

# Measles

*Clifford*

*December 13, 2016*

## INTRODUCTION AND LITERATURE REVIEW

### Introduction

Then the Lord said to Moses and Aaron, “Take handfuls of soot from a furnace and have Moses toss it into the air in the presence of Pharaoh. It will become fine dust over the whole land of Egypt, and festering boils will break out on men and animals throughout the land.”—Exodus 9:8–9.

Mankind’s woes with infectious diseases is one of the few well documented facts in history, as it has greatly imparted on civilizations. Now, Public health personnel, Stakeholders and Society as whole invest resources to study and understand the dynamics of epidemics. Mathematical tools have been used to describe the dynamics of infectious diseases since Daniel Bernoulli’s smallpox model in 1766. The most popular of these are Mechanistic models, in which assumptions are made about the nature of the underlying disease dynamics, based on research and observations. With these assumptions we often have a system of equations, most often these are differential equations. This has been to great success. The eradication of small pox testifies to this fact.

### Mathematical Models

The role Mathematical models play in studying infectious disease dynamics cannot be overstated. The most common of these Mechanistic models are the compartmental models. Compartmental models are defined as systems in which there are flows of material between units called compartments (Jacquez and Simon, n.d.). These compartments partition the population under study at any given time. For each compartment, its elements are time homogeneous (local homogeneity). Its governed by local mass balance, i.e. everthing going out of one compartment will go into the next compartment (Jacquez and Simon, n.d.). In the epidemiological setting, these are models that group hosts, based on their states at infection (Mishra, Fisman, and Boily 2010). Lets take for example tuberculosis, the stages at infection here include; people who do not have the disease, those who have but are not yet spreading the disease, those who have and are spreading the disease, and those who have recovered from the disease. More generally, the most common groups (compartments) are the Susceptible, Exposed, Infectious, Recovered/Removed.

- Susceptible(S): This is the state at which the host can be, but isn’t infected.
- Exposed(E): This is the state at which the host is infected, but not infectious.
- Infectious(I): This is the state at which the host is both infected and infectious
- Recovered/Removed(R): This is the state at which the host is either recovered from the disease, immune or has been isolated. There can be many of these compartments in a model and also serveral additional compartments like that of the vaccinated (V) etc. depending on the assumed disease dynamics.

Assumptions are made about the transition(flow) of hosts from one compartments to the other, given these assumptions we have a system of differential equations whose solution are to reflect the disease dynamics. Sometimes after the development of the model, the realisations can be incompatible with observation, this gives strong indication of a major violation of an assumption, and this can be investigated and rectified, or the model might just not be adequate to describe the disease dynamics . For most of these compartments, the flow out is assumed to be proportional to the size of the compartment. That is if  $K$  is a compartment and  $r$  is the rate of movement out of  $K$ , then  $K' = -rK$  , which implies that  $K(t) = \exp\{-rt\}$ . Thus  $1/r$  is the average length of time in  $K$ . Movements into the compartment can be captured by adding terms to the right hand side of the differential equation.(Hethcote 2000)

However, for the susceptible compartment we can have more complicated outflows, since this is related to infection. By nature, Transition Dynamic Models have force of infection as dependent on prevalence (Mishra, Fisman, and Boily 2010) (The force of infection is the per capita incidence rate of infection per susceptible host). This most often leads to non-linearity in transmission. There are two popular forms of incidence; Standard Incidence and Mass Action.

1. Standard Incidence: Here we assume that the average rate of contact  $\beta$  is constant. Thus by deduction our incidence is  $(\beta SI)/N$ . Where  $N$  is the population size.
2. Mass Action: We assume that the average rate of contact is proportion to the population size  $N$ . This also gives us incidence as  $\beta SI$ . (Hethcote 2000)

Assumptions do not only help in model formulation, they also help in simplifying the model. Even with that, most of the differential equations encountered in epidemiology are complex enough so that, it is difficult to find an analytic solution. However, without the analytic solution it is still possible to study the system using qualitative methods and simulations. One of the main focal points of interest are bifurcation points. These are simply points at which before and after, the system changes its dynamics. The Basic Reproductive number  $R_0$  is such a threshold point. This is the average number of secondary infections from the introduction of an infective into a completely susceptible population. When the Basic reproductive number is less than 1 [greater than 1], we have the disease-free equilibrium (the state of disease eradication) to be asymptotically stable [unstable]. This Bifurcation parameter is important because, through interventions like vaccination, we can change the magnitude of this parameter. And if we are able to reduce it to less than one, as in with smallpox then we are going to have the disease on decline and eventually eradicated.

## Childhood diseases

In quest to adequately studying the dynamics of infectious diseases, childhood diseases give a rare opportunity. Not only are its dynamics interesting as we'd later see, but the rich source of data available, helps us evaluate existing models and build mathematical and statistical tools to improve them and also, fit these models to new data to test their predictive ability. One such well documented childhood disease is measles (*Black 1984; Anderson & May 1991*). Measles is a childhood disease with well-studied dynamics. Its relatively high infection rate, meant that in the pre-vaccination period, almost everyone was infected by age 20 (Quadros 2004). This made it of interest to stakeholders in most countries, hence a rich data is available on measles for most developed countries (Bjornstad, Finkenstadt, and Grenfell 2002). Measles is also interesting to epidemiologists because of its relatively simpler infection and recovery (*Black 1984*), and it also exhibits both episodic and endemic dynamics (*Barlett 1957*).

Before we discuss and compare the existing models for measles dynamics, let's take a look at the biology and history of the disease.

## Measles

Measles is an infectious childhood disease caused by the rubeola virus of the Morbilli group of viruses. The mode of infection can be airborne (coughs and sneezes from an infected person), and also through direct contact to infected or exposed objects. The virus is infectious when it comes into contact with nose, mouth, eyes and skin. This makes it highly contagious. 7-18 days after exposure, symptoms are manifested. Thus the length of the exposed period is usually 8-10 days. The length of the infectious period is 8 days and this is between 4 days after the rash starts till 4 days after the rash. Other symptoms are:

- Fever anorexia, cough and irritability
- Watery and red eyes with running nose.

People often recover fully from measles in 2 weeks, though it sometimes lasts for 3 weeks. Measles on its own doesn't have disease related fatalities, however it can be very fatal to infants, people with weakened immune

system from HIV and Malnutrition. 157700 people are recorded to have died from measles in 2011, combined with it been the leading cause of death with in children show how fatal it can be.

There is no cure for measles, as stated above most people recover naturally from it. However, there is a very effective vaccine with a vaccine efficacy between 95% to 98% for people 6 months and older, since maternal immunity transfered to babies can interfere with vaccine efficacy (Quadros 2004). Yet, measles has made a comeback though it was eradicated from certain regions like the Americas. This is because not everyone is and can be vaccinated:

1. Some people have vaccine allergies
2. Infants can be too young to be vaccinated
3. HIV and other disease can weaken the immune system of host, and this makes vaccination life threatening
4. A much recent campaign by “Anti-Vaxxers” against vaccination because it is rumored to cause autism in infants.

Measles is very infectious; it can take just one infectious person to start an outbreak. An infected person has a 90% chance of infecting unvaccinated people (i.e. up to 12 to 18 person). However, due to the fact that measles vaccination is almost perfect, vaccination below a critical value for a region is enough to attain herd immunity. This has been effective for the Region of the Americas. The dynamics of the disease can be split into three different time frames. The pre-vaccination, transient and post vaccination periods. Though the dynamics of the disease is well studied, what is well understood is the pre-vaccination and post vaccination dynamics. Most models developed for these two periods fit well to data. However, the same cannot be said about the transient period. Below is a plot of the number of reported cases during the period.

