AN EFFECTIVE FRAMEWORK FOR THE MODELING OF SEASONAL PARAMETERS FOR MEASLES USING A CONTINUOUS TIME MODEL

A Senior Scholars Thesis

by

ALI MAHMOUD EL-HALWAGI

Submitted to Honors and Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

May 2012

Major: Chemical Engineering

AN EFFECTIVE FRAMEWORK FOR THE MODELING OF SEASONAL PARAMETERS FOR MEASLES USING A CONTINUOUS TIME MODEL

A Senior Scholars Thesis

by

ALI MAHMOUD EL-HALWAGI

Submitted to Honors and Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by:

Research Advisor:	Carl Laird
Associate Director, Honors and Undergraduate Research:	Duncan MacKenzie

May 2012

Major: Chemical Engineering

ABSTRACT

An Effective Framework for the Modeling of Seasonal Parameters for Measles Using a Continuous Time Model. (May 2012)

Ali Mahmoud El-Halwagi
Department of Chemical Engineering
Texas A&M University

Research Advisor: Dr. Carl Laird
Department of Chemical Engineering

Currently, measles (also known as English measles) is detrimentally affecting the youth of the developing world. The shortage of vaccines and the devastating consequences of measles yield disturbing consequences globally. According to the World Health Organization, “there were 164,000 measles deaths globally – nearly 450 deaths every day or 18 deaths every hour.” In this work an approach was developed for estimating an appropriate continuous time seasonal transmission model for the spread of measles. Estimation of this time-varying parameter is an ill-posed problem and regularization is required. Subsequently, data obtained from New York City over a twenty year period were used to formulate the problem for estimating values of the seasonal transmission parameter in New York City. Finally, the estimation problem was solved and the results were analyzed. This research shows that there is a clear relationship between school terms and seasonal parameters.

DEDICATION

I would like to dedicate this thesis to my amazing family for their unconditional love and support. I am eternally grateful.

ACKNOWLEDGMENTS

I would first and foremost like to acknowledge my family for their unwavering support and unconditional love. I would like to thank my brother Omar and my mom for continuously motivating me to do my best. I would especially like to thank my dad for inspiring me to become a Chemical Engineer and for always being my role model. His support means the world to me. I would also like to thank Dr. Carl Laird for giving me the opportunity to work under his guidance and for proposing the topic. He has been a remarkable advisor and mentor. I am also very thankful to Daniel Word for all of his help and patience. Most importantly, I would like to thank God for providing me with the strength necessary to complete this thesis.

NOMENCLATURE

β	Seasonal Transmission Parameter
ρ	Regularization Term
t	Time
R_0	Reproduction Parameter
$N(t)$	Number of People in Population
$S(t)$	Number of Susceptible Individuals
$I(t)$	Number of Infective Individuals
$R(t)$	Number of Recovered Individuals

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Currently, measles (also known as English measles) is providing a detrimental effect on the youth of the developing world. The shortage of vaccines and the devastating consequences of measles yield disturbing consequences globally. According to the World Health Organization, “there were 164,000 measles deaths globally – nearly 450 deaths every day or 18 deaths every hour.” There is strong motivation to investigate the transmission rate of measles through mathematical modeling to get a better and more quantitative sense of how measles spreads and how it might be controlled.

Potential benefits

My work has many necessary goals. First of all, infectious diseases prove to be fascinating mechanisms because the spread of each disease is not only dependent on the characteristics of the disease, but also on the geography of the disease in question. For example, the seasonal transmission parameter profiles for developed locations like New York City may be different than those in lesser-developed regions like Thailand. Estimation of model parameters provides us insight into the driving forces affecting the spread of the disease.

This thesis follows the style of Chemical Engineering Science.

