

Limit Cycles and Dynamics of Rumor Models

---

A thesis

presented to

the faculty of the Department of Mathematics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

---

by

Geophrey Odero

December 2013

---

Ariel Cintron-Arias, Ph.D., Chair

Jeff Knisley, Ph.D.

Michele Joyner, Ph.D.

Keywords: basic reproduction number, halting rate, sensitivity analysis

## ABSTRACT

### Limit Cycles and Dynamics of Rumor Models

by

Geophrey Odero

This thesis discusses limit cycles and behavior of rumor models. The first part presents the deterministic Daley-Kendall model (DK) with arrivals and departures and comparison of the Susceptibles, Infectives and Removed (SIR) model and the DK model. The second result is a part of the qualitative analysis and the general behavior of extension of the Daley-Kendall model. Here we discuss how the halting rate of spreaders causes the model to change from a stable equilibrium or a stable limit cycle. In the third part we carry out model validation and use both synthetic data and real data sets and fit them to the numerical solutions of the extended Daley-Kendall model. Finally, we find the parameter estimates and standard errors. In this way we shall be able to decide whether the numerical solutions quantifying the relationships between the variables obtained from the qualitative analysis can be accepted as the best description of the data. We discuss sensitivity analysis results and traditional sensitivity functions.

Copyright by Geophrey Odero 2013

All Rights Reserved

## ACKNOWLEDGMENTS

This thesis would have not been successful without the support of many people. First of all, I would like to thank Dr. Ariel Cintron-Arias, my thesis chair, for the suggestion of this topic and continued advice and unrelenting support. His help and guidance made this work successful. I would also like to thank my committee members, Dr. Jeff Knisely and Dr. Michele Joyner for their encouragement, guidance and support. I would also like to thank East Tennessee State University for providing me with conducive environment to complete this thesis. And finally, thanks to my mother, and numerous friends who always offered me encouragement, support and love throughout this process.

## TABLE OF CONTENTS

ABSTRACT . . . . .	2
ACKNOWLEDGMENTS . . . . .	4
LIST OF FIGURES . . . . .	7
1 INTRODUCTION . . . . .	8
1.1 Basic Reproductive Number for the SIR Model versus the DK Model . . . . .	10
2 EXTENSIONS TO THE DALEY-KENDALL MODEL . . . . .	12
2.1 Halting Rate of Spreaders Leading to Stable Equilibrium . . .	12
2.2 Halting Rate of Spreaders Leading to Stable Limit Cycles . . .	16
3 MODEL VALIDATION . . . . .	26
3.1 Forward Solution and Synthetic Data . . . . .	27
3.2 Ordinary Least Squares Estimator . . . . .	29
3.3 Uncertainty Quantification . . . . .	30
3.4 Real Data Sets . . . . .	31
3.5 Parameter Estimates and Standard Errors . . . . .	36
4 SENSITIVITY ANALYSIS . . . . .	53
5 CONCLUDING REMARKS . . . . .	57
BIBLIOGRAPHY . . . . .	60
VITA . . . . .	61

## LIST OF FIGURES

1	Compartment Model for the Naives, Spreaders and Stiflers . . . . .	9
2	The SIR Model versus the DK Model plots . . . . .	11
3	Time Plots and Phase Portrait for Stable Equilibrium . . . . .	15
4	Trapping Region . . . . .	20
5	Trace Curve in the $(a, b)$ Parameter Space . . . . .	24
6	Time Plots and Phase Portrait for a Stable Equilibrium in Panels $(a)$ – $(b)$ , and Stable Limit Cycle in Panels $(c)$ – $(d)$ . . . . .	25
7	True Solution of Equations (9)–(10) and Synthetic Data versus Time	28
8	Plot of the Summer Olympics Data Sets . . . . .	32
9	Plot of the Winter Olympics Data Sets . . . . .	33
10	Plot of the U.S.A. Presidential Elections Data Sets . . . . .	34
11	Plot of the U.S.A. Elections Data Sets . . . . .	35
12	The best Fit Solution for the Summer Olympics Data Sets using the Nelder-Mead optimization algorithm (fminsearch) . . . . .	37
13	The best Fit Solution for the Summer Olympics Data Sets using a pattern search optimization algorithm . . . . .	38
14	The best Fit Solution for the Summer Olympics Data Sets using a global search optimization algorithm . . . . .	39
15	The best Fit Solution for the winter Olympics Data Sets using the Nelder-Mead optimization algorithm (fminsearch) . . . . .	41
16	The best Fit Solution for the Winter Olympics Data Sets using a pat- tern search optimization algorithm . . . . .	42

17	The best Fit Solution for the Winter Olympics Data Sets using a global search optimization algorithm . . . . .	43
18	The best Fit solution for the U.S.A Presidential Elections Data Sets using the Nelder-Mead optimization algorithm (fminsearch) . . . . .	45
19	The best Fit Solution for the U.S.A presidential Elections Data Sets using a pattern search optimization algorithm . . . . .	46
20	The best Fit Solution for the U.S.A. presidential Elections Data Sets using a global search optimization algorithm . . . . .	47
21	The best Fit solution for the U.S.A Elections Data Sets using the Nelder-Mead optimization algorithm (fminsearch) . . . . .	49
22	The best Fit Solution for the U.S.A Elections Data Sets using a pattern search optimization algorithm . . . . .	50
23	The best Fit Solution for the U.S.A. Elections Data Sets using a global search optimization algorithm . . . . .	51
24	Sensitivity curves . . . . .	56

## 1 INTRODUCTION

Different mathematical models for dissemination of information or rumor have been studied in the past. One such model was introduced by Daley and Kendall [4] commonly referred to as the DK model. In this model it is assumed that the population is closed and homogeneously mixing. The individuals are subdivided into three categories or classes: *naives* are those individuals who are not aware of the rumor, *spreaders* are those individuals who are actively involved in spreading the rumor, and *stiflers* are those individuals who know the rumor but have stopped spreading it. It is assumed that the rumor is spread through the population by direct contact between spreaders and naives. Whenever a spreader interacts with a naive, a naive becomes a spreader. When a spreader contacts a stifler, the spreader turns into a stifler and when a spreader meets another spreader, the initiating spreader becomes a stifler. Another fundamental assumption is that a person aware of the rumor will continue telling it until such a time when he or she decides that it is no longer worthy to be called “news” [4].

Each of the classes of individuals can be represented using three compartments as shown in Figure 1 where  $U$  represents the class of naives,  $V$  the class of spreaders and  $W$  the class of stiflers. Figure 1 shows individuals in the  $U$  compartment interacting with the naive individuals; the individuals in the  $V$  compartment who are spreaders interacting with the individuals who are stiflers;  $W$ . The parameter  $\beta$  is the contact rate,  $\mu$  is the arrival/departure rate and  $\gamma$  is the halting rate. The spread of a rumor in a closed population resembles the spread of an epidemic like in the SIR model proposed by Kermack and McKendrick [8]. The spreaders in the DK model



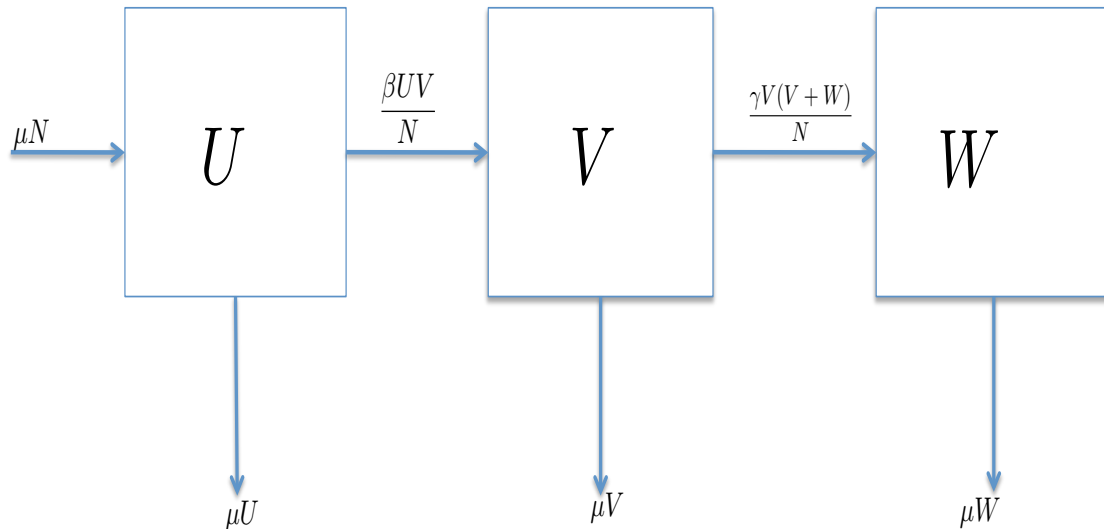


Figure 1: Compartment Model for the Naives, Spreaders and Stiflers

correspond to the infectives in the SIR model. The system of nonlinear ordinary differential equations describing the both the SIR model (equations (1)–(3)) and the DK model (equations (4)–(6)) are given below.