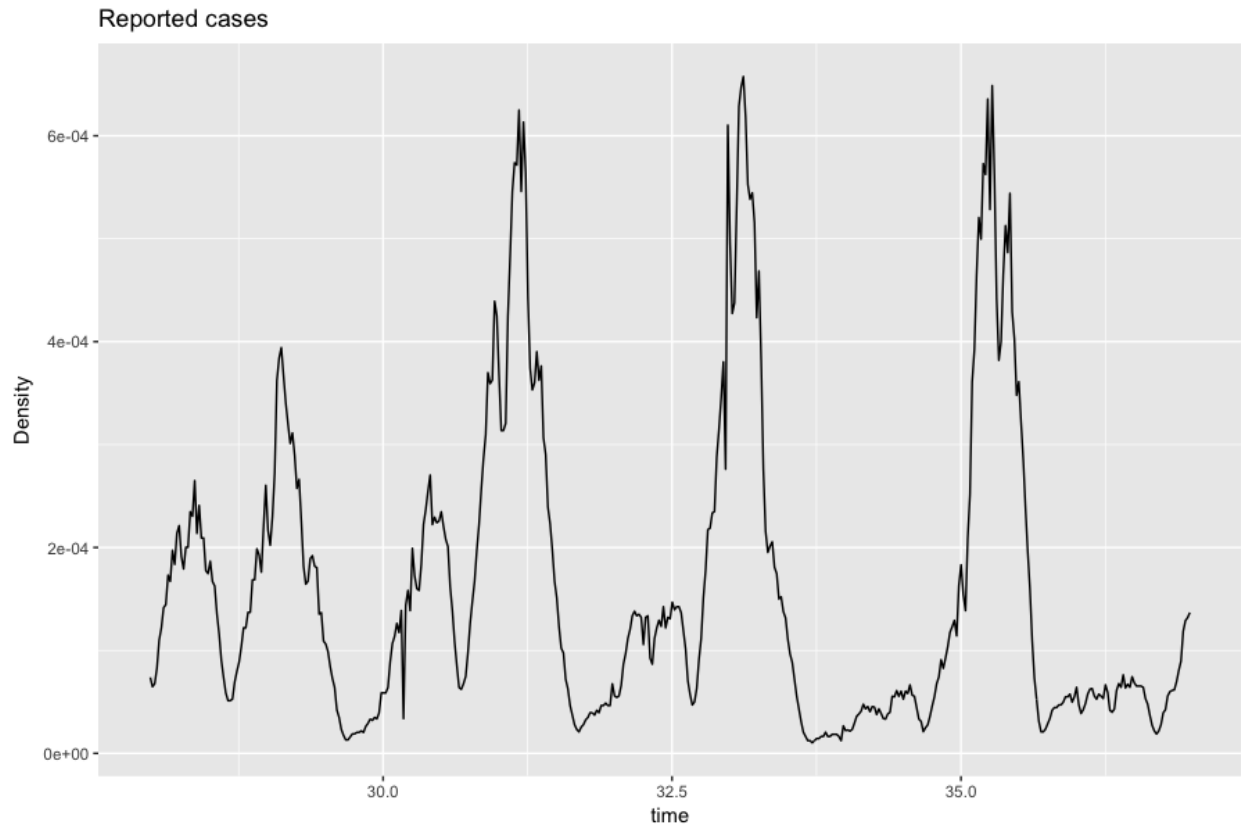
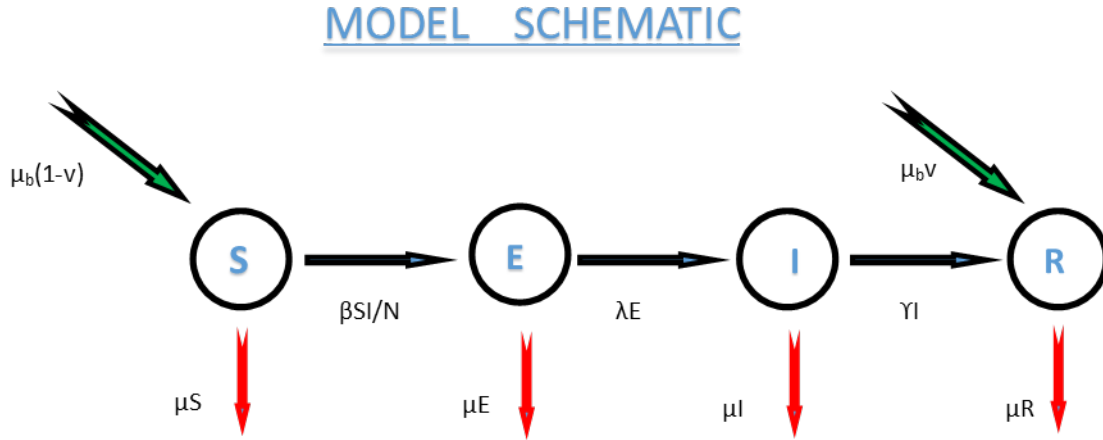


Figure 1:

## Mechanistic model

lets consider a simple mechanistic SEIR model with vaccination data . Let the number of births be  $\mu_b$ , the death rate be  $\mu$ , the transmission rate be  $\beta$  , the lenght of the latent period be  $\lambda$ , and the average lenght of infectiousness be  $\gamma$ . The expression on the right hand side for each state indicates the instantaneous rate of change of the states density. The positive terms represent the growth of that density, likewise, the negative terms represent the decline, thus for the first differential equation, we have  $\mu_b(1 - v)$  to be the inflows into the susceptible compartment i.e. the number of births of unvaccinated children,  $\beta IS/N$  the infected outflow and  $\mu S$  the number of susceptibles that die. We also assume standard incidence. Here we have a model where there is no vertical transimission, thus newborns are born into the susceptible class. Those in the suscepitble class can get infected and thus move to the exposed class, can be vaccinated and move to the recovered class (We refrain from creating a vaccinated class because we are not interested in its density), or can simply remain in the vaccinated class. The exposed eventually get infectious and move into the infected class which then eventually move into the recoverd class. We assume that there is permanent immunity both from the being vaccinated and going and getting infected. There is also a constant death rate for the compartments which implies that there is no disease induced death. Below is the model schematic.



We have the following as our derived differential equation from the model.

$$S' = \mu_b(1 - v) - \beta IS/N - \mu S$$

$$E' = \beta SI/N - (\lambda + \mu)E$$

$$I' = \lambda E - (\gamma + \mu)I$$

$$R' = \gamma I - \mu R + \mu_b v$$

$$S(0) \geq 0, E(0) \geq 0, I(0) \geq 0, R(0) \geq 0$$

All parameters above are non negative.

Before we analyse and evaluate how good the model is, we first have to ensure that the model is well posed and it atleast has a solution.

## INVARIANCE OF THE NON-NEGATIVE CONE

To ensure that the model is well posed we have to verify that these states representing population densities will never be negative given non negative initial condition. That shows that solution curves with non negative initial condition remain in the non-negative cone

$$\{T = (S, E, I, R) \mid S(t) \geq 0, E(t) \geq 0, I(t) \geq 0, R(t) \geq 0, S(t) + E(t) + I(t) + R(t) = N(t)\}$$

. To do so, we examine the behavior of states on the boundary of the positive cone.

1. When  $S(t) = 0$

$$S' = \mu_b(1 - v)$$

Since the right hand side is greater than zero, when the density of susceptibles gets to zero it immediately jumps up. And thus it will never be negative.

2. When  $E(t) = 0$

$$E' = \beta SI/N$$

Since the right hand side is greater than or equal to zero, when the density of exposed gets to zero it immediately jumps up or remains at 0 depending on the density of susceptibles and infected. This assures us of a non-negative density.

3. When  $I(t) = 0$

$$I' = \lambda E$$

Since the right hand side is greater than or equal to zero, when the density of infected decreases to zero it also rises or remains at 0 depending on the density of the exposed.

4. When  $R(t) = 0$

$$R' = \gamma I + \mu_b v$$

Since the right hand side is greater than zero, when the density of recovered gets to zero it immediately jumps up.

Thus, we've show the non-negative cone is invariant.

## Existence and Uniqueness

Let

$$f(S, E, I, R) = \begin{bmatrix} \mu_b(1 - v) - \beta IS/N - \mu S \\ \beta SI/N - (\lambda + \mu)E \\ \lambda E - (\gamma + \mu)I \\ \gamma I - \mu R + \mu_b v \end{bmatrix}$$

Since  $f$  is continuous and  $f'$  is continuous at all points in its domain,  $f$  is **Lipschitz** . By locally Lipschitz and continuity we have that the **solution exists and is unique**.

## Solutions are Bounded

To show that the solutions of this system exist for all time, we show that solutions are bounded. Adding the right hand sides of the above system of differential equations, we obtain

$$N' = \mu_b - \mu N \quad N(0) = N_0$$

. This differential equation has analytical solution as

$$N(t) = N_0 e^{-\mu t} + \frac{\mu_b}{\mu}$$

. Thus since  $N(t) = S(t) + I(t) + E(t) + R(t)$ , and the limiting value of  $N(t)$  is thus  $\frac{\mu_b}{\mu}$ , we have that the solutions are bounded and exist for all time.

A point  $x^*$  is an equilibrium point if  $f(x^*) = 0$ . As this implies that the rate of change in the state are all 0, once the system gets to an equilibrium point it stays there. Stability of a point is determined by how trajectories initialized close enough to the equilibrium point behave. Stability shows how small changes in initial condition will affect the system. The Disease-Free Equilibrium is the equilibrium associated with local disease extinction. That is when all infection related classes are 0 and the system is at equilibrium. Also a system can be at equilibrium with the disease still persistent in the population. This type of equilibrium is referred to as the endemic equilibrium.

The Disease Free equilibrium (D.F.E) is evaluated to be

$$\left(\frac{\mu_b(1-v)}{\mu}, 0, 0, \frac{\mu_b}{\mu}v\right)$$

## PARAMETER ESTIMATES OF MODEL AND REALISATION

This model is however not adequate to describe the dynamics of the disease using prevaccination data. This model exhibits damp oscillations (transient fluctuations) that eventually goes to an equilibrium point. Recalling that the trajectories are initiated much at 1920 to reduce the dependence on initial conditions, this transient “burn-out” period is what exhibits the fluctuation and thus there will almost be no information for minimizers for the parameter estimation. This can be seen in the figure below. Here the system goes to an endemic equilibrium with estimated parameters .