Furthermore, the work may provide significant public health care benefits by guiding policy decisions. First of all, having a reliable dynamic model will better predict future outbreaks which will allow for better preparation of the health care system. This will also encourage stronger and more direct health care measures to pinpoint the major factors affecting the spread of measles. Moreover, the creation of a continuous model will take into account the previous trends of the spread of measles in a particular area which will build a more pertinent and effective framework for determining future measles outbreaks.

Uniqueness

This work will prove to be most unique in the following two manners: it focuses on estimation using a continuous time model rather than the structurally more simplistic discrete time model, and it applies standard techniques for ill-posed problems to remove the periodicity requirement inherent in other estimation techniques.

As mentioned earlier, the use of discrete time models has taken center stage in recent disease model estimation research. The normal viewpoint of a time model utilizes $N(t) = S(t) + I(t) + R(t)$, where $N(t)$ is the total population size, $S(t)$ is the number of susceptible individuals, $I(t)$ is the number of infective individuals, and $R(t)$ is the number of recovered individuals, all at any given point in time (Anderson and May, 1991). The above model can then be used to include other time-varying parameters, including fertility and mortality rates. As more children are born, the total population size increases and these newborns directly enter the susceptible compartment.

Models describing the spread of infection are dependent on both the number of infected and the number of susceptible individuals. This concept was first examined in detail by Hamer who showed that the number of new infections per unit time is proportional to the number of susceptibles times the number of infectives (Hamer, 1906). Essentially, more people contract a particular disease if there are more people who have the disease to spread and more people who are susceptible to the disease. The value of the proportionality constant is dependent on the disease itself and other social and environmental factors. Furthermore, several researchers have demonstrated that the proportionality constant is correlated in time. Thus, a time-varying parameter must be introduced to the model. Considering a discrete time model, the parameter, β_t, is the transmission parameter that affects the rate of new infections between times t and t+1 (Oli et. al., 2006).

However, the discrete time model contains particular inherent flaws with respect to continuous time models (Abbott III et. al., 2009). First of all, infectious diseases like measles are in fact continuous processes. Additionally, continuous time models allow for a more flexible estimation procedures applicable to various sets of data while discrete time procedures that require the reporting interval matches the time discretization (Finkenstadt and Grenfell, 2000).

A more general, applicable approach is necessary for real world data. Models and estimation procedures should be flexible to allow for the variations of particular data, including the serial interval of the disease and the reporting interval of the data itself (Abbott III et. al., 2009). Thus, I will use a continuous time model to estimate the seasonal transmission parameters governing the spread of measles.

Currently, the biggest criticism involved with continuous time models is the difficulty in estimating the transmission rate parameter, β, in the model (McCallum et. al., 2001). Its difficulty lies in the fact that the transmission rate is in fact continuous but all data occur at particular intervals. This means that there are, in essence, an infinite number of parameters, but a finite number of data points, producing an ill-posed problem. There are two methods to handle this. One can select a particular functional form for the transmission parameter and estimate parameters in that function. However, this restricts the form of the transmission parameter profile and results could be biased by our assumptions of that form. Instead, the concept of regularization is used. Regularization produces a well-posed estimation problem while only mildly restricting the form of the transmission parameter. Here, total-variation regularization is used for the transmission parameter.

On a similar note, continuous models have an advantage over discrete models in lesser-developed areas where less data may be available. The discrete time series SIR model includes an extra parameter that may correct for the errors in discrete time approximation. The continuous time model does not require this parameter. Thus, a continuous model would have one less parameter to estimate (Xia et. al., 2004).

As important as it is to model measles, more importance should be placed on the transmission of measles throughout different periods of the year. Previous work has been done on estimating the measles transmission rate parameter given monthly incidence data; in fact, Soper showed that there was a clear relationship between the transmission rate parameter and school terms (Soper, 1929). This concept makes intuitive sense; one would assume that as children are in school together, they are more likely to get each other sick than when outside of school. In another example, Fine and Clarkson examined the time-varying parameters for England and Wales from 1950 – 1965 and again demonstrated the aforementioned correlation between beta and school terms (Fine and Clarkson, 1982). This work makes use of regularization to estimate the time-varying transmission profile with few assumptions on its particular form. We believe that the transmission profile will be correlated with school term holidays.

