Estimation of seasonal transmission parameters in childhood infectious disease using a stochastic continuous time model

ARTICLE *in* COMPUTER AIDED CHEMICAL ENGINEERING · JANUARY 2010

DOI: 10.1016/S1570-7946(10)28039-2

20th European Symposium on Computer Aided Process Engineering – ESCAPE20
S. Pierucci and G. Buzzi Ferraris (Editors)

Estimation of seasonal transmission parameters in childhood infectious disease using a stochastic continuous time model

Daniel P. Word[a], James K. Young[a], Derek Cummings[b], Carl D. Laird[a]

[a]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, MS 3122 TAMU, College Station, TX 77843, USA,*
carl.laird@tamu.edu (corresponding author)
[b]*Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Rm E6541, Baltimore, MD 21205, USA, dcumming@jhsph.edu*

Abstract

The development of accurate disease models is desirable for the purposes of gaining a better understanding of the underlying dynamics of infectious disease spread and for designing and implementing appropriate control measures to curb infectious disease spread. In this work we develop an estimation framework for long-term continuous time infectious disease models that considers both model and estimation noise. We present a nonlinear programming approach for efficient estimation of model parameters, including seasonal transmission profiles. We then demonstrate the effectiveness of this framework using measles data from New York City and Bangkok, and show that a strong correlation exists between estimated seasonal parameters and school term holidays.

Keywords: Measles, continuous, disease models, seasonal transmission parameters

1. Introduction

One goal of public health programs is to control the spread of infectious diseases and minimize the impact of disease on the population through various control measures such as vaccination programs. However, there are several social, environmental, and biological factors affecting the spread of infectious disease, and the observed temporal dynamics are not always well understood. The development of reliable mechanistic models for the spread of infectious diseases is needed both for aiding public health decision-making and for improving our understanding of factors affecting infectious disease spread. Childhood infectious diseases, such as measles and chickenpox, remain a serious public health concern, especially in developing countries, and are commonly used as a test bed for developing disease models.

Compartment-based disease models are commonly used to describe the dynamics of the disease within the population. In these models the population is assumed to be well mixed with individuals placed into various compartments based on their status with respect to the disease. For example, individuals can be classified as being susceptible to the disease (S), infected but not infectious (E), infected and infectious (I), or recovered and immune (R). This type of model is typically classified according the progression of the population through the compartments [1].

Much work has been done to model measles incidence using discrete time models [2-4]. The time-series SIR (TSIR) model, introduced by Finkenstädt and Grenfell [4], is a discrete-time model that incorporates a seasonal transmission parameter and an exponential mixing in the infection term. This model is capable of capturing the biennial dynamics seen in cities with low birth rates, and it can quantitatively explain the annual cycle seen in measles incidence in cities with high birth rates [4-5]. The TSIR model and estimation procedure described for measles assumes a two-week reporting interval, which is similar to the serial interval for measles. If the reporting interval is different than the serial interval of the disease a different estimation procedure is needed. A continuous time model and estimation framework, on the other hand, can accommodate the common situation where the reporting interval and the serial interval are not similar. In previous work, we addressed the estimation of continuous time deterministic models for infectious disease spread with seasonal transmission parameters [7]. While deterministic models can reasonably capture incidence dynamics in large cities where the disease is endemic, infectious diseases are inherently stochastic in nature, and in communities below around 300,000 people stochastic fadeout of measles cases is commonly observed [6]. Deterministic models are incapable of capturing the disease dynamics in these cases. For these reasons, a model with both measurement and dynamic noise is desired. In this paper, we present a framework for estimation of continuous time infectious disease models from long-term time series data that considers both model and measurement noise. We demonstrate the effectiveness this approach using measles data from New York City and Bangkok.

2. Problem Formulation

The classic SIR framework model with a seasonal transmission parameter is used in our study. This model is sufficient to capture the key features of measles dynamics since life-long immunity is typically retained following infection. It has long been observed that measles incidence exhibits a seasonal pattern that appears to be correlated with school terms [3,9]. In the continuous time SIR model, we include a seasonal transmission parameter $\beta(y(t))$, also called the contact rate. Here, the function $y(t)$ maps the overall horizon time to the elapsed time within the current year. This forces β to have a yearly periodicity.