$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} - \mu S \quad (1)$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \mu I \quad (2)$$

$$\frac{dR}{dt} = \gamma I - \mu R \quad (3)$$

$$\frac{dU}{dt} = \mu N - \frac{\beta UV}{N} - \mu U \quad (4)$$

$$\frac{dV}{dt} = \frac{\beta UV}{N} - \frac{\gamma V(V+W)}{N} - \mu V \quad (5)$$

$$\frac{dW}{dt} = \frac{\gamma V(V+W)}{N} - \mu W \quad (6)$$

One of the main similarities between the two models represented by equations (1)–(3) and (4)–(6), respectively, is in the creation of new cases. For example, compare  $\frac{\beta SI}{N}$  in equation (2) with  $\frac{\beta UV}{N}$  in equation (5). On the other hand, the main difference is in the halting rate, that is  $\gamma I$  in equation (2) versus  $\frac{\gamma V(V+W)}{N}$  in equation (5).

The Jacobian matrix corresponding to system equations (4)–(6) is given by

$$J = \begin{bmatrix} \frac{-\beta V}{N} - \mu & \frac{-\beta U}{N} & 0 \\ \frac{\beta V}{N} & \frac{\beta U}{N} - \frac{\gamma(V+W)}{N} - \frac{\gamma V}{N} - \mu & \frac{-\gamma V}{N} \\ 0 & \frac{\gamma(V+W)}{N} + \frac{\gamma V}{N} & \frac{\gamma V}{N} - \mu \end{bmatrix}.$$

The Jacobian matrix evaluated at  $(N, 0, 0)$  reduces to

$$J = \begin{bmatrix} -\mu & -\beta & 0 \\ 0 & \beta - \mu & 0 \\ 0 & 0 & -\mu \end{bmatrix}.$$

If  $\beta < \mu$  then this implies that all eigenvalues have negative real parts. The basic reproduction number for the *DK* model  $\mathcal{R}_0 = \frac{\beta}{\mu}$ , ([3]) implying that when  $\mathcal{R}_0 < 1$ , then the equilibrium point  $(N, 0, 0)$  is stable.

### 1.1 Basic Reproductive Number for the SIR Model versus the DK Model

The Basic Reproductive Number of the *SIR* model is  $\mathcal{R}_0^{SIR} = \frac{\beta}{\gamma + \mu}$  [8], while for the *DK* model the basic reproductive number is  $\mathcal{R}_0^{DK} = \frac{\beta}{\mu}$ . Therefore we can see that

$$\mathcal{R}_0^{SIR} < \mathcal{R}_0^{DK}.$$

The numerical solutions for the SIR model versus the DK model are shown in Figure 2 below. The SIR model represented by Figure 2 (a) shows that the infective has suffered extinction and all individuals in the population are susceptible hence the disease free equilibrium is stable. The DK model represented by Figure 2 (b) shows that the naives, spreaders, and stiflers are in coexistence hence the endemic equilibrium is stable. The same parameter values and initial conditions were employed while computing numerical solutions of equations (1)–(3) and (4)–(6), respectively. Figure 2 (a) shows that the disease free equilibrium is stable in the SIR model while Figure 2 (b) shows that the endemic equilibrium is stable in the DK model. The parameter values are  $\beta = 1.5$ ,  $\mu = 1.0$ ,  $\gamma = 1.0$ . The initial conditions  $N(0) = 100$ ,  $S(0) = 0.5 \cdot N$ ,  $I(0) = 0.5 \cdot N$ ,  $R(0) = 0$  and  $U(0) = 0.5 \cdot N$ ,  $V(0) = 0.5 \cdot N$ ,  $W(0) = 0$ .

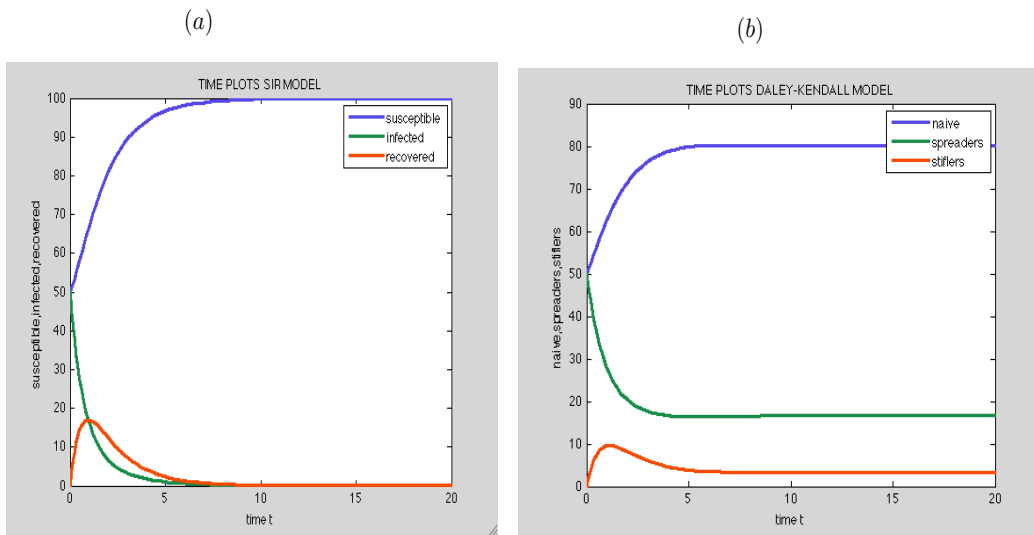


Figure 2: The SIR Model versus the DK Model plots

## 2 EXTENSIONS TO THE DALEY-KENDALL MODEL

In order to understand the dynamics of the Daley-Kendall model [4], we consider two models represented by systems of equations: one with a linear halting rate  $\alpha v$  and the other with a non-linear halting rate  $\frac{bv}{1+v}$  for spreaders. We look at the effect the halting rate has in the qualitative behavior of the models. The qualitative behavior of these models will provide a basis for the understanding of the DK model and its extensions; in particular it will establish a dynamic transition: from stable equilibrium to stable limit cycle.

### 2.1 Halting Rate of Spreaders Leading to Stable Equilibrium

Consider the non-linear model given by the system of differential equations (7)–(8)

$$\frac{du}{dt} = au\left(1 - \frac{u}{K}\right) - \beta uv \quad (7)$$

$$\frac{dv}{dt} = \beta uv - \alpha v \quad (8)$$

where  $u$  and  $v$  denote the naive individuals and spreaders respectively, while  $a$ ,  $\beta$ ,  $K$ , and  $\alpha$  are the intrinsic factor, contact rate, carrying capacity and arrival/departure rate, respectively. We can rule out the existence of limit cycles for this model by using Dulac's criterion [9].

Let the Dulac function be given by

$$B = \frac{1}{uv}.$$

By using the Dulac's Criterion, the system represented by equations (7) and (8) reduces to

$$\frac{\partial}{\partial u} \left( \frac{auK - au^2}{uvK} - \frac{\beta uv}{uv} \right) + \frac{\partial}{\partial v} \left( \frac{\beta uv - \alpha v}{uv} \right)$$

This reduces to  $-a/(Kv) < 0$ . Given  $a, k > 0$ , by application of Dulac's Criterion the system does not support stable limit cycles.

We can compute the equilibrium points [7] for the model represented by the system of equations (7)–(8) and also investigate the behavior of the model at each of the equilibrium points [5]. This system has the following equilibrium points:  $(0, 0)$ ,  $(K, 0)$  and  $(\frac{\alpha}{\beta}, \frac{a(\beta K - \alpha)}{\beta^2 K})$ .