Another crucial output from the continuous time model is R_0, the basic reproduction number. This parameter shows the average number of people who get infected due to one infected person. Basically, it measures secondary infections. This concept is essential in determining whether a particular disease will survive and thrive. If $R_0 > 1$, the disease has the devastating potential of invading and thriving in a particular population.

Some properties of measles (and other common childhood diseases) make them effective case studies for modeling and estimation. First, it is important to note that the reported case counts in the data under-represent the actual number of cases (some children will not be taken to the hospital, some hospitals will not report, etc.). However, prior to widespread vaccination, almost every individual in the population would eventually contract the disease. This allows us to perform a time-averaged balance between births and case counts to approximate the reporting fraction.

CHAPTER II

METHOD

The following general methodology is used to estimate the disease model (in this case, measles). Table 1 discusses the four major steps required to build an accurate disease model formulation.

Table 1- Methodology for Measles Modeling

Develop an accurate and applicable model formulation
↓
Implement regularization to produce a well-posed estimation problem
↓
Prepare the case count data and include in the estimation
↓
Solve and Examine Solution

Step 1: Developing an accurate model

Each of the four steps listed above is crucial in creating a reliable model. The first step requires first defining the important variables in the system. In the case of measles, the important variables include S(t), R(t), and I(t). The estimation formulation requires an objective function and, here, we minimize the squared-error. We utilizes AMPL, a mathematical modeling language that takes in an objective function, various parameters, variables, sets, and constraints.

Step 2: Implementing regularization

To effectively estimate the profile of the transmission parameter β, the effects of the parameter ρ (regularization term) must be examined. For a particular set of data, there is a value of ρ that appropriately balances the estimation error with the restriction on the transmission profile. This value is found using the L-curve method which graphs part of the objective function versus the total variation regularization term. Every L-curve has a corner, hence the name L-curve. The value of ρ that corresponds to the corner will be used throughout the estimation. We carefully

Step 3: Preparing the data

For this study, the data were obtained from New York City over a twenty year period. This is to examine the behavior of the seasonal transmission parameter in New York City. A twenty year period was chosen. The data needed to be entered from original documents, a susceptible reconstruction procedure was done to estimate the level of underreporting. Then, the corrected values were transferred to AMPL dat files.

Step 4: Solve and examine the solution

After formulation in AMPL, the problem was solved using the nonlinear interior-point solver IPOPT. This is an open-source package available for solution of large-scale nonlinear programming problems.

After solving the estimation problem, we investigate the temporal profile of the transmission parameter. We hypothesize that our solution will show that there is a clear correlation between school terms and the seasonal transmission parameter. Moreover, our hypothesis includes that the measles is spread more rapidly during the school year than during the summer months, providing crucial data on when to supply vaccines or prepare health care facilities.

CHAPTER III

RESULTS

Regularization is absolutely necessary to form a well-posed estimation problem. However, care must be taken when selecting the regularization parameter. In this chapter, the methodology of determining the proper regularization term will be examined and the appropriate value in this model will be shown.

The objective function includes the regularization parameter $1/\rho$ as a factor in front of the regularization term (based on total variation). The methodology used to determine the correct ρ value is as follows.

The process to determine ρ is known as the L-curve method as shown in Figure 1. The L-curve graphs optimal values of the objective function against the value of the regularization term for different values of the regularization parameter. The corner of the graph indicates a good balance between fitting the data and overly restricting the shape of the transmission parameter profile. Thus, too high or too low ρ values will produce a strictly horizontal or vertical graph; with no true curve, the proper ρ value cannot be determined.

For our estimation problem, the regularization parameter values were varied from 1e-4 to 1e-6 by 1e-7, and the following graph was obtained.

