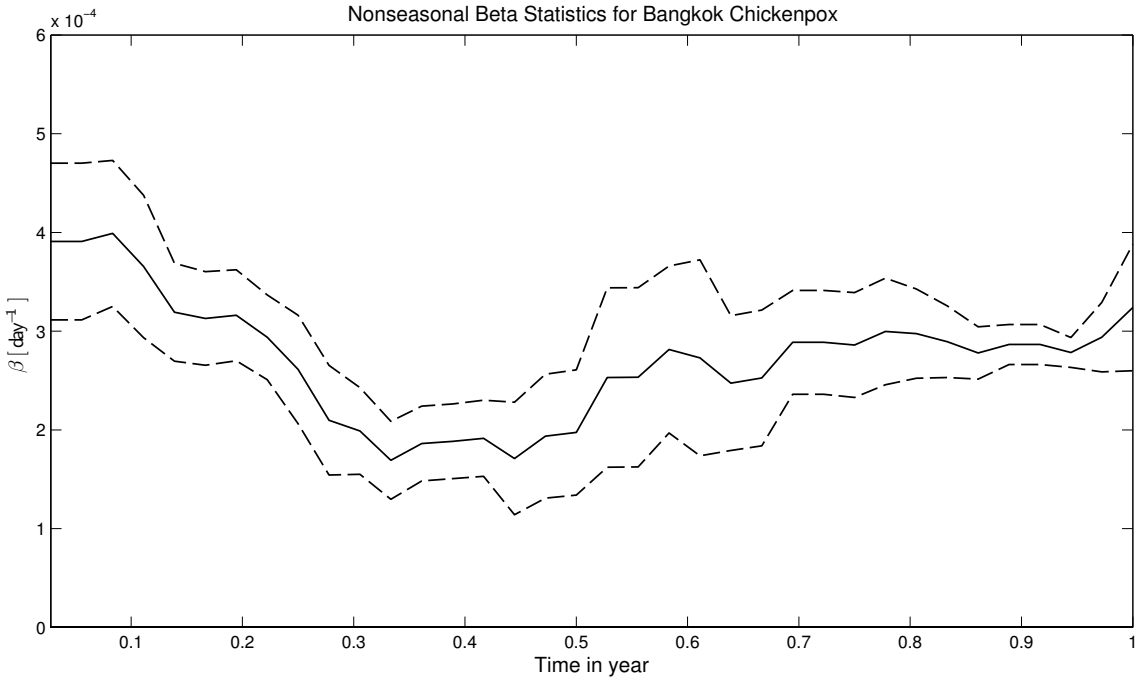

Fig. 27. Bangkok chickenpox 1985-1995 estimated $\beta(t)$ without periodicity constraint; (–) is the mean and (- -) is one standard deviation from the mean.

approach is also very efficient compared to previous methods. The MCMC method employed by Cauchemez and Ferguson [41], while it was able to effectively estimate seasonality patterns, took over 20 hours on a desktop computer for a single run. This method is vastly more efficient. This work sets the stage for new development in model estimation and optimization.

C. Future Work

This continuous-time model was formulated using an Explicit Euler method for initialization, orthogonal collocation on finite elements with Radau points as a discretization method for the ODEs, total variation regularization on the parameter, and an interior point primal-dual filter search method for optimization of the objective. These meth-

ods were chosen because they were believed to most effectively capture the estimation goals.

The future work to consider would be to explore different discretization techniques, both single-step and multi-step in order to evaluate the accuracy they might offer in the estimation formulation. Various methods with increasing degrees of order accuracy would give valuable insight as to the necessary degree of accuracy required in estimation and future modeling efforts. A comparison of these methods might also provide a good comparison of degree of accuracy versus time efficiency of the calculation.

Regularization of the $\beta(\tau)$ is necessary to restrict the "shape" of the seasonal transmission parameter function. This research assumed that the estimated profile might be discontinuous, however, this might not prove to be the only case that accurately keeps the case data. An exploration and comparison of various regularizations of $\beta(\tau)$ would help understand the best representation of the yearly patterns of these infections.

This work was based on one of the simplest seasonal transmission models possible, a seasonal SIR model. The SIR model contains several assumptions that might not be valid for specific case studies. Fortunately the nonlinear programming formulation presented is both flexible and efficient. Several model improvements should be investigated, including different recovery functions and mixing assumptions.

This entire work assumed deterministic dynamics. It is remarkable that these models can so effectively capture the observed state profiles. Nevertheless, in smaller communities stochastic fadeout is evident and stochastic models must be considered.