##	mub	beta	mud	lambda	gamma
##	0.02500000	600.00000000	0.01666667	45.62500000	26.07142857
##	va	s	e	i	r
##	0.00000000	0.04347826	0.00000000	0.00010000	0.95642174

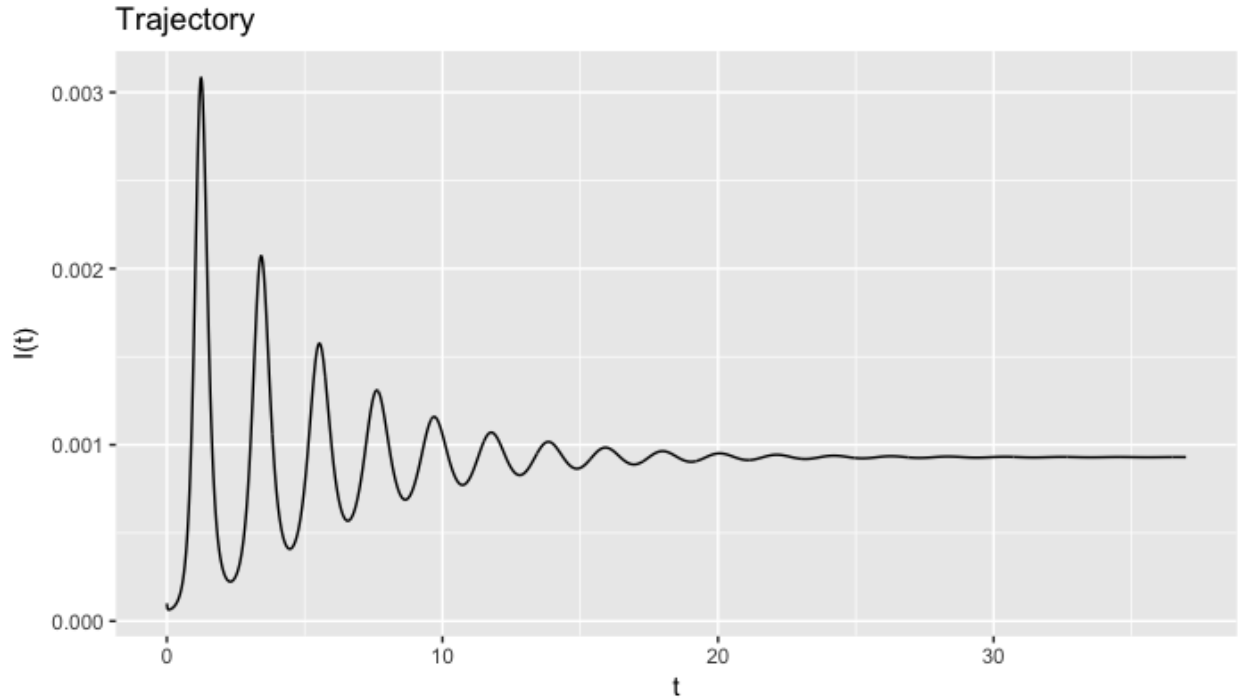


Figure 2:

## Sensitivity analysis

Sensitivity analysis determine how responsive the model states are to parameters. This helps us ascertain how uncertainties affect our model and its solutions. For models with the closed forms of the analytic solutions known, the sensitivity functions are the derivatives of the states in respect to time. However not all models have anaylic solutions and most often to get the sentivity function we differentiate our model in respect to the parameters, this gives new differential eqautions (of sensitivities) which we solve alongside out main differential equation. For the SEIR model without forcing, giving the parameter values above, below is a plot of its sensitivities.

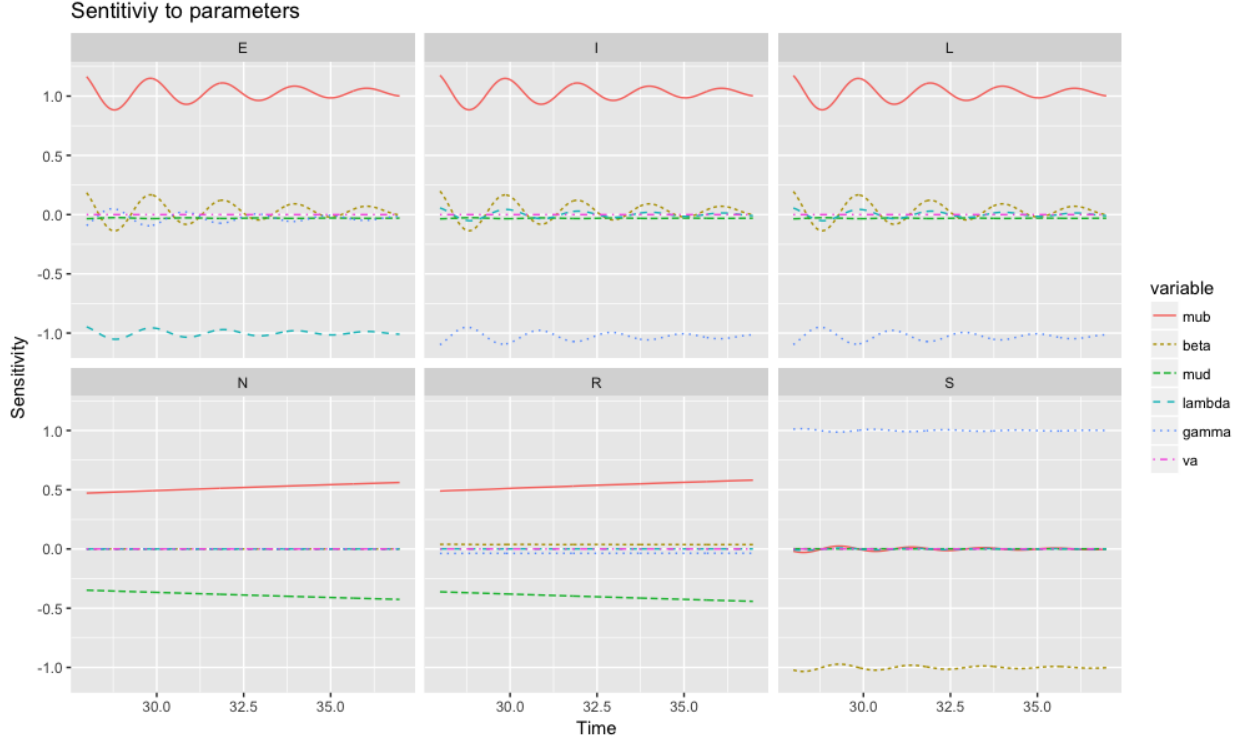


Figure 3:

We see that the models states are almost insensitive to parameters , except for  $\mu_b$  and  $\mu_d$ . However from the magnitude, shape and directions of these, we realise that the the combined effect also cancels out.

## The Basic Reproduction Number

With states at infection being  $\{E, I\}$  we use the next generation matrix to find the basic reproductive number. Consider

$$E' = \beta SI/N - (\lambda + \mu)E$$

$$I' = \lambda E - (\gamma + \mu)I$$

if we decompose the above as  $X' = [T + \Sigma]X$ , where  $T$  is for transmission i.e. the birth of newly infected, and  $\Sigma$  is for transition from one compartment to the other, the  $R_0$  is the dominant eigenvalue of  $K = -T\Sigma^{-1}$  evaluated at the D.F.E. Hence, we have our  $R_0$  as

$$\frac{\lambda\beta(1-v)}{(\lambda+\mu)(\gamma+\mu)}$$

## Model with Seasonal Forcing

As can be seen from the above the dynamics of prevalence from the model doesn't match that observed, the seasonally (i.e. periodic nature) of prevalence isn't captured. We consider a modification of the model with seasonal forcing. Seasonal forcings illustrates the dynamic nature of certain parameters (in this case incidence). In the above model we assumed standard incidence with average rate of contact constant. We now use a sinusoidal function for the average rate of contact. Adding seasonal forcing to this model adds its ability to capture the periodicity of the observed data. There are several types of forcing that can be used. For measles a realistic forcing will be one that corresponds to the school year, this is because measles is a childhood disease and is often transmitted at school. Examples of these are school-term forcing, where we mostly have a piece-wise function with high magnitude when the school year begins in August, and 0 or lesser values during the breaks. We however use a sinusoidal function to capture this pattern.  $\beta_1$  is the average rate of contact,  $\beta_1 \in (0, 1)$  is an amplitude parameter, and  $\phi \in (0, 1)$  is a parameter for adjusting periodicity. We have the new model to be:

$$\begin{aligned} S' &= \mu_b(1 - v) - \beta_1(1 + \beta_2 \cos(2\pi(t - \phi))) \frac{SI}{N} - \mu S \\ E' &= \beta_1(1 + \beta_2 \cos(2\pi(t - \phi))) \frac{SI}{N} - (\lambda + \mu)E \\ I' &= \lambda E - (\gamma + \mu)I \\ R' &= \gamma I - \mu R + \mu_b v \end{aligned}$$

$$S(0) \geq 0, E(0) \geq 0, I(0) \geq 0, R(0) \geq 0$$

All parameters above are non negative.