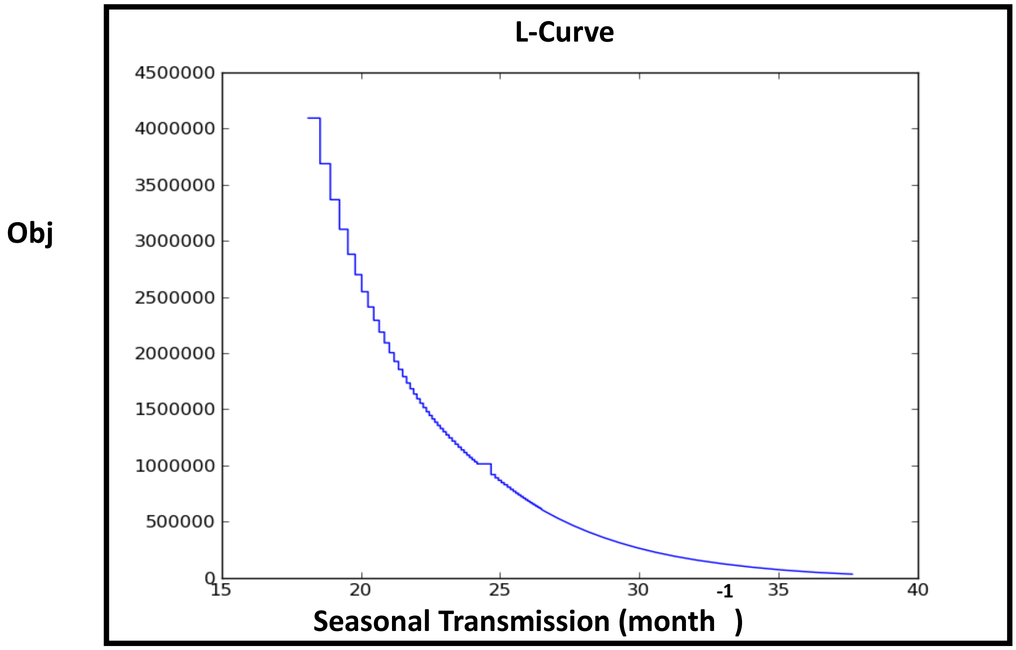

Fig. 1 L-Curve

By examining the L-curve, it is evident that the corner exists at ~24-25 on the x-axis. This value corresponds to a regularization parameter value of $6.5*10^{-6}$.

With this newly found regularization value, the formal estimation problem can now be solved and the estimated seasonal transmission parameter profiles can be examined. Figure 2 shows how the seasonal transmission parameter varies every month for a twenty year period. From this figure, a clear relationship exists between the time of the year and school transmission parameter since β oscillates annually, hitting its lowest value in the summer.