2.1. Stochastic Continuous Time Formulation

The differential equations describing the seasonal stochastic continuous time SIR model can be written as,

$$\frac{dS}{dt} = \frac{-\beta(y(t))S(t)I(t)}{N} + \mu(t)N + \varepsilon_S \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta(y(t))S(t)I(t)}{N} - \gamma N + \varepsilon_I \tag{2}$$

where S is the number of susceptibles and I is the number of infectives. System parameters include the birth rate, $\mu(t)$, which is known and time varying, and N and γ, the reported population and the recovery rate respectively, which are known scalar inputs. Dynamic noise terms ε_S and ε_I are included for the susceptible balance and the infective balance equations respectively.

Prevalence refers to the number of individuals in the population who are infected at a given point in time, whereas incidence is the number of new infectious occurring over a given time interval. The available reported case data is measles incidence, but the state variable $I(t)$ represents the measles prevalence. Over a particular reporting interval, *i*, the incidence can be calculated by integrating the rate of infection,

$$\int_{t_{i-1}}^{t_i} \frac{\beta(y(\lambda))I(\lambda)S(\lambda)}{N} d\lambda. \tag{3}$$

To include this in the estimation formulation, a new state variable $\phi(t)$ is introduced to represent the cumulative incidence at time *t*.

$$\frac{d\phi}{dt} = \frac{\beta(y(t))S(t)I(t)}{N} \tag{4}$$

Since not every individual who becomes infected will seek medical assistance, not every infection is reported. This underreporting can be significant and must be considered in the estimation. The output equation for the reported cases is given by,

$$\Phi_i = \eta_i\left(\phi_i - \phi_{i-1}\right) + \varepsilon_\phi \tag{5}$$

where $\varepsilon\phi$ is the measurement noise term, ϕ is the estimated cumulative incidence at a point in time, Φ_i is the reported incidence over a given time interval, and η_i is the time varying reporting factor that accounts for the degree of underreporting. We use a standard susceptible reconstruction procedure to estimate this reporting factor.

The estimation problem can be written as a nonlinear programming problem with the differential and algebraic constraints described above. There are a number of techniques for solving this class of optimization problem. Here, we discretize all the state and algebraic variables and include the complete set of discretized equations as constraints in the nonlinear programming problem. The Explicit Euler technique was used to discretize the system, however, we have used Radau collocation techniques with similar results. Without further restriction of $\beta(y(t))$, this estimation problem has all the challenges of classic inverse problems including ill-conditioning and non-uniqueness. The seasonal parameter $\beta(y(t))$ was discretized less finely than the differential equations and the profile was regularized. Total variation regularization is used since it allows for discontinuous jumps as expected from a seasonal transmission parameter correlated with school term holidays. Combining the regularization term with a least-squares objective for the noise terms and initial state conditions gives the following objective function,

$$\min\ a\sum_{i\in\Im}\varepsilon_{Si}^{\ 2} + b\sum_{i\in\Im}\varepsilon_{Ii}^{\ 2} + c\sum_{i\in\Re}\varepsilon_{\phi i}^{\ 2} + d\left(S_0 - S_{init}\right)^2 + e\left(I_0 - I_{init}\right)^2 + \frac{1}{\rho}\left\|\Delta\beta\right\|_1 \tag{6}$$

where $\Im$ is the set of finite elements, $\Re$ is the set of reporting intervals, *a*, *b*, and *c* are weighting terms based on assumed standard deviations of the residuals, and *d* and *e* are weights placed on the residuals of the initial conditions. In the regularization term, $\Delta\beta$ is a first order approximation of $d\beta/dt$, and ρ is the regularization parameter that is calculated using the standard L-curve method [10]. As written, the regularization term is non-differentiable, however, it is easily reformulated by writing,

$$\frac{1}{\rho}\left\|\Delta\beta\right\|_1 = \frac{1}{\rho}\sum_{j\in B}\Delta\beta_j^+ + \Delta\beta_j^- \tag{7}$$

and including the following constraints,

$$\Delta\beta_j - \Delta\beta_j^+ + \Delta\beta_j^- = 0 \quad (8)$$

$$\Delta\beta_j^+ \text{ , } \Delta\beta_j^- \geq 0 \;\; \forall\, j \in B \quad (9)$$

where B is the set of discretizations for β within the year. This reformulated objective function with constraints (8) and (9), along with the constraints arising from the discretization of (1), (2), and (4), and the reporting factor adjustment (5) give rise to a large-scale nonlinear programming problem with purely algebraic constraints.