REFERENCES

[1] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, pp. 599–653, October 2000.

[2] B. F. Finkenstädt and B. T. Grenfell, "Time series modelling of childhood diseases: A dynamical systems approach," *Journal of the Royal Statistical Society, Series C*, vol. 49, pp. 187–205, 2000.

[3] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*. New York: Oxford University Press Inc., 1991.

[4] D. Bernoulli, "Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir," *Histoires et Mémoires de l'Académie Royale des Sciences de Paris*, pp. 1–45, 1760.

[5] W. H. Hamer, *Epidemic Disease in England*. The Milroy Lectures, London: Bedford Press, 1906.

[6] R. Ross, *The Prevention of Malaria*. New York: E.P. Dutton & Company, 1910.

[7] N. T. Bailey, *The Mathematical Theory of Infectious Diseases*. Hafner, NY: A Charles Griffin Book, 1975.

[8] A. J. Lotka, "The stability of the normal age distribution," *Proceedings of the National Academy of Sciences of the USA*, vol. 8, pp. 339–345, 1922.

[9] K. Dietz, "Epidemics and rumours: A survey," *Journal of the Royal Statistical Society*, vol. 130, no. 4, pp. 505–528, 1967.

[10] K. Dietz, "The first epidemic model: A historical note on P.D. En'ko," *Australian & New Zealand Journal of Statistics*, vol. 30A, no. 1, pp. 56–65, 1988.

[11] A. G. McKendrick, "Applications of mathematics to medical problems," *Proceedings of the Edinburgh Mathematical Society*, vol. 44, pp. 98–130, 1925.

[12] W. O. Kermack and A. G. McKendrick, "Contributions to the mathematical theory of epidemics – I," *Proceedings of the Royal Society of London Series A*, vol. 115, pp. 700–721, 1927.

[13] N. Becker, "The uses of epidemic models," *Biometrics*, vol. 35, no. 1, pp. 295–305, 1979.

[14] K. Dietz and D. Schenzle, "Mathematical models for infectious disease statistics," in *A Celebration of Statistics,* A. C. Atkinson and S. E. Feinberg (eds.), New York: Springer-Verlag, 1985, pp. 167-204.

[15] V. Isham and G. Medley, *Models for Infectious Human Diseases.* Cambridge: Press Syndicate of the University of Cambridge, 1996.

[16] K. Dietz, "Transmission and control of arbovirus diseases," in *Epidemiology,* D. Ludwig and K. L. Cooke (eds.), Philadelphia: SIAM, 1975, pp. 104-121.

[17] J. A. Jacquez, C. P. Simon, and J. S. Koopman, "The reproduction number in deterministic models of contagious diseases," *Current Topics in Theoretical Biology*, vol. 2, pp. 159–209, 1991.

[18] H. E. Soper, "The interpretation of periodicity in disease prevalence," *Journal of the Royal Statistical Society*, vol. 92, pp. 34–73, 1929.

[19] P. E. Fine and J. A. Clarkson, "Measles in England and Wales I: An analysis of factors underlying seasonal patterns," *International Journal of Epidemiology*, vol. 11, no. 1, pp. 5–14, 1982.

[20] Y. Xia, O. N. Bjornstad, and B. T. Grenfell, "Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics," *The American Naturalist*, vol. 164, no. 2, pp. 267–281, 2004.

[21] M. Soroush, "State and parameter estimations and their applications in process control," *Computers & Chemical Engineering*, vol. 23, no. 2, pp. 229–245, 1998.

[22] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*. CRC Applied Mathematics and Nonlinear Science Series, Boca Raton, FL: Chapman & Hall, 2004.

[23] U. M. Ascher and L. R. Petzold, *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. Philadelphia, PA: SIAM, 1998.

[24] D. Simon, *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*. New York: John Wiley & Sons, 2006.

[25] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control: Theory and Design*. Madison, WI: Nob Hill Publishing, 2009.

[26] L. T. Biegler, A. M. Cervantes, and A. Wächter, "Advances in simultaneous strategies for dynamic process optimization," *Chem. Eng. Sci.*, vol. 57, pp. 575–593, 2002.

[27] S. Kameswaran and L. T. Biegler, "Simultaneous dynamic optimization strategies: Recent advances and challenges," *Computers & Chemical Engineering*, vol. 30, pp. 1560–1575, 2006.

[28] A. Cervantes and L. T. Biegler, "Optimization strategies for dynamic systems," in *Encyclopedia of Optimization,* C. Floudas and P. Pardalos (eds.), Dordrecht: Kluwe Academic Publishers, 2000, pp. 487-488.