Evaluating the Jacobian matrix of the system represented by equations (7)–(8), at the first equilibrium point  $(0, 0)$  we obtain

$$J(0, 0) = \begin{bmatrix} a & 0 \\ 0 & -\alpha \end{bmatrix}.$$

For any value of  $\alpha, a > 0$ , the determinant  $\Delta < 0$ , implying that the equilibrium point  $(0, 0)$  is unstable.

Evaluating the Jacobian matrix of the system represented by equations (7)–(8), at the second equilibrium point  $(K, 0)$ , we obtain

$$J(K, 0) = \begin{bmatrix} -a & -\beta K \\ 0 & \beta K - \alpha \end{bmatrix}.$$

For any value of  $\alpha, \beta, K > 0$ , if  $\beta K < \alpha$  then we have all eigenvalues with negative real parts. Moreover, if  $\beta K < \alpha$  then the determinant  $\Delta > 0$  and trace  $T < 0$ , hence the equilibrium point  $(K, 0)$  is stable.

Evaluating the Jacobian matrix of the system represented by equations (7)–(8) at the third equilibrium point  $\left(\frac{\alpha}{\beta}, \frac{a(\beta K - \alpha)}{\beta^2 K}\right)$ , we obtain

$$J\left(\frac{\alpha}{\beta}, \frac{a(\beta K - \alpha)}{\beta^2 K}\right) = \begin{bmatrix} a - \frac{2a\alpha}{\beta K} - \frac{a(\beta K - \alpha)}{\beta K} & -\alpha \\ \frac{\alpha(\beta K - \alpha)}{\beta K} & 0 \end{bmatrix}.$$

The determinant  $\Delta = \frac{\alpha^2(\beta K - \alpha)}{\beta K}$  and trace  $T = a - \frac{2a\alpha}{\beta K} - \frac{a(\beta K - \alpha)}{\beta K}$ . For any value of  $\alpha, a, \beta, K > 0$ , if  $\beta K - \alpha > 0$  and  $a < \left(\frac{2a\alpha}{\beta K} + \frac{a(\beta K - \alpha)}{\beta K}\right)$  then  $\Delta > 0$  and  $T < 0$  hence the equilibrium point  $\left(\frac{\alpha}{\beta}, \frac{a(\beta K - \alpha)}{\beta^2 K}\right)$  is stable.

Figure 3 shows numerical solutions illustrating the existence of stable equilibrium for the model represented by the system of equations (7)–(8). The graphs in Figure 3 (b) are trajectories in the phase plane i.e., plots of  $v(t)$  versus  $u(t)$  common graphs when looking at equilibria. All the trajectories converge at the stable equilibrium point. On the other hand, Figure 3 (a) displays time solution plots, i.e.,  $u(t)$  versus  $t$ . The parameter values used are  $a = 2, K = 3, \beta = 2$  and  $\alpha = 1.5$ . The initial conditions  $u(0) = 3$  and  $v(0) = 2$ . The determinant;  $\Delta = 1.69$  and trace;  $T = -1$ .

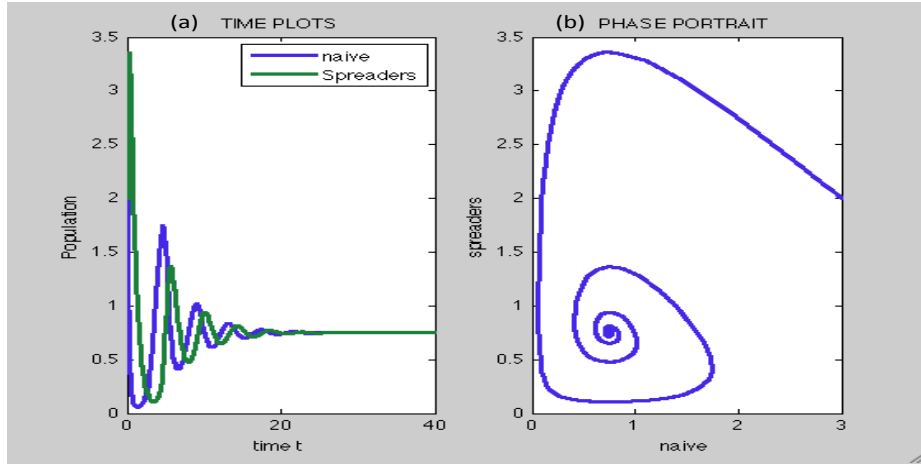


Figure 3: Time Plots and Phase Portrait for Stable Equilibrium

## 2.2 Halting Rate of Spreaders Leading to Stable Limit Cycles

Consider the non-linear model given by the system of differential equations (9)–(10) below. Let us analyze the behavior of the system after the introduction of a non-linear halting rate  $\frac{bv}{1+v}$  in equation (10).

$$\frac{du}{dt} = au\left(1 - \frac{u}{K}\right) - \beta uv \quad (9)$$

$$\frac{dv}{dt} = \beta uv - \frac{bv}{1+v}. \quad (10)$$

We can compute the equilibrium points for this system and also investigate the behavior of the model at each of the equilibrium points. This system has the following equilibrium points:  $(0, 0)$ ,  $(K, 0)$ ,  $\left(\frac{C+D}{2a}, \frac{b}{\beta\left(\frac{C+D}{2a}-1\right)}\right)$  and  $\left(\frac{-C+D}{2a}, \frac{b}{\beta\left(\frac{-C+D}{2a}-1\right)}\right)$  where  $C = \sqrt{(-aK - \beta K)^2 - 4abK}$  and  $D = aK + \beta K$ .

Evaluating the Jacobian matrix of the system of equations (9)–(10) at the first equilibrium point  $(0, 0)$ , we obtain

$$J(0, 0) = \begin{bmatrix} a & 0 \\ 0 & -b \end{bmatrix}.$$

This Jacobian matrix has two real eigenvalues with opposite signs. For any value of  $a, b > 0$ , the determinant  $\Delta = -ab < 0$  hence the equilibrium point  $(0, 0)$  is unstable [9].



Evaluating the Jacobian matrix of the system of equations (9)–(10) at the second equilibrium point  $(K, 0)$ , we obtain

$$J(K, 0) = \begin{bmatrix} -a & -\beta K \\ 0 & \beta K - \alpha \end{bmatrix}.$$

For any value of  $\alpha, a, \beta, K > 0$ , if  $\beta K < \alpha$  then all eigenvalues have negative real parts and determinant  $\Delta > 0$  and trace  $T < 0$  hence the equilibrium point  $(K, 0)$  is stable.

Similarly, at the third equilibrium point  $\left(\frac{C+D}{2a}, \frac{b}{\beta(\frac{C+D}{2a}-1)}\right)$  we evaluate the Jacobian matrix of the system of equations (9)–(10) and we obtain

$$J\left(\frac{C+D}{2a}, \frac{b}{\beta(\frac{C+D}{2a}-1)}\right) = \begin{bmatrix} \frac{-a\bar{u}_3}{K} & -\beta\bar{u}_3 \\ \beta\bar{v}_3 & \frac{\beta\bar{v}_3}{(1+\bar{v}_3)^2} \end{bmatrix},$$

where  $\begin{bmatrix} \bar{u}_3 \\ \bar{v}_3 \end{bmatrix} = \begin{bmatrix} \frac{C+D}{2a} \\ \frac{b}{\beta(\frac{C+D}{2a}-1)} \end{bmatrix}$ .

The determinant  $\Delta = \beta^2\bar{u}_3\bar{v}_3 - \frac{a\bar{u}_3\bar{v}_3}{K(1+\bar{v}_3)^2}$  and trace  $T = \frac{\beta\bar{v}_3}{(1+\bar{v}_3)^2} - \frac{a\bar{u}_3}{K}$ . For any value of  $\alpha, a, \beta, K > 0$ , if  $\frac{a\bar{u}_3\bar{v}_3}{K(1+\bar{v}_3)^2} < \beta^2\bar{u}_3\bar{v}_3$  and  $\frac{a\bar{u}_3}{K} > \frac{\beta\bar{v}_3}{(1+\bar{v}_3)^2}$  then  $\Delta > 0$  and  $T < 0$ , hence this equilibrium point  $\begin{bmatrix} \frac{C+D}{2a} \\ \frac{b}{\beta(\frac{C+D}{2a}-1)} \end{bmatrix}$  is stable. However, we must comment on the fact that  $\begin{bmatrix} \bar{u}_3 \\ \bar{v}_3 \end{bmatrix}$  lies on the fourth quadrant of the  $u - v$  plane, and therefore has no applicable meaning. Because  $\bar{v}_3 < 0$ , we are only interested in the positive value of the spreaders.