2.2. Data

The data sets used in this work contain yearly population and birth rate data, monthly measles case count data from New York City from the years 1947-1965 [11], and monthly measles case count data from Bangkok for the years 1975-1984. In the Thai data, there is regular passive surveillance for measles coupled with active surveillance to assess the performance of the passive surveillance system. All data is anonymized, and laboratory confirmation is reported when available. These two locations have very different school term holidays, allowing us to show the correlation between school terms and seasonal transmission. New York city has a long summer school holiday lasting from the end of June to mid September, while Bangkok has two long school holidays: one from the beginning of March to the end of April and one the entire month of October.

An additional challenge in the Bangkok data is missing information for the year 1979. To account for this, our model is integrated through this period, however the estimation is weighted to exclude these points from the objective function. Both data sets suffer from significant under-reporting. Susceptible reconstruction techniques have shown that about 1 in 9 cases are reported in New York City across the entire time horizon studied. Bangkok, however, has a varying reporting fraction that for this work is assumed to be linear over the time horizon and ranging from about 1 in 100 cases reported at the start of the time horizon to about 1 in 20 cases at the end.

3. Estimation Results

Estimations were performed for New York City and Bangkok using 240 finite elements per year and 60 discretizations of the seasonal transmission parameter per year. The problems were formulated in AMPL [12] and solved using the nonlinear interior-point method IPOPT [13].

3.1. New York City

The estimation for New York City produced essentially zero mean profiles for all model and measurement noise terms. The characteristics of these estimated noise terms are shown in Table 1. The model noise in the susceptible balance equation (ε_S) however, showed an apparent correlation in time. This would indicate that the susceptible dynamics are not being appropriately captured with this simple model, and future work will address improvements in this area. Nevertheless, estimated mean percentage of susceptibles in the population was 4.3%, which is similar to other literature values for measles [8].

Table 1. Noise terms for New York City

Residuals	ε_S (1/day)	ε_I (1/day)	ε_ϕ (1/day)
Mean	-14.973	-0.896	-0.477
Stand. Dev.	172.331	41.076	10.478

Fig. 1 shows the estimated seasonal transmission profiles, β. The profile shows seasonality that coincides almost perfectly with the school term summer holiday that occurs from the end of June to mid September.

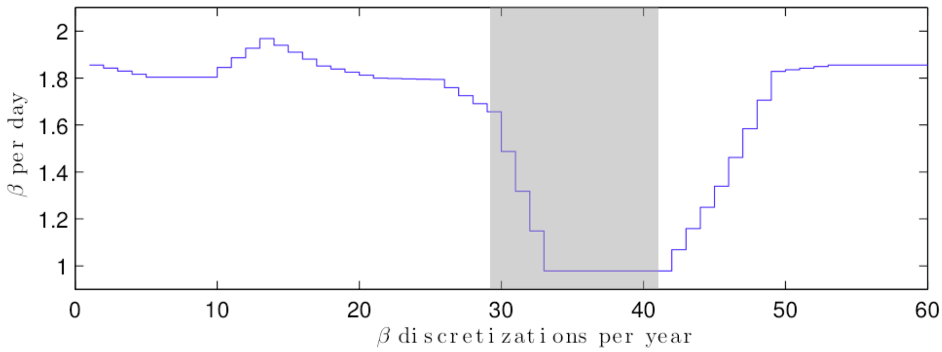

Figure 1: Seasonal transmission parameter for NYC with the school holiday indicated in grey

3.2. Bangkok

The estimation for Bangkok gave results similar to the New York City estimation. The reporting fraction is very low for Bangkok, and the data contains significant noise, but the model and measurement noise terms all have near zero mean as shown in Table 2. Again, the model noise corresponding to the susceptible balance equation do not appear to be independent in time, however, the estimated mean percentage of susceptibles in the population was still 4.4%.

Table 2. Noise terms for Bangkok

Residuals	ε_S (1/day)	ε_I (1/day)	ε_ϕ (1/day)
Mean	-14.262	-0.480	-0.104
Stand. Dev.	225.630	24.386	1.902

The estimated profile for the seasonal transmission parameter is shown in Fig. 2. The profile shows strong agreement with the school holidays that occur from the beginning of March through the end of April and the whole of October, although a delay of approximately one month is apparent. This may be due to delayed reporting or an artifact of the long reporting interval.