[29] J. T. Betts, "Practical methods for optimal control using nonlinear programming," *Applied Mechanics Reviews*, vol. 55, no. 4, p. B68, 2002.

[30] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems.* Burlington, MA: Elsevier Academic Press, 2005.

[31] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Review*, vol. 34, no. 4, pp. 561–580, 1992.

[32] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems.* Philadelphia, PA: SIAM, 1995.

[33] T. J. Cornwell and K. F. Evans, "A simple maximum entropy deconvolution algorithm," *Astronomy and Astrophysics*, vol. 143, no. 1, pp. 77–83, 1985.

[34] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming.* Danvers, MA: The Scientific Press, 1993.

[35] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, 2006.

[36] J. Åkesson, *Languages and Tools for Optimization of Large-Scale Systems.* Lund, Sweden: Media-Tryck, 2007.

[37] G. Bader and U. M. Ascher, "A new basis implementation for mixed order boundary value ode solvers," *SIAM Journal of Scientific and Statistical Computation*, vol. 8, no. 4, pp. 483–500, 1987.

[38] V. M. Zavala and L. T. Bielger, "Large-scale parameter estimation in low-density polyethylene tubular reactors," *Industrial & Engineering Chemistry Research*, vol. 45, pp. 7867–7881, 2006.

[39] G. H. Abbott III, D. P. Word, D. Cummings, and C. D. Laird, "Estimating seasonal drivers in childhood infectious diseases with continuous time and discrete-time models," American Control Conference, 2010. Submitted.

[40] A. J. K. Conlan and B. T. Grenfell, "Seasonality and the persistence and invasion of measles," *Proceedings of the Royal Society B*, vol. 274, no. 1, pp. 113–1141, 2007.

[41] 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*, vol. 5, no. 25, pp. 885–897, 2008.