Let us now consider the fourth equilibrium point  $\left(\frac{-C+D}{2a}, \frac{b}{\beta(\frac{-C+D}{2a}-1)}\right)$ . This is the only equilibrium point in the first quadrant denoted as

$$\begin{bmatrix} \bar{u}_4 \\ \bar{v}_4 \end{bmatrix} = \begin{bmatrix} \frac{-C+D}{2a} \\ \frac{b}{\beta(\frac{-C+D}{2a}-1)} \end{bmatrix}.$$

We now need to prove by Poincare' Bendixson theorem [9] that at the equilibrium point  $\left(\frac{-C+D}{2a}, \frac{b}{\beta\left(\frac{-C+D}{2a}-1\right)}\right)$ , the system undergoes a stable limit cycle.

Poincare' -Bendixson theorem is explained in [9] and here we quote this theorem:

“Suppose that

- $R$  is a closed, bounded subset of a plane;
- $\frac{dx}{dt} = f(x)$  is a continuously differentiable vector field of an open set containing  $R$ ;
- $R$  does not contain any fixed points and;
- There exists a trajectory  $P$  that is ”confined” in  $R$ , in the sense that it starts in  $R$  and stays in  $R$  for all future time. Therefore, either  $P$  is a closed orbit as  $t \rightarrow \infty$  or it spirals towards a closed orbit [2]. In either case,  $R$  contains a closed orbit.”

Note that Poincare-Bendixson Theorem [9] is used to establish the existence of closed orbits in particular systems. When applying the Poincare-Bendixson Theorem [9], it is easy to satisfy the first three conditions:

- $R$  is a closed, bounded subset of a plane;
- $\frac{dx}{dt} = f(x)$  is a continuously differentiable vector field of an open set containing  $R$ ;
- $R$  does not contain any fixed points.

What we need to do is to show that all solutions are trapped in a bounded region away from the equilibrium point  $\left(\frac{-C+D}{2a}, \frac{b}{\beta\left(\frac{-C+D}{2a}-1\right)}\right)$  by showing that it is repelling.

We should note that a vector field is an assignment of a vector to each point in a subset of Euclidean space and can be visualized as a collection of arrows with a given magnitude and direction each attached to a point in the plane. The nullclines are sets of points in the direction field such that the derivative is equal to zero [7].

For this model, equations (9)–(10), we first need to start by considering the direction of the vector fields (note that they are made by assigning a three dimensional vector to every point in region); then we construct the trapping region. Note that in plotting Figure 4 below, by definition, the arrows are vertical on the  $\frac{du}{dt} = 0$  nullcline, and horizontal on the  $\frac{dv}{dt} = 0$  nullcline. The direction of flow is determined by the signs of  $\frac{du}{dt}$  and  $\frac{dv}{dt}$ . Thus, we construct the trapping region.

Figure 4 below is a plot of the nullclines and the trapping region in the first quadrant denoted by dotted lines and vector field direction. Every trajectory starts within the trapping region and moves to its interior and remains there all the time as the system evolves.

Figure 4 has regions  $R_1, R_2, R_3$  and  $R_4$ . We shall look at the direction of flow of vector field in each of the four regions and finally concentrate on the trapping region  $R_1$  corresponding to  $R$  (a closed, bounded subset of a plane). Figure 4 shows the trapping region taking into consideration the boundaries made by the respective nullclines corresponding to the naives and spreaders in the model (9)–(10).

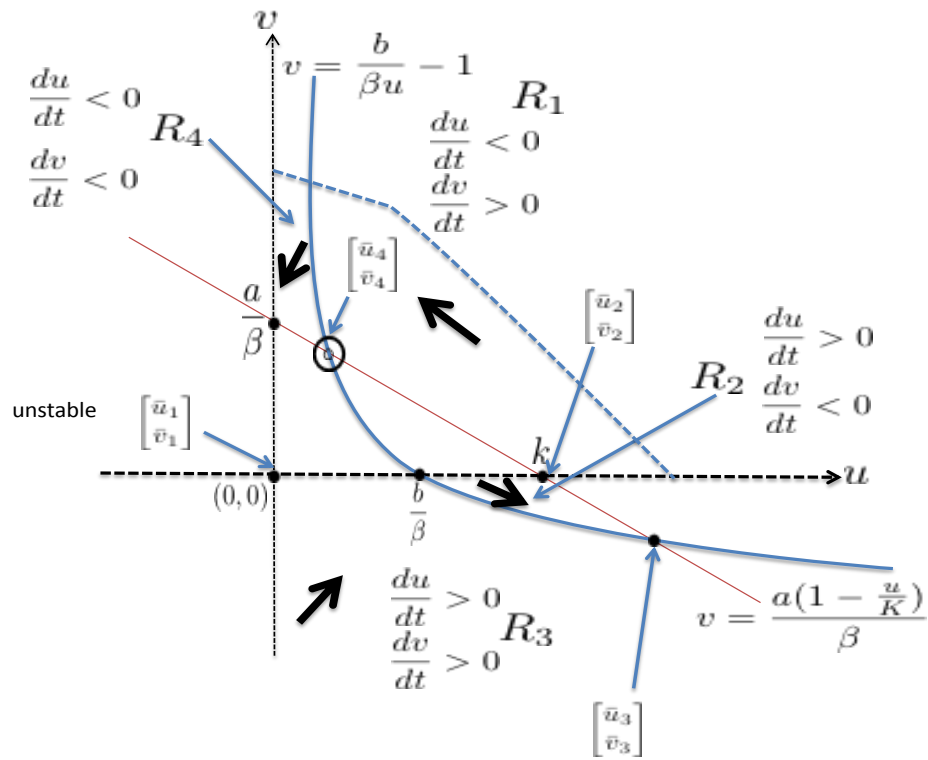


Figure 4: Trapping Region

Now consider the region bounded by the dashed lines shown in figure 4 above. We claim that it is a trapping region  $R$ . To show this, we have to show that all vectors move away from the boundary point into the box. Consider the Rumor model given by the system of differential equations (9) and (10). We need to prove that a closed orbit exists if  $a, b, \beta$  and  $K$  satisfy an appropriate condition other than  $a, b, \beta, K > 0$ . We now have to find a condition under which the equilibrium point is a repeller, i.e., an unstable node or spiral.

Equilibria are computed by setting the right-side of equations (9)–(10) to zero and and by solving for  $(u, v)$ :

$$u \left[ a \left( 1 - \frac{u}{K} \right) - \beta v \right] = 0 \quad (11)$$

$$v \left[ \beta u - \frac{b}{1+v} \right] = 0 \quad (12)$$

By equations (11)–(12), we see that the trivial equilibrium given by  $\begin{bmatrix} \bar{u}_1 \\ \bar{v}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$  is a solution.

Excluding the trivial equilibrium point  $(0, 0)$ , the equilibrium equations reduce to

$$a \left( 1 - \frac{\bar{u}_4}{K} \right) - \beta \bar{v}_4 = 0 \quad (13)$$

$$\beta \bar{u}_4 - \frac{b}{1 + \bar{v}_4} = 0. \quad (14)$$

Expressing both equations (13) and (14) in terms of  $\bar{v}_4$ , gives

$$\bar{v}_4 = \frac{a(K - \bar{u}_4)}{\beta K} \quad (15)$$

$$\bar{v}_4 = \frac{b}{\beta \bar{u}_4} - 1. \quad (16)$$

where  $\bar{v}_4 = \frac{b}{\beta(\frac{-C+D}{2a}-1)}$ .