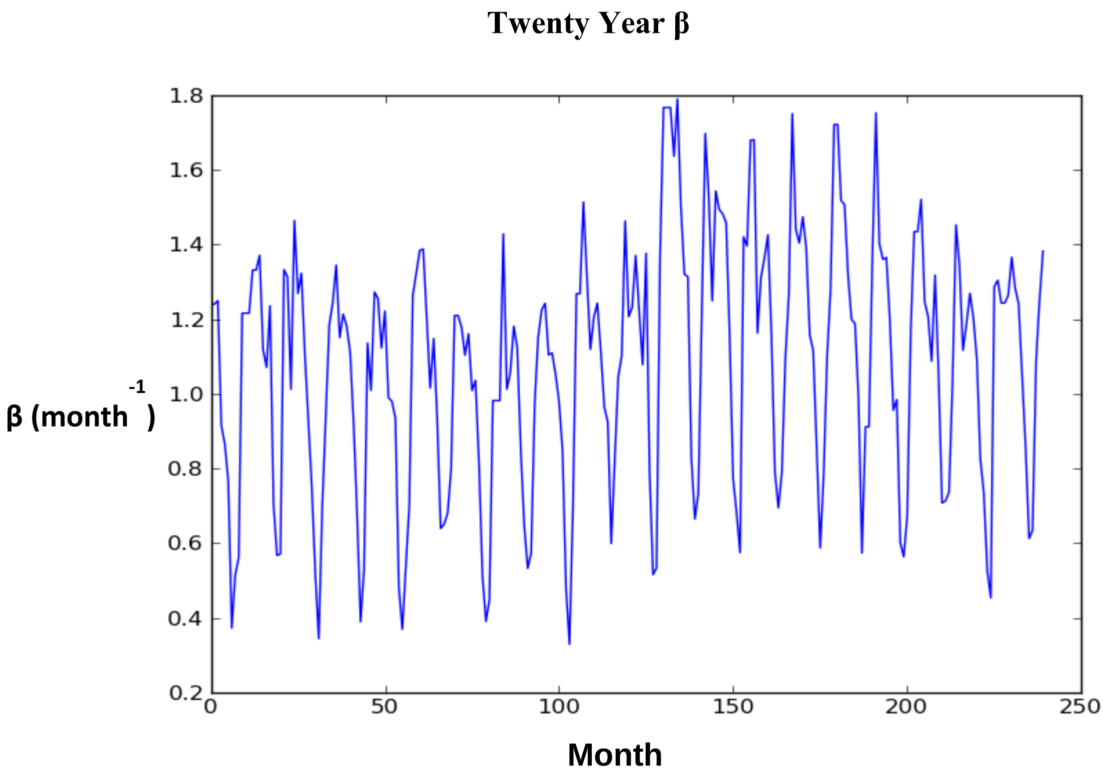

Fig. 2 Twenty Year β

To show that β exhibits similar behavior from year to year, β was graphed annually as shown below in Figure 3. Figure 3 demonstrates how β starts off large at the beginning of the year, hits its lowest point in the summer, and returns to its highest value near the end of the year.