The analysis of the system with seasonal forcing is almost identical to the one without seasonal forcing except for the stability analysis which requires a more complicated method (Floquet Multipliers). However in this case the model is (very) sensitive to parameters and with that parameter estimation is plausible.

With estimated parameters as

```
gamma = 365/14

(pars <- c(mub=1/60, beta= 600, mud = 1/60, lambda = 365/8, gamma = 365/14,
          va = 0, s = 1/23, e = 0, i = 1e-4, r = 1-1/23-1e-4, beta2 = .2, phi = .5))

##          mub          beta          mud          lambda          gamma
## 0.01666667 600.00000000 0.01666667 45.62500000 26.07142857
##          va          s          e          i          r
## 0.00000000 0.04347826 0.00000000 0.00010000 0.95642174
##          beta2          phi
## 0.20000000 0.50000000
```

we obtain a least-square estimate of

```
## [1] 2.246786e-05
```

We clearly see the periodic nature of the trajectory in the figure below. The dynamics of the disease here mostly exhibits bi-annual cycles.

We carry out local sensitivity analysis for our parameter set. The model will be inadequate if the parameters are insensitive. However, and contrary to that observed when using standard forcing we have that the states are very sensitive to parameters. The figure below shows the sensitivity plot of the states in respect to the parameters across time. A sensitivity value approximately 1(-1) means the parameter has a strong positive(negative) relation with the state at that time. An almost 0 signifies a no relation. Because the birth and death rate was set to make the initial condition constant, we have as expected that the total population



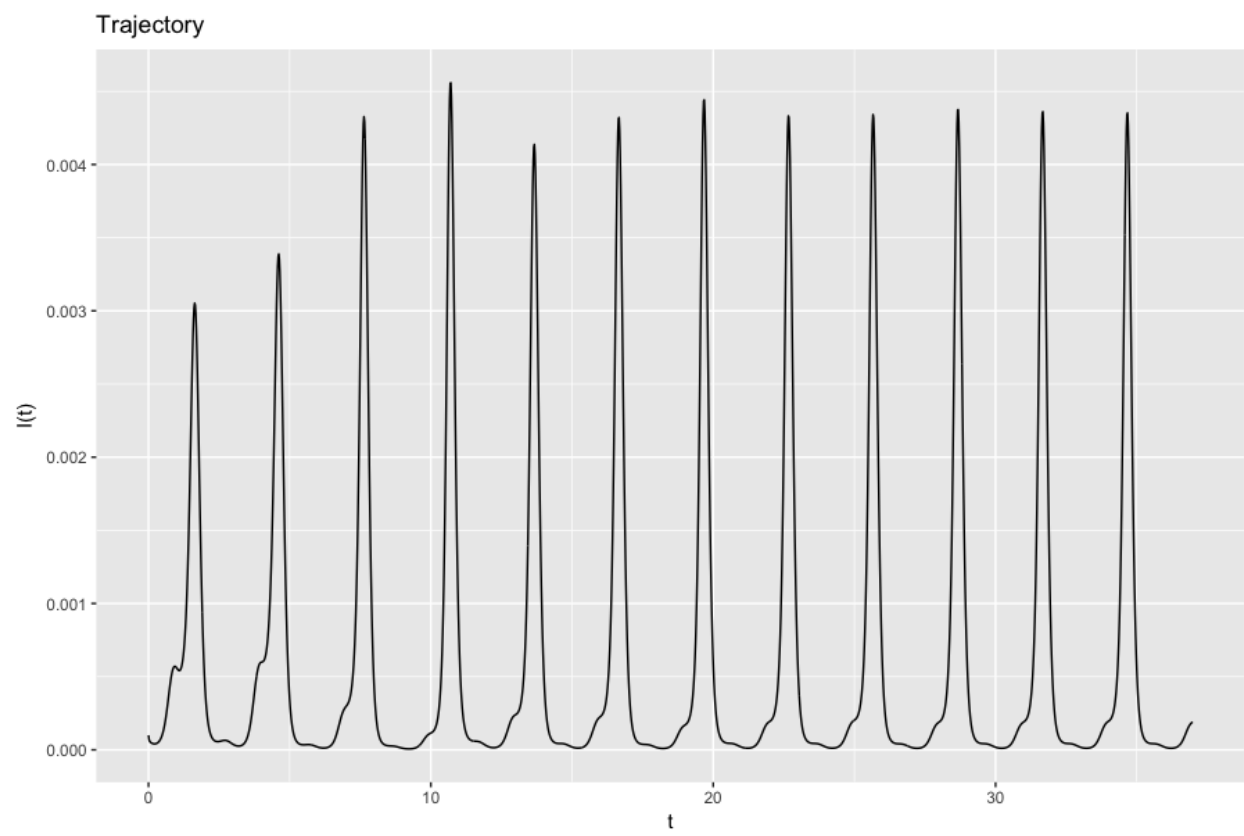


Figure 4:

size  $N(t)$  is insensitive to any of the parameters, the next insensitive state is the recovered state  $R(t)$ . The other states however show periodic sensitivities in respect to parameters.

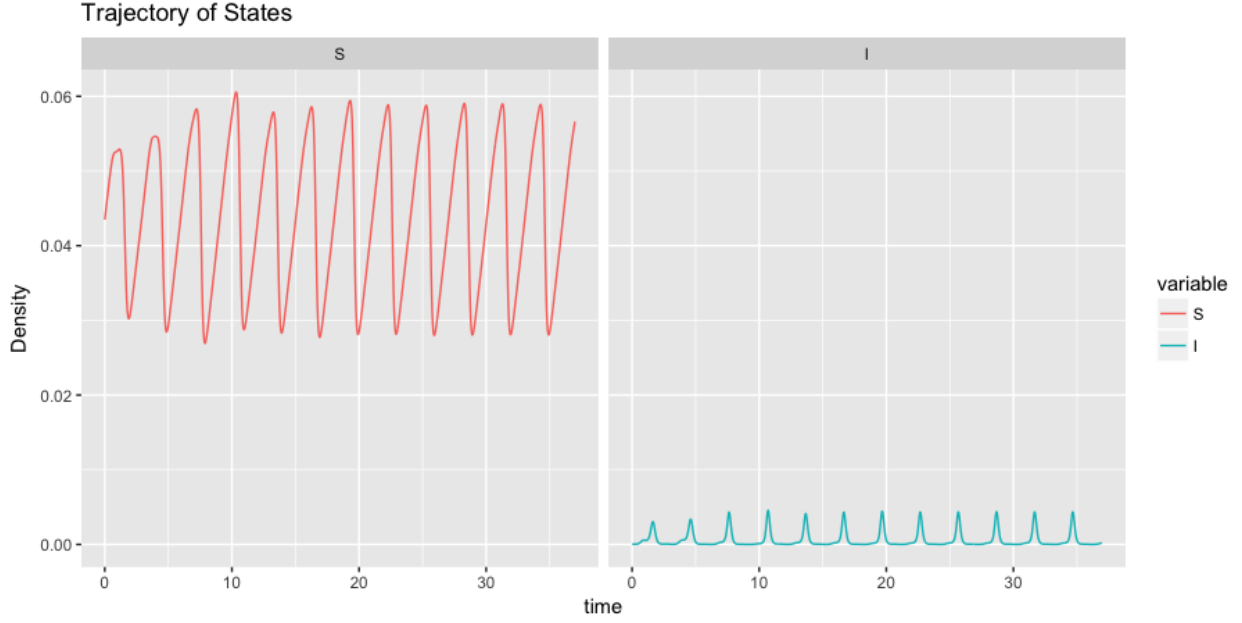


Figure 5:

### Modeling Measles Outbreaks (Jansen and Stollenwerk 2005)

As mention above the dynamics of an infectious disease is largely dependent on the Basic reproductive number  $R_0$ . When the  $R_0 > 1$  the disease establishes itself in the host population and remains endemic, when  $R_0 < 1$  the disease eventually dies out of the population. Considering the  $R_0$  of the model below (i.e.  $R_0 = \beta/\lambda$ ),

$$\begin{aligned} S' &= -\beta IS \\ I' &= \beta SI - \gamma I \\ R' &= \gamma I \end{aligned}$$

with  $S + I + R = 1$  (A scaled model).