Evaluating the Jacobian matrix at the interior equilibrium  $\left(\frac{-C+D}{2a}, \frac{b}{\beta(\frac{-C+D}{2a}-1)}\right)$

i.e.  $\bar{u}_4 > 0$  and  $\bar{v}_4 > 0$ , the Jacobian matrix reduces to

$$J \left( \frac{-C+D}{2a}, \frac{b}{\beta(\frac{-C+D}{2a}-1)} \right) = \begin{bmatrix} -\frac{a\bar{u}_4}{K} & -\beta\bar{u}_4 \\ \beta\bar{v}_4 & \frac{\beta\bar{v}_4}{(1+\bar{v}_4)^2} \end{bmatrix}.$$

In order to determine an expression/equation for the trace of  $J$  we need to simplify the Jacobian matrix by expressing it in terms of  $\bar{u}_4$ , where  $\bar{u}_4 = \frac{-C+D}{2a}$ . Using equations (13) and (14) reduces the Jacobian matrix now to

$$J = \begin{bmatrix} -\frac{a\bar{u}_4}{K} & -\beta\bar{u}_4 \\ \frac{a(K-\bar{u}_4)}{K} & \frac{\beta\bar{u}_4(b-\beta\bar{u}_4)}{b} \end{bmatrix},$$

and therefore its trace is given by

$$T = \frac{\bar{u}_4}{Kb} [K\beta b - K\beta^2\bar{u}_4 - ab]. \quad (17)$$

We should note that the equilibrium point  $\left(\frac{-C+D}{2a}, \frac{b}{\beta(\frac{-C+D}{2a}-1)}\right)$  is unstable for  $T > 0$  and stable for  $T < 0$  [9], whenever the determinant is positive. We need to find an expression that satisfies the condition for the dividing line  $T = 0$ . Using the expression for the trace (17), we see a condition that the trace  $T$  equals zero if and only if  $\bar{u}_4 = \frac{\beta Kb - ab}{\beta^2 K}$ .

From equations (13) and (14) we have that  $\bar{v}_4 = \frac{aK - a\bar{u}_4}{\beta K}$  or  $\bar{v}_4 = \frac{b}{\beta\bar{u}_4} - 1$ , implying that

$$\frac{aK - a\bar{u}_4}{\beta K} = \frac{b}{\beta\bar{u}_4} - 1 \quad (18)$$

Substituting  $\bar{u}_4 = \frac{-C+D}{2a}$  into equation (18), gives the following expression:

$$ab\beta^3 K^3 - ab^2\beta^2 K^2 + 2a^2b^2\beta K - a^2b\beta^2 K - a^3b^2 - b\beta^3 K^3 + b\beta^4 K^3 - ab\beta^3 K^2 = 0 \quad (19)$$

Solving for the parameter  $b$  in terms of other parameters we find that

$$b_c = -\frac{\beta^2 K^2 (a^2 + a(\beta - \beta k) - (\beta - 1)\beta K)}{a(a - \beta K)^2} \quad (20)$$

Using equation (20), we show that the equilibrium point is unstable for  $T > 0$ , if  $b > b_c$  and stable for  $T < 0$  if  $b < b_c$ , respectively. The dividing line  $T = 0$  therefore occurs if and only if equation (20) is satisfied. An important result we have seen here is that by using Poincare'-Bendixson theorem, we were able to identify a trapping region and proved that the equilibrium point  $\left(\frac{-C+D}{2a}, \frac{b}{\beta(\frac{-C+D}{2a}-1)}\right)$  is actually a repeller, and thus the trapping region can sustain a stable limit cycle [9].

To have a clear picture of the qualitative behavior of the model around the point  $b_c$  above, we can reduce  $b_c$  into a function of  $a$  by assigning values to the free variables  $\beta$  and  $K$ . For parameter values  $\beta = 1$  and  $K = 5$  the equation for  $b_c$  reduces to

$$b_c = \frac{-25(a^2 - 4a)}{a(a - 5)^2}. \quad (21)$$

and this defines a curve in  $(a, b)$  space, as shown by Figure 5. It shows the parameter space, i.e.,  $(a, b)$  and the curve defined by equation (20). Two regions are identified in the parameter space, one that supports stable equilibrium, and another one with stable limit cycle.

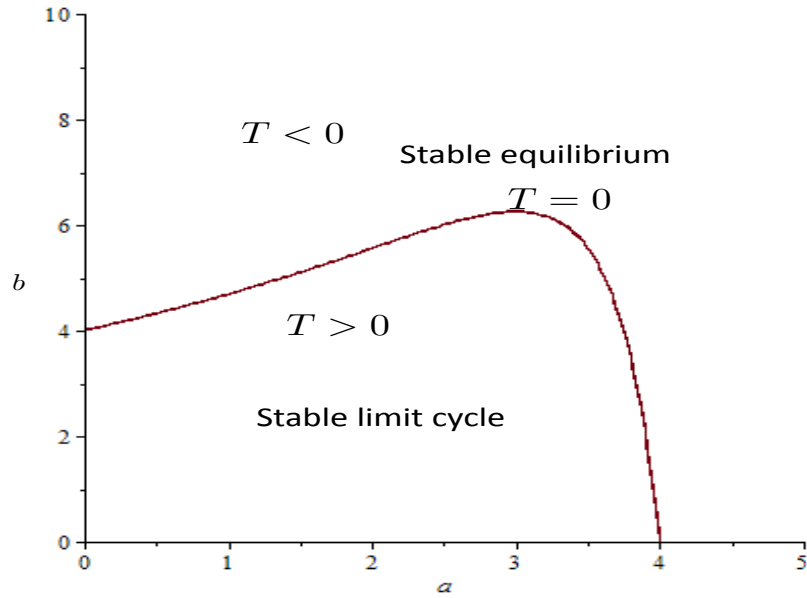


Figure 5: Trace Curve in the  $(a, b)$  Parameter Space



For parameters in the region corresponding to  $T > 0$  in Figure 5, we should note that there is a guarantee that the model represented by the system of equations (9)–(10) has a stable limit cycle [9] and the numerical solutions for both the systems with linear halting rate in equation (9) and non-linear halting rate in equation (10) are considered. The time solutions and phase portraits for the system are then displayed. The stable equilibrium is displayed by Figures 6 (a)–6 (b) while the initial conditions and parameter values are  $u(0) = 5, v(0) = 2$  and  $a = 1, b = 3, \beta = 2, K = 2$  respectively. The stable limit cycle is displayed by Figures 6 (c)–6 (d) while the initial conditions and parameter values used are  $u(0) = 5$  and  $v(0) = 2$   $a = 1, b = 2, \beta = 1, K = 3$  respectively.

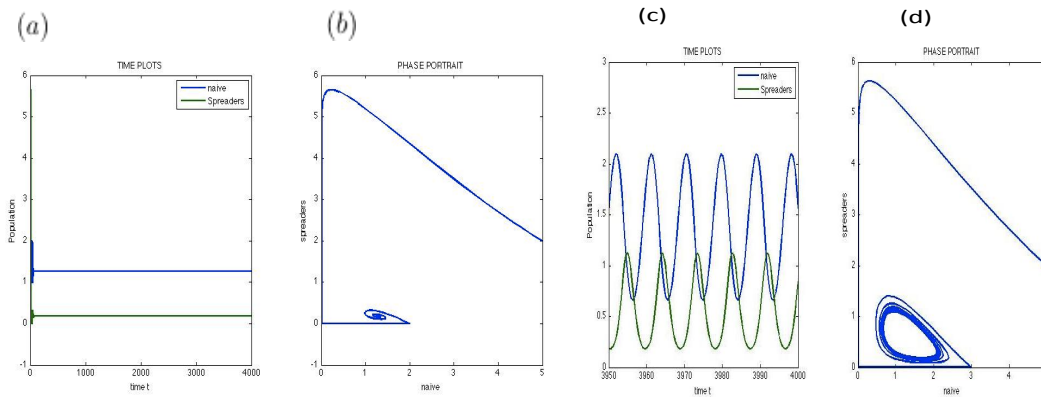


Figure 6: Time Plots and Phase Portrait for a Stable Equilibrium in Panels (a) – (b), and Stable Limit Cycle in Panels (c) – (d)

### 3 MODEL VALIDATION

In this section, we shall discuss how to estimate the parameters of the model given a data set by using ordinary least squares (OLS) [3]. We shall also perform uncertainty quantification and look at a statistical observation process that describes a data set. One of the main objectives of model validation is to compare estimates from a mathematical model of a system to the measured behavior of the system. In this section, we shall also use Ordinary Least Squares (OLS) method to validate our model represented by equations (9)–(10).