APPENDIX A

AMPL FILES OF THE ESTIMATION FORMULATION

The main *.run file

```
#--------------------------------------------------------------------#
# AMPL routine for SIR disease model                                 #
# George Abbott Fall 2009                                            #
#                                                                    #
# 1 reset model  (clean memory)                                      #
# 2 set model name                                                   #
# 3 set data  file                                                   #
# 4 set solver and options                                           #
# 5 solve the optimization problem                                   #
# 6 include any additional files needed                              #
#                                                                    #
#                                                                    #
#--------------------------------------------------------------------#

reset                                    ; # 1
model SIR_RCFE_LIbeta.mod                ; # 2
data SIR_RCFE_LI.dat                     ; # 3


let beta[0] := P_BETA[1];
let beta[fepy] := P_BETA[1];


############### Initialization of the functions
for {n in ye diff {0}}
{
    let beta[n-1]  :=  P_BETA[n];
}

for {i in fe}
{
    let h[i] := 1/nfe;
}

################ forward Euler method
```

```
let inf[1,1]  :=  inf_init_spec;
let sus[1,1]  :=  sus_init_spec;
let phi[1,1]  :=  0;

let inf[1,2]  :=  inf[1,1] + (cpoints[2]-cpoints[1])*time/nfe*
(beta[1]*inf[1,1]*sus[1,1]/P_POP - P_GAMMA*inf[1,1]);
let sus[1,2]  :=  sus[1,1] + (cpoints[2]-cpoints[1])*time/nfe*
(-beta[1]*inf[1,1]*sus[1,1]/P_POP + P_BIRTHS[1]);
let phi[1,2]  :=  phi[1,1] + (cpoints[2]-cpoints[1])*time/nfe*
(beta[1]*inf[1,1]*sus[1,1]/P_POP);

let inf[1,3]  :=  inf[1,2] + (cpoints[3]-cpoints[2])*time/nfe*
(beta[1]*inf[1,2]*sus[1,2]/P_POP - P_GAMMA*inf[1,2]);
let sus[1,3]  :=  sus[1,2] + (cpoints[3]-cpoints[2])*time/nfe*
(-beta[1]*inf[1,2]*sus[1,2]/P_POP + P_BIRTHS[1]);
let phi[1,3]  :=  phi[1,2] + (cpoints[3]-cpoints[2])*time/nfe*
(beta[1]*inf[1,2]*sus[1,2]/P_POP);

for {i in fe diff{1} }
{
let inf[i,1]  :=  inf[(i-1),1] + cpoints[1]*time/nfe*
(beta[(1+(i-2)mod(fepy))]*inf[(i-1),3]*sus[(i-1),3]/P_POP
- P_GAMMA*inf[(i-1),3]);
let sus[i,1]  :=  sus[(i-1),1] + cpoints[1]*time/nfe*
(-beta[(1+(i-2)mod(fepy))]*inf[(i-1),3]*sus[(i-1),3]/P_POP
+ P_BIRTHS[i]);
let phi[i,1]  :=  phi[(i-1),1] + cpoints[1]*time/nfe*
(beta[(1+(i-2)mod(fepy))]*inf[(i-1),3]*sus[(i-1),3]/P_POP);

let inf[i,2]  :=  inf[i,1] + (cpoints[2]-cpoints[1])*time/nfe*
(beta[(1+(i-1)mod(fepy))]*inf[i,1]*sus[i,1]/P_POP - P_GAMMA*inf[i,1]);
let sus[i,2]  :=  sus[i,1] + (cpoints[2]-cpoints[1])*time/nfe*
(-beta[(1+(i-1)mod(fepy))]*inf[i,1]*sus[i,1]/P_POP + P_BIRTHS[i]);
let phi[i,2]  :=  phi[i,1] + (cpoints[2]-cpoints[1])*time/nfe*
(beta[(1+(i-1)mod(fepy))]*inf[i,1]*sus[i,1]/P_POP);

let inf[i,3]  :=  inf[i,2] + (cpoints[3]-cpoints[2])*time/nfe*
(beta[(1+(i-1)mod(fepy))]*inf[i,2]*sus[i,2]/P_POP - P_GAMMA*inf[i,2]);
let sus[i,3]  :=  sus[i,2] + (cpoints[3]-cpoints[2])*time/nfe*
(-beta[(1+(i-1)mod(fepy))]*inf[i,2]*sus[i,2]/P_POP + P_BIRTHS[i]);
let phi[i,3]  :=  phi[i,2] + (cpoints[3]-cpoints[2])*time/nfe*
(beta[(1+(i-1)mod(fepy))]*inf[i,2]*sus[i,2]/P_POP);
}
```

```
let inf_init := inf_init_spec;
let sus_init := sus_init_spec;
fix beta;

######################### start of solving

option  solver ipopt                    ; # 4
solve                                   ; # 5

unfix beta                              ; # 6
option  solver ipopt                    ; # 7
solve                                   ; # 8


#============= Beginning of the Matlab file output =============#
printf "cphi = [ " >SIR_results.m;
for {n in pe diff{1}}
{
printf "%f ", 0.1083*(phi[nep*n,ncp]-phi[nep*(n-1),ncp]) >SIR_results.m;
}
printf "];\n\n" >SIR_results.m;

printf "phi = [ " >SIR_results.m;
for {n in pe diff{1}}
{
printf "%f ", P_PHI[n] >SIR_results.m;
}
printf "];\n\n" >SIR_results.m;

printf "infected = [ " >SIR_results.m;
for {n in fe}
{
for {m in cp}
{
printf "%f ", inf[n,m] >SIR_results.m;
}
}
printf "];\n\n" >SIR_results.m;

printf "cumuphi = [ " >SIR_results.m;
for {i in fe}
{
printf "%f ", phi[i,ncp] >SIR_results.m;
}
```

```
printf "];\n\n" >SIR_results.m;

printf "susceptible = [ " >SIR_results.m;
for {n in fe}
{
for {m in cp}
{
printf "%f ", sus[n,m] >SIR_results.m;
}
}
printf "];\n\n" >SIR_results.m;

printf "beta = [ " >SIR_results.m;
for {n in ye}
{
printf "%f ", beta[n] >SIR_results.m;
}
printf "];\n\n" >SIR_results.m;


printf "plot(phi,'r--','LineWidth',2); hold on;
plot(cphi,'LineWidth',2);xlabel('Finite elements (36 per year)');
ylabel('Infected Individuals');title('Infected Cases for Rho = ');
legend('Recorded Data','This work');" >SIR_results.m;
printf "\nfigure;plot(beta(1:length(beta)-1),'LineWidth',2);hold on;
plot(betatrue,'r--','LineWidth',2);xlabel('Finite elements (36 per year)');
ylabel('Beta value per day');title('Beta for Rho = ');
legend('This work','Previous results');" >SIR_results.m;


#====================-- end of the run file -===============#
```