After introducing vaccination into the model the Basic reproductive number reduces to  $R_c = \beta(1 - c)/\lambda = R_0(1 - c)$ , where  $c$  is the number of vaccinated people in the population. If we extend this model to include demography and vaccination rate ( $v$ ), we get

$$S' = \mu N - \beta IS/N - (\mu + v)S$$

$$I' = \beta SI/N - (\gamma + \mu)I$$

$$R' = \gamma I + vS - \mu R$$

with  $S + I + R = N$  (Unscaled). This system however displays stochastic effects when  $R_0$  approaches 1. For a supercritical system with  $R_0$  approaching 1 we have an initially stochastic phase with phases out to a deterministic one. On the other hand, for the subcritical system we have the system demonstrating strong stochastic dynamics. To study the stochastic model for measles outbreaks, a stochastic model was formed with the following assumptions on the latter model:

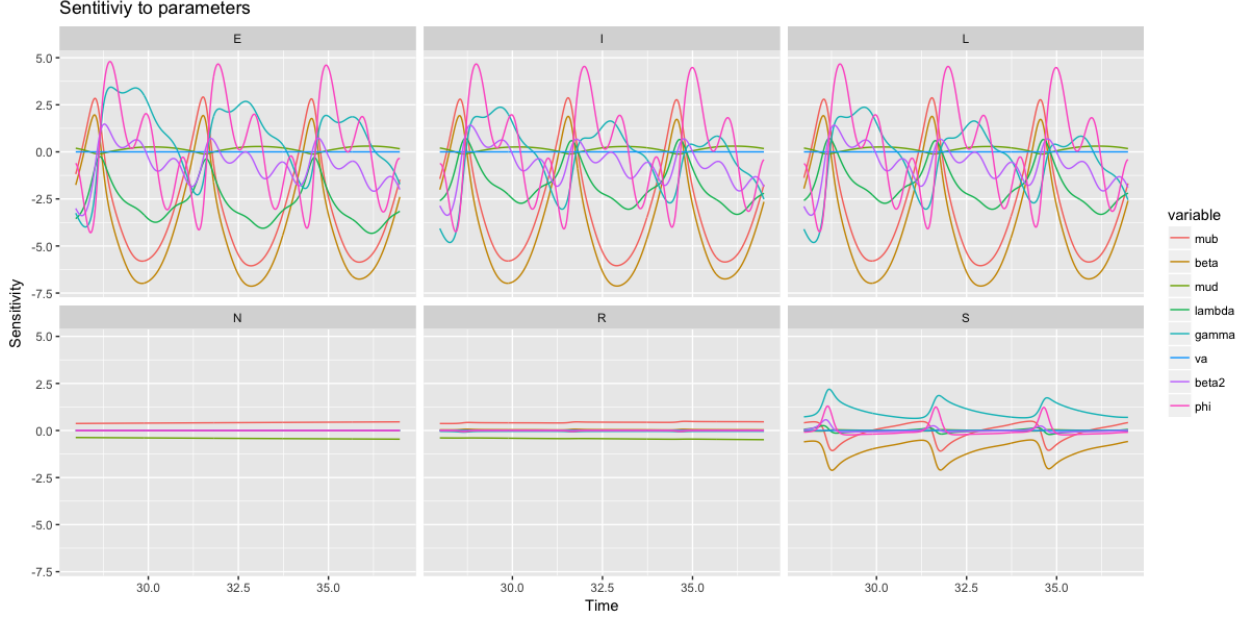


Figure 6:

1. There is a large host population and its size will not be affected by the disease dynamics
2. There are no vital dynamics and vaccination rate is set to 0 during the epidemic. ie the proportion of the vaccinated remains ( $c$ ).

Assumption 2 is because of the relatively short span of epidemics as compared to demographic processes. This assumption reduces the model to a birth and death process with birth as  $\lambda R_c$  and death as  $\lambda$ . And it is easy to derive an expression for the expectation and variance of the number of infected hosts,  $I(t)$ .

Though the time between jumps is exponentially distributed, to derive the size distribution of outbreaks, this was discarded and the time between jumps was fixed at 1, to simplify the model. The Gillespie algorithm was then used to obtain the probabilities of entering and leaving the initial state, which was  $\frac{R_c}{R_c+1}$  and  $\frac{1}{R_c+1}$  resp. The probability of an outbreak a certain size  $x$  or larger is given by

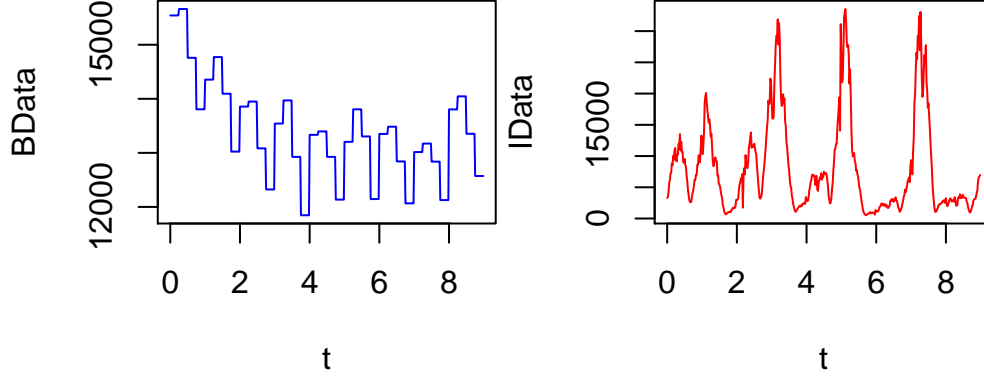
$$\sum_{x=n}^{\infty} q(x) = \max\left(0, 1 - \frac{1}{R_c}\right) + \frac{(4R_c)^{n-1}(\Gamma(n-.5))}{(1+R_c)^{2n-1} * \sqrt{\pi}n!} {}_2F_1\left(1, n-.5, n+1, \frac{4R_c}{(1+R_c)^2}\right)$$

Where  ${}_2F_1$  is the hypergeometric function, and  $q(x)$  is the probability of an outbreak a certain size  $x$ . A recent number of measles outbreak in UK as a result of the MMR scare can well be explained by the above dynamics. As the estimated  $R_c$  jumps from 0.35 to 0.7, before and during the scare.

## Time series modelling of childhood diseases: a dynamical systems approach (Finkenstadt and Grenfell 2000)

### Susceptible Reconstruction

We will be using measles incidence data of the UK and Wales. The data used in the Grenfell Discrete time Model from 1944-1966. However the data available to us is from 1948-1966. Thus the analysis will be done on the later years. However because of a setback with initial conditions of the susceptible reconstruction, we use the data of 1948 as the “burn-out” period. Hence data from 1949 onward will be considered in our estimation. Below is a graph of Incidence and Births for the given periods.



Data observed is only of reported cases, thus, there is no data available for the number of susceptibles at a particular point in time, and even the incidence data collected is only partially observed, ie there is underreporting, and not everyone infected by the disease reports it. To address this, Susceptible Reconstruction Algorithm is used to reconstruct the dynamics of the unobserved susceptibles from birth and partially observed incidence data. This is done by regressing cumulative births and cumulative case notification. An assumption is first made of a constant reporting rate through time and this is estimated about 56.8%.

The dynamics of the susceptible class can be described by

$$S_t = B_{t-d} + S_{t-1} - I_t + U_t$$

$B_t$  being the number of Births at time  $t$ , with  $d$  corresponding to a small delay that allows maternal immunity to dwindle. For measles its approximately 8 weeks.  $S_t$  and  $I_t$  are the number of susceptibles and infected at time  $t$  respectively. We also have additive error factor  $U_t$ , with  $E(U_t) = 0$  and  $Var(U_t) = \sigma_U^2$ . For the transmission equation we have

$$I_t = r_t I_{t-1}^{\alpha_1} S_{t-1}^{\alpha_2} \epsilon_t$$

. Here we assume multiplicative error for the total error is most likely to depend on the magnitude of the model values. We also have  $E(\ln(\epsilon_t)) = 0$  and  $Var(\ln(\epsilon_t)) = \sigma_\epsilon^2$ .  $r_t$  is a factor of proportionality that varies with time with a period of a year.  $\alpha_1$  and  $\alpha_2$  are mixing parameters, when  $\alpha_1 = \alpha_2 = 1$  we have homogeneous mixing.

### Reconstructing the Susceptible dynamics.

The Exposed and Infected period for measles last for 2 weeks then its followed by a life long immunity. Because the disease time scale is 2 weeks we aggregate the data bi-weekly. Thus the infected persons at time  $t$  are related to contacts from time  $t - 1$ , and also effect the number of infected at  $t + 1$ .