4. Conclusions

The usefulness of reliable disease models for further understanding the dynamics of infectious diseases and planning public health policy is apparent. The model described here appears to capture the dynamics of measles effectively for two diverse cities, and the continuous time formulation allows us to make immediate use of the data with larger reporting intervals. The estimated transmission parameter profiles for both cities demonstrate strong seasonality correlated with school term holidays despite the very different school holiday schedules for these locations. These estimations not only help improve our understanding of infectious disease spread, but also help quantify the effect of closing schools, a commonly proposed control measure for emerging infectious diseases.

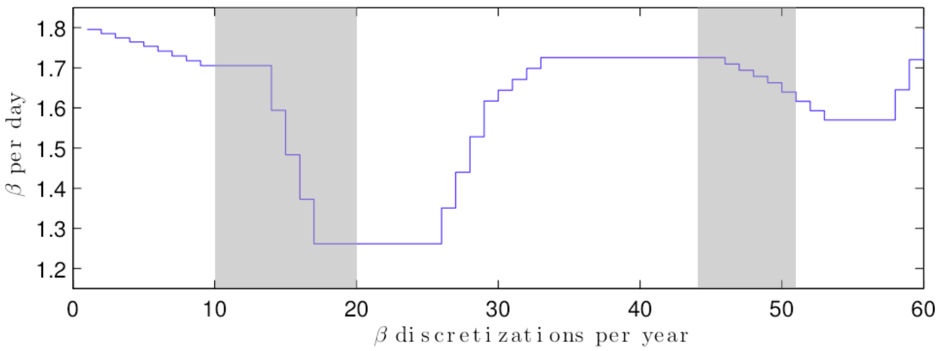

Figure 2: Seasonal transmission parameter for Bangkok with school holidays indicated in grey

For large cities where diseases are endemic, deterministic models can effectively reproduce the observed dynamics, however, in smaller communities were stochastic fadeout is evident, stochastic models are necessary. Future work will include a thorough analysis of the estimation framework on smaller community sizes. Furthermore, the approach is suitably efficient and flexible, and future work will investigate more complex model structures. Continued study will improve our understanding of the system, improve prediction, and improve our ability to control endemic and emerging infectious diseases.

5. Acknowledgements

We thank Dr. Sopon Iamsirithaworn and the Thai Ministry of Public Health for contributing the Bangkok measles data used in this work.

References

[1] H.W. Hethcote, 2000, The mathematics of infectious diseases, SIAM Review, 42, 4, 599-653
[2] W.H. Hamer, 1906, Epidemic Disease in England, Bedford Press
[3] P.E. Fine, J.A. Clarkson, 1982, Measles in England and Wales I: An Analysis of Factors Underlying Seasonal Patterns, Int. J. Epidemiology, 11, 1, 5-14
[4] B.F. Finkenstädt, B.T. Grenfell, 2000, Time series modelling of childhood diseases: a dynamical systems approach, J. Royal Stat. Soc., Ser. C, 49, 187-205
[5] A.J.K. Conlan, B.T. Grenfell, 2007, Seasonality and the persistence and invasion of measles, Proc. R. Soc. B, 274, 1133-1141
[6] M.J. Keeling, 1997, Modelling the Persistence of Measles, Trends Microbiol., 5, 12, 513-518
[7] G.H. Abbott III, D.P. Word, D. Cummings, C.D. Laird, 2009, Estimating Seasonal Drivers in Childhood Infectious Diseases with Continuous Time and Discrete-Time Models, submitted to the 2010 American Control Conference (invited)
[8] RM. Anderson, R.M. May, 1991, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press
[9] H.E. Soper, 1929, The interpretation of periodicity in disease prevalence, J. Royal Stat. Soc., 92, 34-73
[10] R.C. Aster, B. Borchers, C.H. Thurber, 2005, Parameter Estimation and Inverse Problems, Elseiver Academic Press
[11] J.A. Yorke, W.P. London, 1973, Recurrent outbreaks of measles, chickenpox, and mumps, Am. J. Epidemiology, 98, 6, 469-482
[12] R. Fourer, D.M. Gay, B.W. Kernighan, 1993, AMPL: A Modeling Language for Mathematical Programming, The Scientific Press
[13] A. Wächter, L.T. Biegler, 2006, On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming, Math. Programming, 106, 25-57