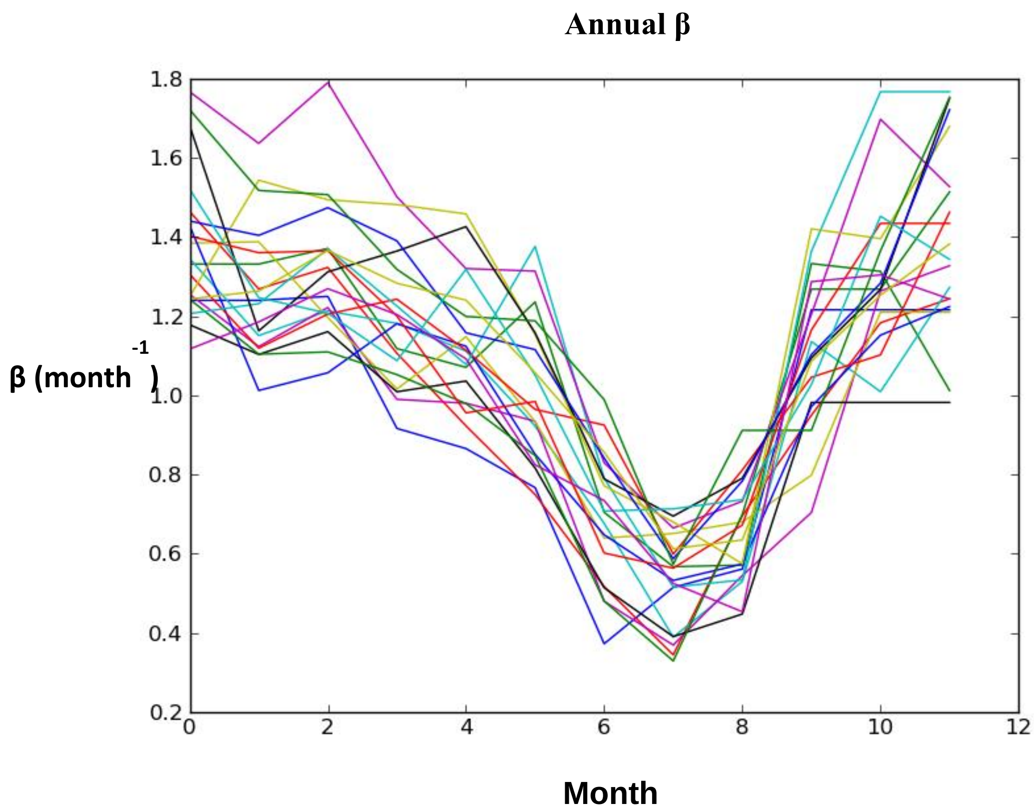

Fig. 3 Annual β

If the value of the regularization parameter was too low, then variation in the transmission parameter would be heavily penalized and the curves would be significantly flatter. However, if the value of the regularization parameter is too high, then the problem is ill-posed and the estimated transmission parameters would be very noise. By properly utilizing the L-curve, the appropriate regularization value is determined, and the seasonal transmission profiles are efficiently estimated. This entire procedure is general, and it can be applied to other data sets from alternate regions.

CHAPTER IV

SUMMARY AND CONCLUSIONS

After developing a procedure for estimation of time-varying transmission profiles, the method was applied to New York City data over a twenty year period. Many conclusions can be drawn. First of all, this work has demonstrated the power available in large-scale nonlinear programming techniques and the ability of these techniques to estimate time-varying parameters like β. More specifically, these results showed a strong correlation between the seasonal transmission parameter and school term holidays as estimated from New York City data. These results are consistent over the twenty year period.

Since many countries lack central health care facilities and the proper tools to continuously care for the sick, determining the optimal time to give measles vaccines is crucial. From the results above, the seasonal transmission parameter was at its lowest each year during the summer time. Thus, we propose giving vaccines toward the end of the summer to prevent the large spike in measles transmissions that occur at the beginning of the school year. This will limit the number of infectives and mitigate the devastating effects of measles worldwide.

REFERENCES

Abbott, G.H., Word, D.P., Cummings, D., Laird, C.D., 2009. Estimating seasonal drivers in childhood infectious diseases with continuous time and discrete-time models. In: Proceedings of American Control Conference, Baltimore, MD, June.

Anderson, R.M., May, R.M., 1991. Infectious diseases of humans: dynamics and control. Oxford University Press Inc., New York.

Fine, P.E., Clarkson, J.A., 1982. Measles in England and Wales I: an analysis of factors underlying seasonal patterns. International Journal of Epidemiology, 11(1):5-14.

Finkenstadt, B.F., Grenfell, B.T., 2000. Time series modeling of childhood diseases: a dynamical systems approach. Journal of the Royal Statistical Society, 49(2):187-205.

Hamer, W.H., 1906. Epidemic disease in England. Bedford Press, London.

McCallum, H. Barlow, N.D., Hone, J., 2001. How should pathogen transmission be modeled? Trends in Ecology & Evolution, 16(6):295-300.

Oli, M.K., Venkataraman, M., Klein, P.A., Wendland, L.D., Brown, M.B., 2006. Population dynamics of infectious diseases: a discrete time model. Ecological Modelling, 198:183-194.

Soper, H.E., 1929. The interpretation of periodicity in disease prevalence. Journal of the Royal Statistical Society, 92(1):34-73.

Xia, Y. Bjornstad, O.N., Grenfell, B.T., 2004. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. The American Naturalist, 164(2):267-281.

CONTACT INFORMATION

Name:	Ali Mahmoud El-Halwagi
Professional Address:	c/o Dr. Carl Laird Department of Chemical Engineering MS 236 Texas A&M University College Station, TX 77843
Email Address:	ali.elhalwagi@gmail.com
Education:	B.S., Chemical Engineering, Texas A&M University, May 2012 Undergraduate Research Scholar