The number of true cases is proportional to the number of reported cases, i.e.  $I_t = \rho_t C_t$ , with  $C_t$  the number of reported cases and  $\rho_t \geq 1$  a random variable with  $E(\rho_t) = \rho$ . Thus we have our susceptible equation as

$$S_t = B_{t-d} + S_{t-1} - \rho_t C_t + U_t$$

Letting  $Z_t = S_t - \bar{S}$ , where  $E(S_t) = \bar{S}$  we get  $E(Z_t) = 0$ , subtracting  $\bar{S}$  from both hannd side of the equation (The deviations of  $S_t$  are recursively the same for  $Z_t$ ) gives

$$Z_t = B_{t-d} + Z_{t-1} - \rho_t C_t + U_t$$

Successively iterating the above equation gives us,

$$Z_t = Z_0 + \sum_{i=1}^t B_{i-d} - \sum_{i=1}^t \rho_i C_i + \sum_{i=1}^t U_i$$

This is thus free from the autoregressive nature of our initial subsebtible reequation. We reparametrize our model with

$$X_t = \sum_{i=1}^t C_i, Y_t = \sum_{i=1}^t B_{i-d}, U_t = \sum_{i=1}^t u_i, R_t = \sum_{i=1}^t (\rho_i - \rho) C_i.$$

for  $t = 1, \dots, n$  then the following are random processes.

$$R_t = R_{t-1} + (\rho_t - \rho) C_t U_t = U_{t-1} + u_t$$

With the new notation we have, our regression equation as

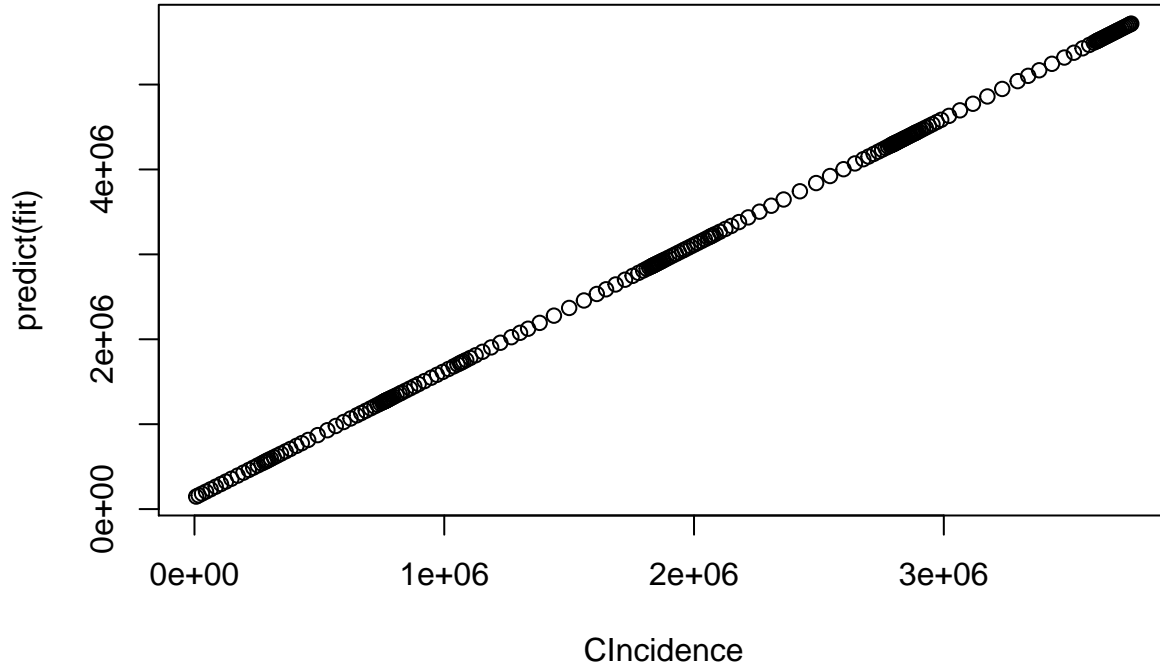
$$Y_t = -Z_t + \rho X_t + R_t + Z_t - U_t$$

A subset of the Dataset can be seen from the Data below, with time, Culmulative Incidence, Culmulative Birth and Culmulative Error.

##	Time	CIncidence	CBirths	U
## 1	0.00000000	6755	31322.0	1.5824386
## 2	0.03982301	16581	62644.0	2.2289468
## 3	0.07964602	30015	93966.0	0.7977002
## 4	0.11946903	46213	125288.0	1.0609918
## 5	0.15929204	64790	156610.0	1.2934610
## 6	0.19911504	85056	186126.6	1.1219288

Assuming a constant reporting rate and noise free data i.e  $U_t \approx 0$  we have a simple linear relation of culmulative incidence on culmulative birth. A point estimate of the average reporting rate is the reciprocal of the estimate of the slope. A summary of the simple linear regression and Analysis of Variance can be seen below.