Consider a time-evolving process described by the mathematical model

$$\frac{dx}{dt} = f(x(t, \theta), \theta).$$

The function  $f$  is continuous and is defined by the right side of equations (9)–(10). We consider the model output to be a re-scaling of the number of spreaders by a scalar  $\alpha$ . Other model parameters we shall consider in our analysis are  $a$ ,  $b$ ,  $\beta$  and  $K$ . For example, in the extension model of the  $DK$  model,  $x(t, \theta) = (u(t, \theta), v(t, \theta))$ , and  $\theta = (a, b, \beta, K)$ . Then, the statistical model for the observation process for  $i = 1, \dots, n$ . is given by

$$Y_i = \alpha v(t_i, \theta_0) + \varepsilon_i.$$

where errors denoted by  $\varepsilon_i$  are assumed to be random variables satisfying the following assumptions [1]:

- the errors  $\varepsilon_i$  have a mean of zero,  $E = [\varepsilon_i] = 0$
- the errors  $\varepsilon_i$  have finite common variance,  $\text{var}(\varepsilon_i) = \sigma_0^2 < \infty$ ;

- the errors  $\varepsilon_i$  are independent identically distributed (i.e  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$  whenever  $i \neq j$ ).

### 3.1 Forward Solution and Synthetic Data

A realization of the observation process can be computed using independent draws from a normal distribution [3]. Suppose both  $\theta_0$  and  $\sigma_0^2$  are known. Then, a realization  $y_i$  is given by

$$y_i = \alpha v(t_i, \theta_0) + \varepsilon_i.$$

where  $\alpha v(t_i, \theta_0)$  is the numerical solution to the deterministic model given by equations (9)–(10) (for spreaders) evaluated at the true value of the parameter,  $\theta_0$ , and the  $\varepsilon_i$  is the error,  $\varepsilon_i \sim \mathcal{N}(0, \sigma_0^2)$ , for  $i = 1, \dots, n$ .

To explore the parameter estimation, we can use simulated data. These data are generated by setting values for  $\theta_0$  and  $\sigma_0$ . For example, let  $a = 1$ ,  $b = 2$ ,  $\beta = 1$ ,  $K = 3$  and  $\sigma_0 = 0.1203$ . In this way, we ensure that we are fitting a correct model to the data and a correct method is being used. Figure 7 displays numerical solutions  $\alpha v(t, \theta_0)$  with initial conditions  $u(0) = 5, v(0) = 2$ .

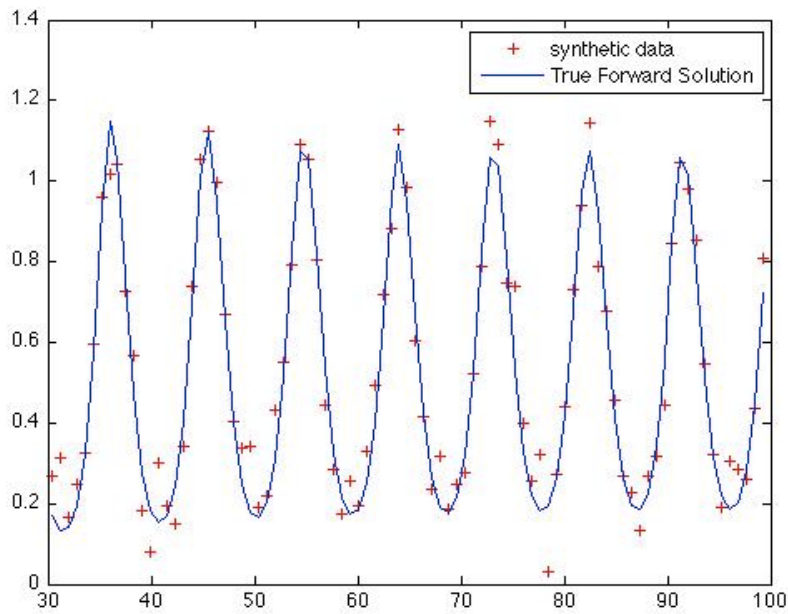


Figure 7: True Solution of Equations (9)–(10) and Synthetic Data versus Time

### 3.2 Ordinary Least Squares Estimator

The ordinary least squares method (OLS) is used to compute estimations of parameters and to fit data [1]. In this subsection we shall discuss how to use (OLS) to estimate parameters of the DK from datasets.

The estimator  $\Theta_{OLS}$  minimizes

$$\sum_{i=1}^n [Y_i - \alpha v(t_i, \theta)]^2.$$

In the limit as  $n \rightarrow \infty$ ,  $\Theta_{OLS}$  is approximately distributed according to a multivariate normal distribution, i.e.

$$\Theta_{OLS} \sim \mathcal{N}_p(\theta_0, \Sigma_0),$$

where  $\Sigma_0 = \sigma_0^2 [n\Omega_0]^{-1} \in \mathbb{R}^{p \times p}$  and  $\Omega_0 = \lim_{n \rightarrow \infty} \frac{1}{n} \chi(\theta_0)^T \chi(\theta_0)$  [1].

The  $n \times p$  matrix  $\chi(\theta_0)$  is known as the sensitivity matrix of the system and is defined as (see [3] for additional details)

$$\chi_{ij}(\theta_0) = \frac{\partial}{\partial \theta_j} \alpha v(t_i, \theta_0) |_{\theta=\theta_0}.$$

Numerical values of  $\chi(\theta)$  are readily calculated, for particular  $\theta$ , by solving both equations below (forward sensitivity equations) simultaneously from  $t = t_0$  to  $t = t_n$ :

$$\frac{dx}{dt} = f(t, x(t, \theta), \theta) \tag{22}$$

$$\frac{d}{dt} \frac{\partial x}{\partial \theta} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial \theta} \tag{23}$$

### 3.3 Uncertainty Quantification

Note that that  $\hat{\theta}_{OLS}$  is a realization of the estimator [3]  $\Theta_{OLS}$ , and is calculated using a realization  $\{y_i\}_{i=1}^n$  of the observation process  $\{Y_i\}_{i=1}^n$ . An approximation for the sampling distribution of the estimator is given by

$$\Theta_{OLS} \sim \mathcal{N}_p(\theta_0, \Sigma_0) \approx \mathcal{N}_p(\hat{\theta}_{OLS}, \hat{\Sigma}_{OLS}).$$

The covariance matrix  $\Sigma_0$  is approximated by

$$\hat{\Sigma}_{OLS} = \hat{\sigma}_{OLS}^2 [\chi(\hat{\theta}_{OLS})^T \chi(\hat{\theta}_{OLS})]^{-1}.$$

where

$$\hat{\sigma}_{OLS}^2 = \frac{1}{n-p} \sum_{i=1}^n [y_i - \alpha v(t_i, \hat{\theta}_{OLS})]^2$$

Standard errors can be used to quantify the uncertainty in the estimation

$$SE_k(\hat{\theta}_{OLS}) = \sqrt{(\hat{\Sigma}_{OLS})_{kk}} \tag{24}$$

for  $k = 1, \dots, p$ . To quantify uncertainty in the estimation we need to calculate  $\hat{\Sigma}_{OLS}$ .

For more details see [1] and references therein.

### 3.4 Real Data Sets

Real data sets can be found from google trends (<http://www.google-trend.com>). Google trends is a free public service that provides charts and data of how often words, phrases, personalities or topics providing temporal patterns about searches in which an individual can find different data sets. To get a real data set from using google trends, a person has to sign in to a google account and type in a query. This opens up a chart with a series frequency over time or in a particular region. In order to get the the right data set one has to be specific in the search term to use. For the data to be saved for model analysis purposes then it is to be downloaded as a csv file. This data set can then be copied/pasted and saved in a specific file for the purposes of analysis.

Let us now consider different sample of data sets from google trends. The data sets we shall compare include the following:

- Summer olympics
- Winter olympics
- U.S.A. presidential elections and
- U.S.A. elections (any other election not necessarily the presidential election, used to represent an expected period slightly different from four years).

We shall plot a display for all the above sample data sets from google trends and also the respective zoomed-in image for clarity as shown in Figures 8–11.

Figure 8 shows a display of the summer Olympics data sets from google trends. Panel (a) shows the actual plot while panel (b) the zoomed image for better resolution.