The model *.mod file

```
# ====================================================================
# Dynamic optimization formulation for SIR disease model
# by George Abbott Fall 2009
#
# ====================================================================

# define indexes and general variables

param nfe >= 1 integer;
param ncp >= 1 integer;
param pfe >= 1 integer;
param fepy >= 1 integer;
param years >=0;
param meas >=0;

# define mathematical model parameters

param time;
param P_POP;
param P_GAMMA;
param P_RHO;
param inf_init_spec;
param sus_init_spec;
param life_exp;
param nep;


# define dimensions for all indexed variables

set fe := 1..nfe ;  # number of finite elements
set fewithzero := 0..nfe;
set cp := 1..ncp ;  # number of collocation points
set ye := 0..fepy ordered;  # number of finite elements per year
set pe := 1..pfe ;  # number of elements for phi measurements
#set mfe := meas..nfe;  # the specific finite elements for measure

param a{cp,cp} ;    # collocation matrix
param h{fe}    ;    # finite element length, this is defined in the .run file
param timedom{0..nfe*ncp} ; # total number of collocation points
param INF_MEAS{1..nfe*ncp};
param P_BETA{ye diff {0}};
param P_INF_INIT{fewithzero};
param P_SUS_INIT{fewithzero};
```

```
param P_BIRTHS{1..nfe*ncp};
param P_PHI{1..nfe};
param cpoints{cp};


# define the decision variables

var inf {fe,cp}  >= 0;
var sus {fe,cp}  >= 0;
var phi {fe,cp}  >= 0;
var beta{ye}     >= 0;
var dbeta{ye};
var betapos{ye}  >= 0 ,:=0;
var betaneg{ye}  >= 0 ,:=0;
var inf_init     >= 0;
var sus_init  >= 0;
var r_naught  >= 0;
var average_age  >= 0;



# states first order derivatives

var infdot{i in fe, j in cp} =  (beta[(1+(i-1)mod(fepy))-1]
+cpoints[j]*(beta[(1+(i-1)mod(fepy))]
-beta[(1+(i-1)mod(fepy))-1]))*inf[i,j]*sus[i,j]/P_POP - P_GAMMA*inf[i,j];
var susdot{i in fe, j in cp} =  -(beta[(1+(i-1)mod(fepy))-1]
+cpoints[j]*(beta[(1+(i-1)mod(fepy))]
-beta[(1+(i-1)mod(fepy))-1]))*inf[i,j]*sus[i,j]/P_POP + P_BIRTHS[i];
var phidot{i in fe, j in cp} =  (beta[(1+(i-1)mod(fepy))-1]
+cpoints[j]*(beta[(1+(i-1)mod(fepy))]
-beta[(1+(i-1)mod(fepy))-1]))*inf[i,j]*sus[i,j]/P_POP;

# collocation equations

fecolI{i in fe diff{1},j in cp}: inf[i,j] = inf[i-1,ncp]+
time*h[i]*sum{k in cp} a[k,j]*infdot[i,k];
fecolS{i in fe diff{1},j in cp}: sus[i,j] = sus[i-1,ncp]+
time*h[i]*sum{k in cp} a[k,j]*susdot[i,k];
fecolP{i in fe diff{1},j in cp}: phi[i,j] = phi[i-1,ncp]+
time*h[i]*sum{k in cp} a[k,j]*phidot[i,k];

fecolI0{j in cp}:       inf[1,j] = inf_init+time*h[1]*
sum{k in cp} a[k,j]*infdot[1,k];
fecolS0{j in cp}:       sus[1,j] = sus_init+time*h[1]*
```

```
sum{k in cp} a[k,j]*susdot[1,k];
fecolP0{j in cp}:       phi[1,j] = time*h[1]*
sum{k in cp} a[k,j]*phidot[1,k];

# Regularization of beta

defofdbeta{i in ye diff {0}}: dbeta[i] = beta[i] - beta[i-1];
equalityconst{i in ye}: dbeta[i] + betapos[i] - betaneg[i] = 0;
dbeta0: dbeta[0] = beta[0] - beta[last(ye)];
betacont: dbeta[0] = 0;



####### objective function...

minimize Error: (1/P_RHO)*sum{i in ye} (betapos[i] + betaneg[i])
+ sum {i in pe diff{1}} (0.1083*(phi[nep*i,ncp]-phi[nep*(i-1),ncp])
- P_PHI[i])^2/P_POP
;

#-- end of the *.mod file --
```

VITA

Name:	George Henry Abbott III
Address:	Artie McFerrin Department of Chemical Engineering Texas A&M University Jack E. Brown Engineering Building 3122 TAMU Room 200 College Station, TX 77843-3122
Email Address:	gabbottiii@gmail.com
Education:	B.S., Mathematics, University of North Texas, 2006 B.A., Music, University of North Texas, 2006 M.S., Chemical Engineering, Texas A&M University, 2010

This document was typeset in LaTeX by George H. Abbott III.