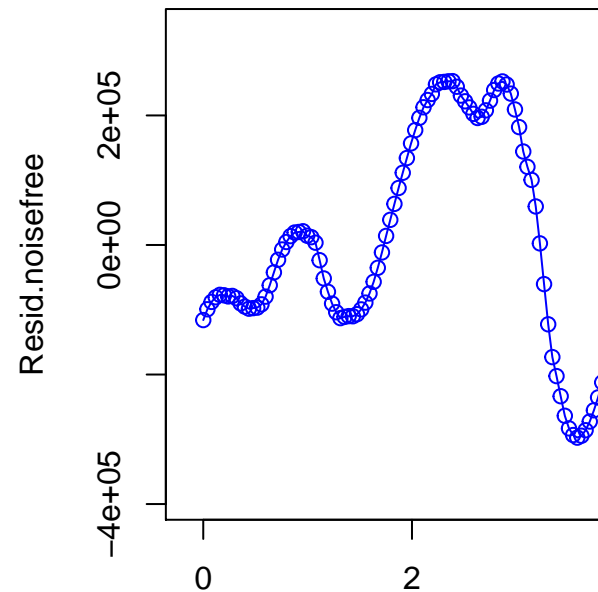
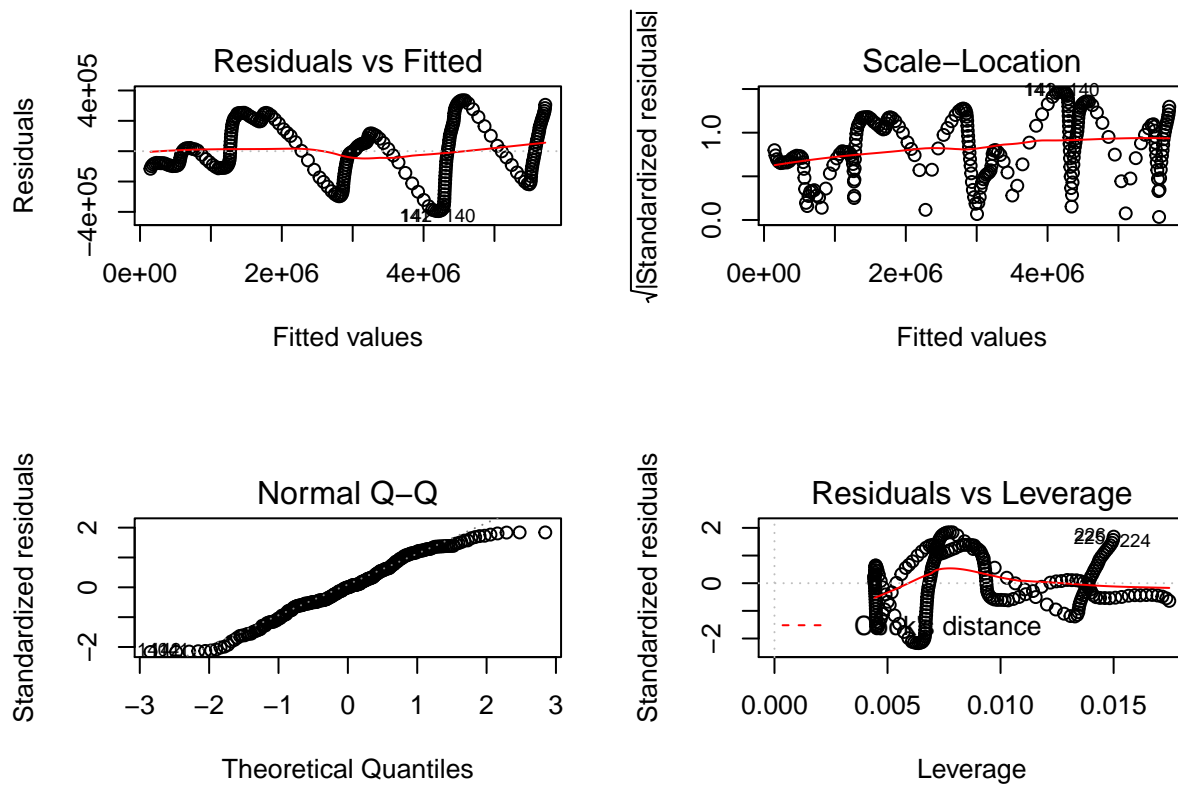
```
##
## Call:
## lm(formula = CBirths ~ CIncidence, data = NewData2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -395028 -109442    -528   144142   335083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.372e+05  2.423e+04   5.663 4.54e-08 ***
## CIncidence  1.488e+00  1.060e-02 140.293 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 182900 on 224 degrees of freedom
## Multiple R-squared:  0.9887, Adjusted R-squared:  0.9887
## F-statistic: 1.968e+04 on 1 and 224 DF, p-value: < 2.2e-16
```



```
##               2.5 %       97.5 %
## (Intercept) 89474.085945 1.849815e+05
## CIncidence   1.466752 1.508544e+00

## Analysis of Variance Table
##
## Response: CBirths
##           Df      Sum Sq   Mean Sq F value    Pr(>F)
## CIncidence  1 6.5853e+14 6.5853e+14   19682 < 2.2e-16 ***
## Residuals 224 7.4946e+12 3.3458e+10
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We have our estimate of the slope to be 1.488, corresponding to a 67.2% average reporting rate. A 95% confidence interval for the average reporting rate is given as (66.29%, 66.28%). All regression related tests were significant, and R-squared: 0.9887 signifies a strong positive linear relation, with the regression model explaining 98.87 percent of the error. The purpose of the regression equation is to estimate the Residuals, which about 0 has the same dynamics as  $S_t$  about  $\bar{S}$ . And also the reporting rates  $\rho_t$ . The estimates by (Finkenstadt and Grenfell 2000) is a slope of 1.76 corresponding to a reporting rate of 56.8%.

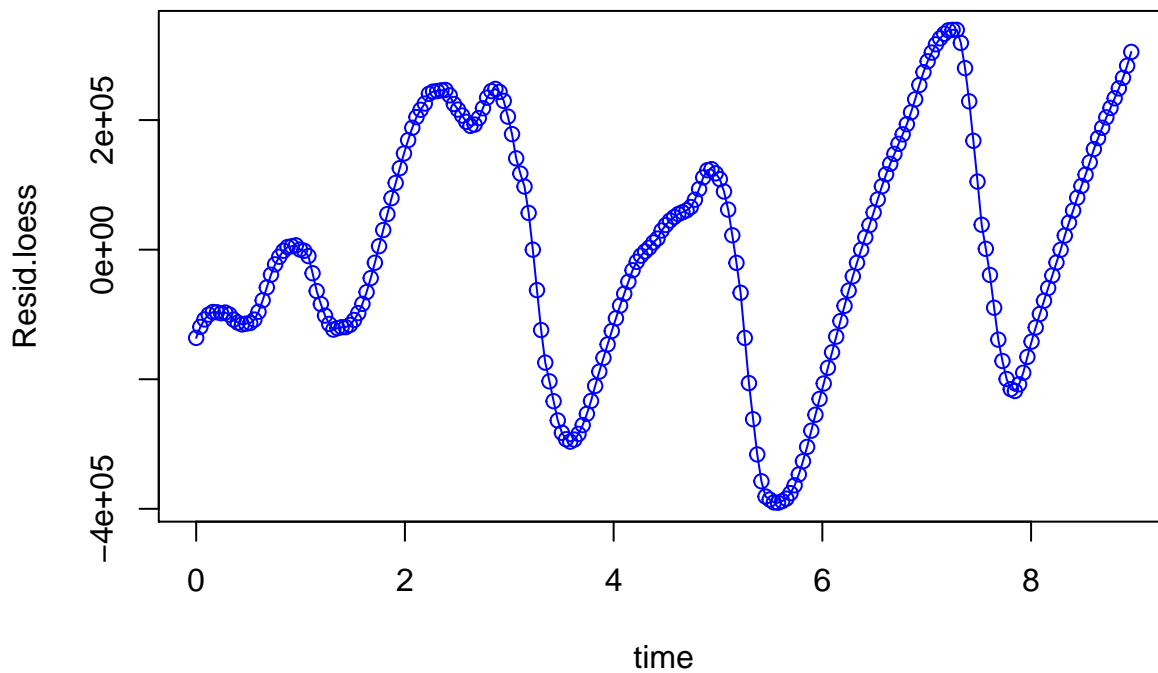


Below are plots of the estimated residuals both with noise and without noise.

## Local Regression analysis

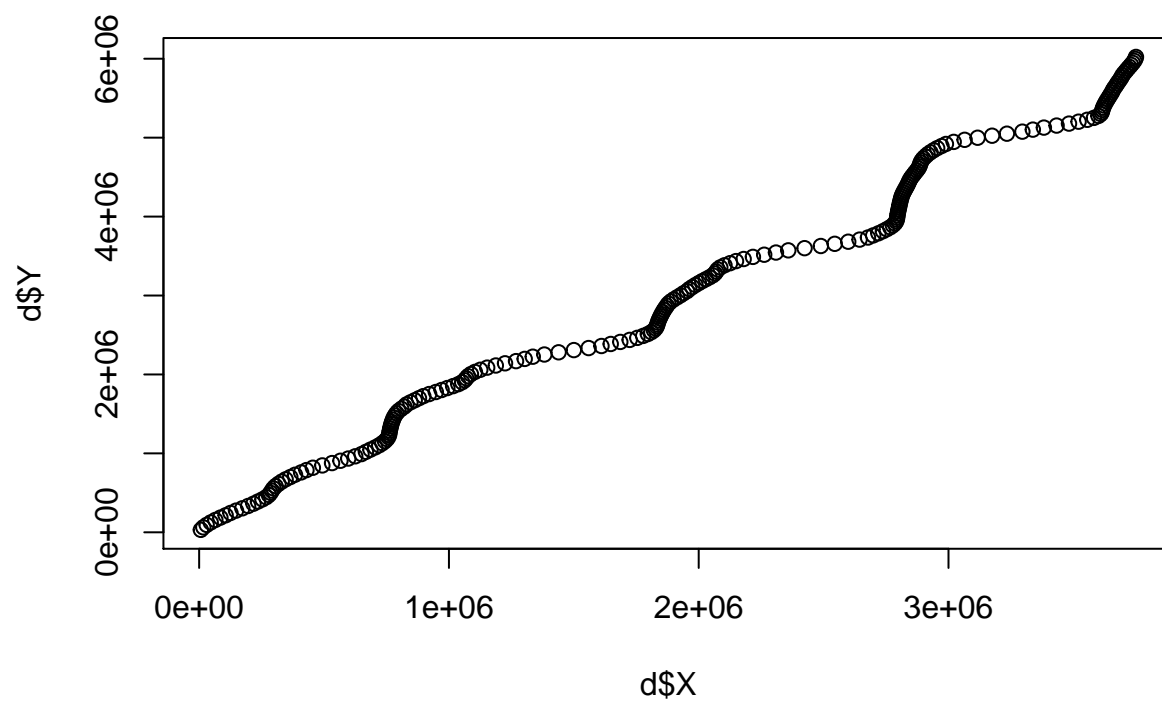
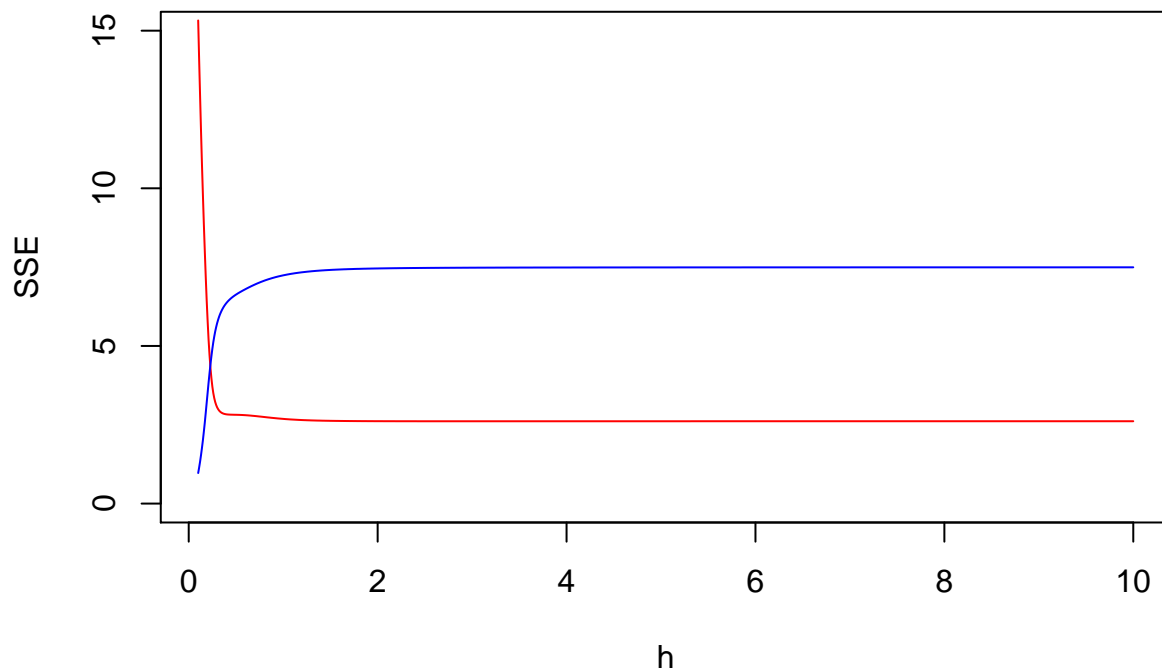
With the simple linear regression, we assumed a constant reporting rate. This however isn't realistic, to get a better address this we rather restrict the constant reporting rates to within each of the bi-weeks. We therefore perform local regression analysis and this also gives us a better fit. " references " and LRA.