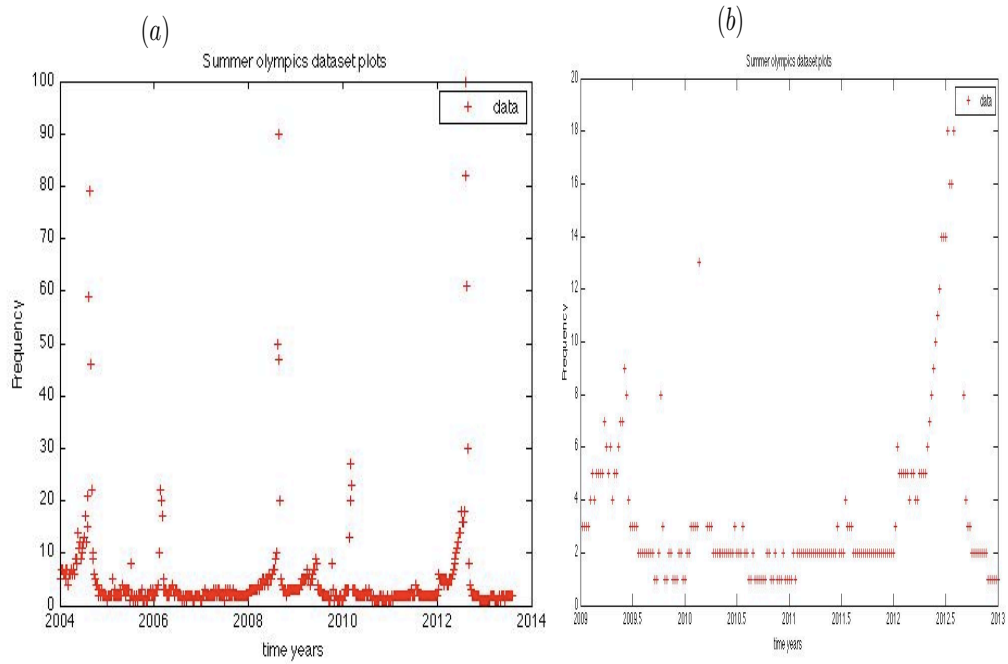


Figure 8: Plot of the Summer Olympics Data Sets



Figure 9 is a display of the winter Olympics data sets from google trends. Panel (a) shows the actual plot while panel (b) the zoomed image for better resolution.

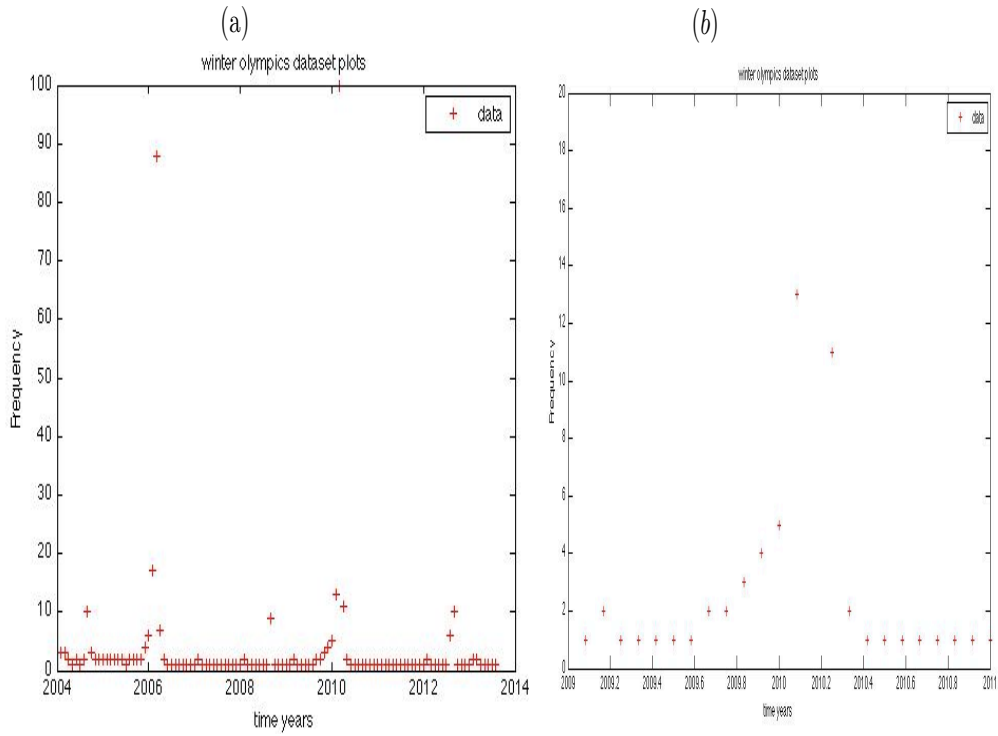


Figure 9: Plot of the Winter Olympics Data Sets

Figure 10 is a display of the U.S.A. presidential elections data sets from google trends. Panel (a) shows the actual plot while panel (b) the zoomed image for better resolution.

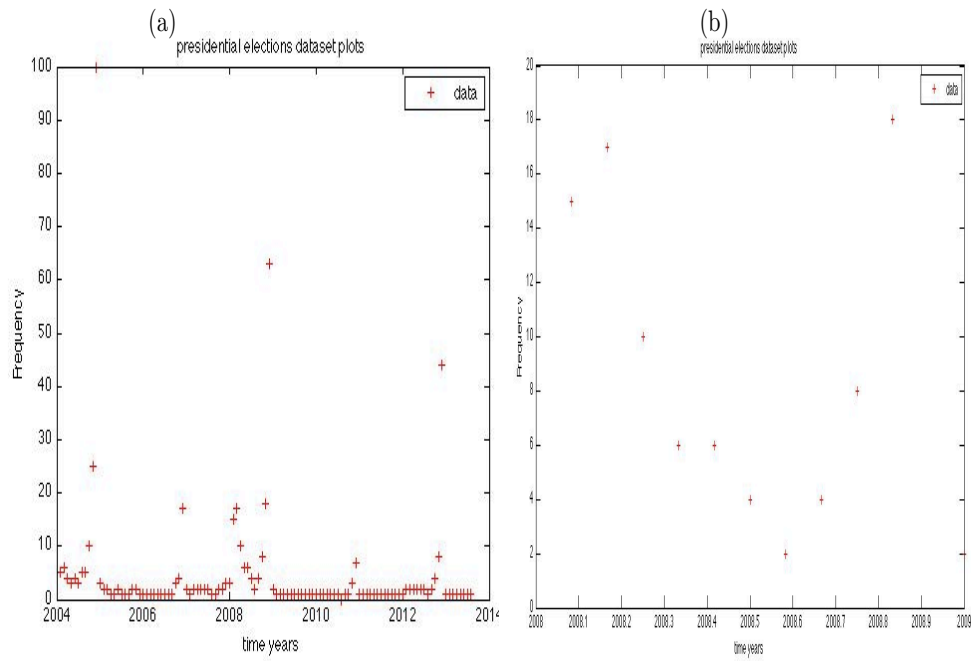


Figure 10: Plot of the U.S.A. Presidential Elections Data Sets

Figure 11 is a display of the U.S.A. elections data sets from google trends. Panel (a) shows the actual plot while panel (b) the zoomed image for better resolution.