```
## Call:
## loess(formula = CBirths ~ CIncidence, data = NewData2, span = 1.49992,
##       degree = 1)
##
## Number of Observations: 226
## Equivalent Number of Parameters: 2.14
## Residual Standard Error: 183000
## Trace of smoother matrix: 2.29 (exact)
##
## Control settings:
##   span      : 1.49992
##   degree     : 1
##   family     : gaussian
##   surface    : interpolate      cell = 0.2
##   normalize  : TRUE
##   parametric : FALSE
##   drop.square: FALSE
```



```
## Loading required package: locpol
```





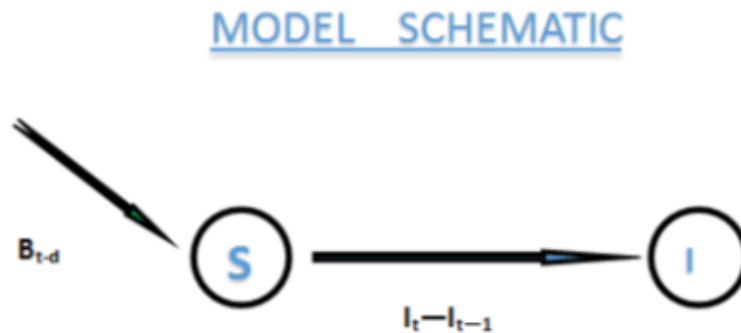
```
#Data needed is in Fdat.d and mydat

Data = data.frame( Fdat.d[,-3], mydat )
Data[,9] = Data[,1] * Data[,6] #Creating I(t)
Data[,10] = c( NA, Data[,8])[1:226] #Creating Z(t-1)
Data[,11] = c( NA, Data[,9])[1:226] #Creating I(t-1)
colnames(Data)[c(1:2,8,9,10,11)] = c("C","B","Z","I","Zt-1","It-1")

source('~/.GitHub/Measles/TSIR parameter estimation.R', echo=F)
```

#Fitit(10)

“Work in progress!!!!!!”



### Dynamics of Measles Epidemics (Estimation and scaling of transmission Rates using a TSIR Model) (Bjornstad, Finkenstadt, and Grenfell 2002)

A study of the pre-vaccination era in UK shows two main dynamics; the endemic dynamics with periodicity which is exhibited in Large populations and the episodic outbreaks exhibited in smaller populations. The threshold differentiating this large and small population of hosts is estimated to be  $\sim 300000-500000$ . It is realized that endemism (The persistence of the disease in the host population) is dependent on the  $R_0$  and also the susceptible replenishment rate (Grenfell and Harwood 1997). However, for non-persistent diseases, new outbreaks are dependent on reintroduction of the disease into the host population (Cliff et al 1993). A TSIR (Time Series SIR) Model was use to explore the balance between noise and episodic dynamics. The advantages of this model include 1. It can be scaled by population size to produce endemic and episodic dynamics 2. It bridges the gap between theoretical model and empirical data.

This model includes **seasonal forcing corresponding to the start of the school year** (*Seasonal forcing explains the periodicity of endemic dynamics*) . Also non-linearity in contact rates are allowed. The most relevant state variables consider are the number of infected and susceptible at any time  $t$ . The approach used was to generate the number of susceptible using the Susceptible Reconstruction Algorithm proposed by Finkenstadt and Grenfell (2000), and the record of birth and deaths during the fitting period.

There are some obstacles with this model, one of this is the Susceptible Reconstruction Algorithm is less reliable in the early part of the time series, because the “initial conditions” are difficult to estimate without some allowances. The model was fitted to weekly data from 60 cities in the UK, collected weekly from 1944-1966. The data was aggregated on a 2 wk. interval to match the time scale of the disease. The rates are

assumed fixed within this interval, thus the number of secondary cases from a single infection is geometrically distributed (Kendall 1949). Together with the assumption of independence the birth-death process will follow a negative binomial distribution (Kendall 1949).

$$I_{t-1} \sim NB(\lambda_{t+1}, I_t)$$

with expectation  $a$  and clumping parameter  $b$ . and

$$S_{t+1} = S_t + B_t - I_{t+1}$$

with this the model is fitted to data.

## Likelihood-based estimation of continuous-time epidemic models from time-series data (Cauchemez and Ferguson 2007)

Likelihood based estimation of parameters for our mechanistic models will be straight forward if infection and removal times are observed for each incident, however this isn't possible because the available data is partially observed and aggregated often weekly. Though data augmentation methods have been used to tackle the missing data problems, this increases the complexity of likelihood based estimation when the number of missing data is huge.

TSIR model though fits well to data in some cases, makes strong assumptions which are not always applicable. One of such assumptions is that the epidemic and data collection processes should have the same time scales. This results from the discretization of a continuous time system because the case counts are usually weekly, and this is quite good if each time period is one generation period. Since TSIR models are unable to estimate the generation times (delay between infection of a case and their typical secondary case) of the disease they assume that the generation time is equal to the observation interval (or a constant multiple of it), which for the measles model above is taken to be two weeks. However, this is often not the case and continuous-time models are free of such constraint.

This model is applicable when we either have temporally aggregated datasets, and the number of size of underreporting is too large to use data augmentation techniques or the epidemic and data collection have different time scales.

The approach involves tackling missing data by augmenting the data with latent state at the beginning of each observational period. The relation between the latent states are derived from a Cox-Ingersoll-Ross process which copies the usual SIR process. The Metropolis-Hastings MCMC sampling is used to explore the joint posterior distribution of augmented data and parameters.

This continuous-time series model is unable to estimate generation times of measles accurately from observation data, though it does well for simulated data if the observation interval is less than 2.5 times the generation time.

## Plug-and-play inference for disease dynamics (He, Ionides, and King 2010)

Continuous-time SIR models adopts diffusion processes to implement MCMC, this breaks down in small populations. Several models make assumptions, and this makes them only locally valid. On the other, plug-and-play techniques makes little or no assumptions that interfere with the hypothesis of interest. This is because it operates basely on simulation from the dynamic model, and that brings its only disadvantage, it is computationally expensive. Two distinctive features about Plug-and-play methods have been considered in the optimization and complex system theory; gradient-free and equation-free.

Given a stochastic dynamic model (partially observed Markov processes), sequential Montecarlo is used to derive the log likelihood function. This means only path simulations are needed. This is done through iterated filtering contained in the R Software Package *POMP*. If identifiability issues are encountered further

assumptions might resolve that issue. Otherwise, profile likelihood will aid in parameter estimation and identifiability. The model has force of infection as

$$\mu_{SE}(t) = \frac{\beta(t)(I + i)^\alpha}{N(t)}$$

where  $\beta(t)$  is the transmission rate which is seasonally forced to match the school year.

With the refinement in modeling tools, we are better equipped to describe the dynamics of measles, mostly separately since the transitions between the episodic and endemics dynamics, is not well understood. Before mass vaccination measles displayed regular endemic cycles which later changed to irregular ones. Most models using the normal SEIR setup and seasonal variation of transmission rate account for this transitions by endogenous dynamical reasons, such as chaos and stochastic shifts in existing stable cycles. This makes predicting transitions almost impossible. But the realization that susceptible replenishment rate affects the disease dynamics gives us a credible and exogenous factor which aids in predicting transitions. This is taken for granted when a constant birth rate is used in models for it ignores slow changes in this parameter. With the appropriate seasonal forcing of transmission rates (i.e. term time forcing), a simple SEIR framework well explains the disease dynamics without a complex age-structured host model [2].

## References

- Bjornstad, Ottar N, Barbel F Finkenstadt, and Bryan T Grenfell. 2002. “Dynamics of Measles Epidemics: Estimation Scaling Of Transmission Rates Using A Time Series SIR Model.” *The Ecological Monographs*.
- Cauchemez, Simon, and Neil M Ferguson. 2007. “Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London.” *The Royal Society*.
- Finkenstadt, Barbel F, and Bryan T Grenfell. 2000. “Time series modelling of childhood diseases: a dynamical systems approach.” *Applied Statistics*.
- He, Daihai, Edward L Ionides, and Aaron A King. 2010. “Plug-and-play inference for disease dynamics: measles in large and small populations as a case study.” *The Royal Society*.
- Hethcote, Herbert W. 2000. “The Mathematics of Infectious Diseases.” *Society for Industrial and Applied Mathematics*.
- Jacquez, J A, and C P Simon. n.d. “Qualitative theory of compartmental systems.” doi:10.1137/1035003.
- Jansen, V A A, and N Stollenwerk. 2005. “Modeling Measles Outbreaks.” *Branching Processes: Variation, Growth, and Extinction of Populations*.
- Mishra, Sharmistha, David N Fisman, and Marie-Claude Boily. 2010. “The ABC of terms used in mathematical models of infectious diseases.” *J Epidemiol Community Health*.
- Quadros, Ciro A de. 2004. “Can measles be eradicated globally?” *Bulletin of World Health Organization*.