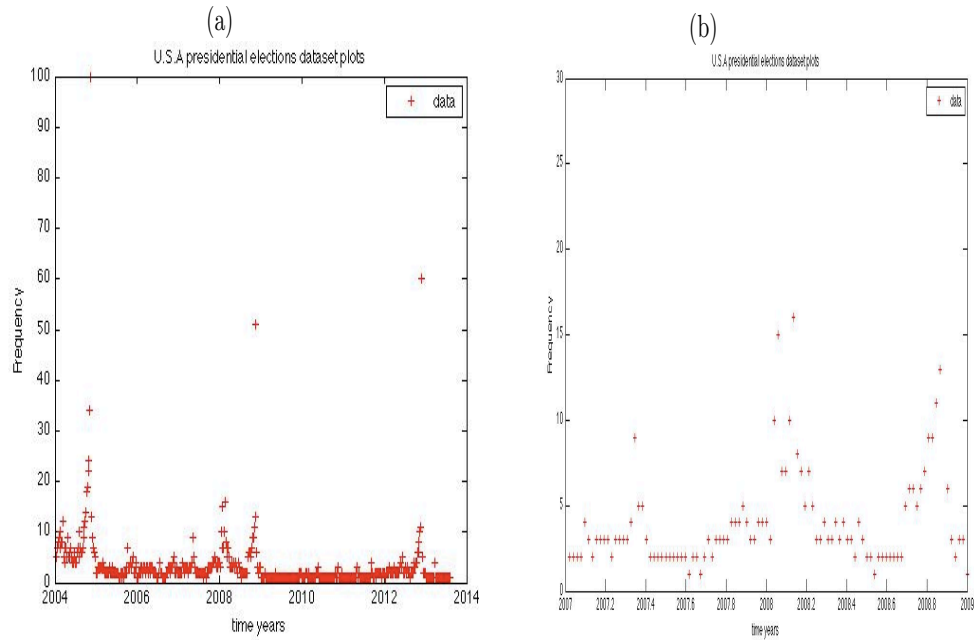


Figure 11: Plot of the U.S.A. Elections Data Sets

The time points in the data sets displayed by Figures 8–11 were measured in weeks. We notice in the general trend that there are times when the measurements remain at a low level, i.e., between 2 and 5, while at other times they reach a maximum value ranging between 60 and 90. The scale in the  $y$ -axis corresponding to frequency is in percentages. In general, these data sets were chosen because they peak after every four years. Note that we need data sets that peak every four years because the numerical solutions of our model has a period of four years.

### 3.5 Parameter Estimates and Standard Errors

Parameter estimation helps to accurately describe the behavior of the system of equations (9)–(10). We shall fit several data sets from google trends observed to be of approximately a period of four years, similar to the period of numerical solutions of the model represented by the equations (9)–(10). The following are some of the optimization algorithms we shall use to estimate the parameters and compare the respective standard errors.

- **Fminsearch:** It is the name of the MATLAB built-in function that uses the Nelder-Mead simplex algorithm and finds the minimum of a scalar function of several variables, starting at an initial estimate [6]. It uses the simplex search method; a direct search method that does not use numerical or analytic gradients. It has some disadvantages. For example, it cannot often handle discontinuity, particularly if it does not occur near the solution. It may only give local solutions and it only minimizes over the real numbers, that is, the input must only consist of real numbers and the output function must only return real numbers [6]. When the input has complex variables, they must be split into real and imaginary parts.
- **Global optimization:** As the name suggests, it finds globally the best solution of nonlinear models, i.e, it tries to find global solutions of a constrained optimization model. In this manner, it uses a global optimization tool box by generating a number of initial/starting points [6]. One such example we shall use here is pattern search.

Now we would like to display the best fit plots of numerical solutions and samples of real data sets from google trends using different optimization algorithms that we have discussed above. This, together with zoomed-in image will help us to determine how good the best fit plots are. Figure 12 shows best fit solution for the summer Olympics data set using the Nelder-Mead algorithm (fminsearch). Figure 12 (a) shows the best fit plot of the data set while Figure 12 (b) is the zoomed-in image of Figure 12 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

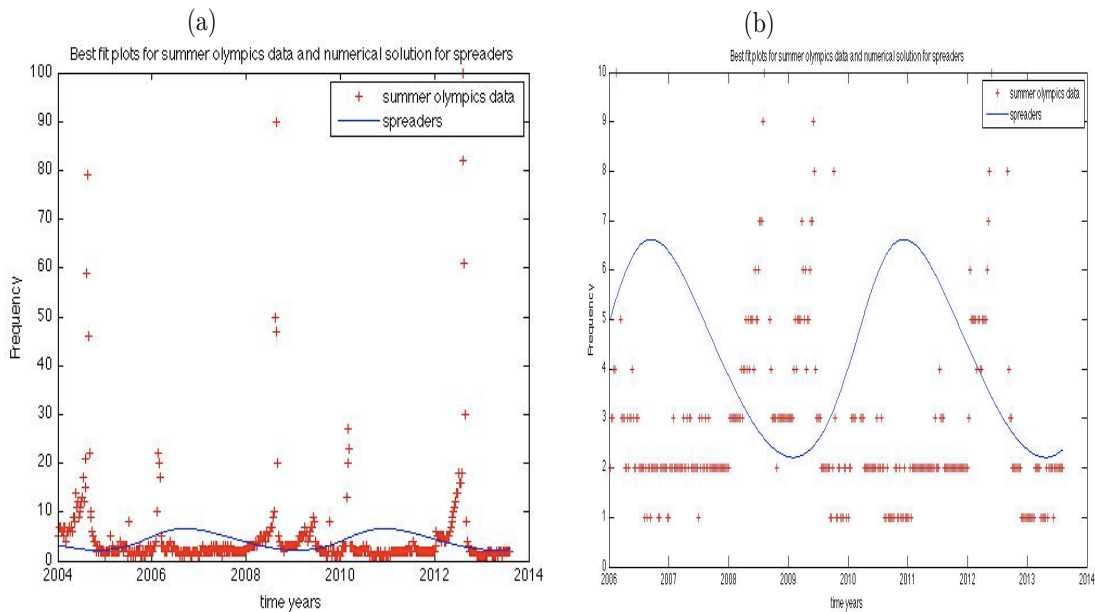


Figure 12: The best Fit Solution for the Summer Olympics Data Sets using the Nelder-Mead optimization algorithm (fminsearch)

Figure 13 shows the best fit solution for the summer Olympics data sets using a pattern search optimization. Figure 13 (a) shows the best fit plot using pattern search algorithm for the data set while Figure 13 (b) is the zoomed-in image of Figure 13 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $U(0) = 5, V(0) = 2$ .

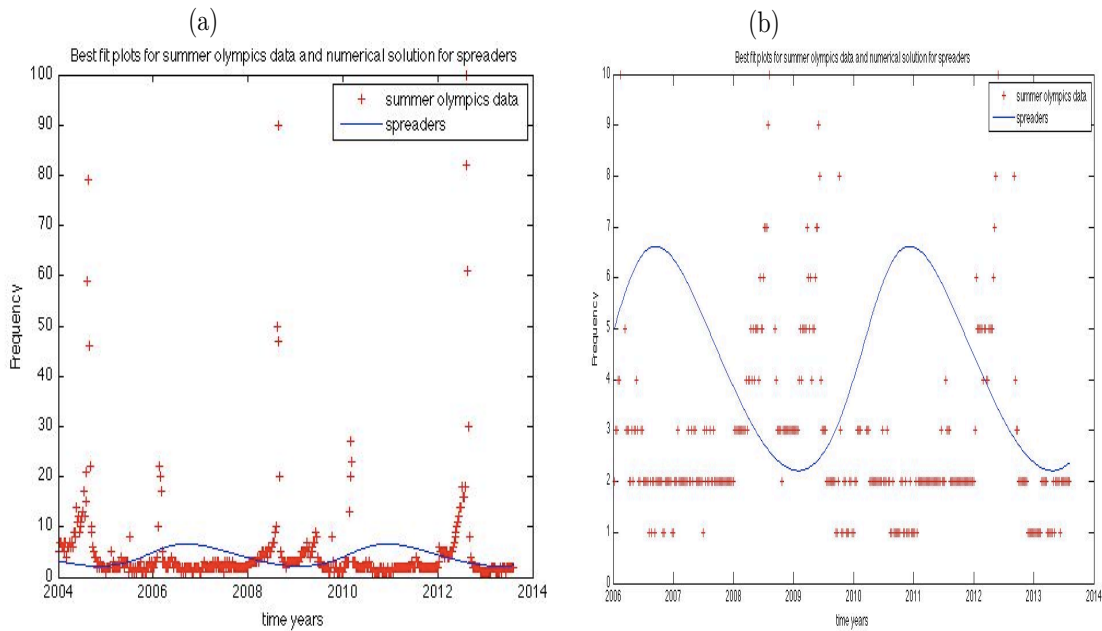


Figure 13: The best Fit Solution for the Summer Olympics Data Sets using a pattern search optimization algorithm

Figure 14 shows the best fit solution for the summer Olympics data sets using a global search optimization algorithm. Figure 14 (a) shows the best fit plot of the data set while Figure 14 (b) is the zoomed-in image of Figure 14 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

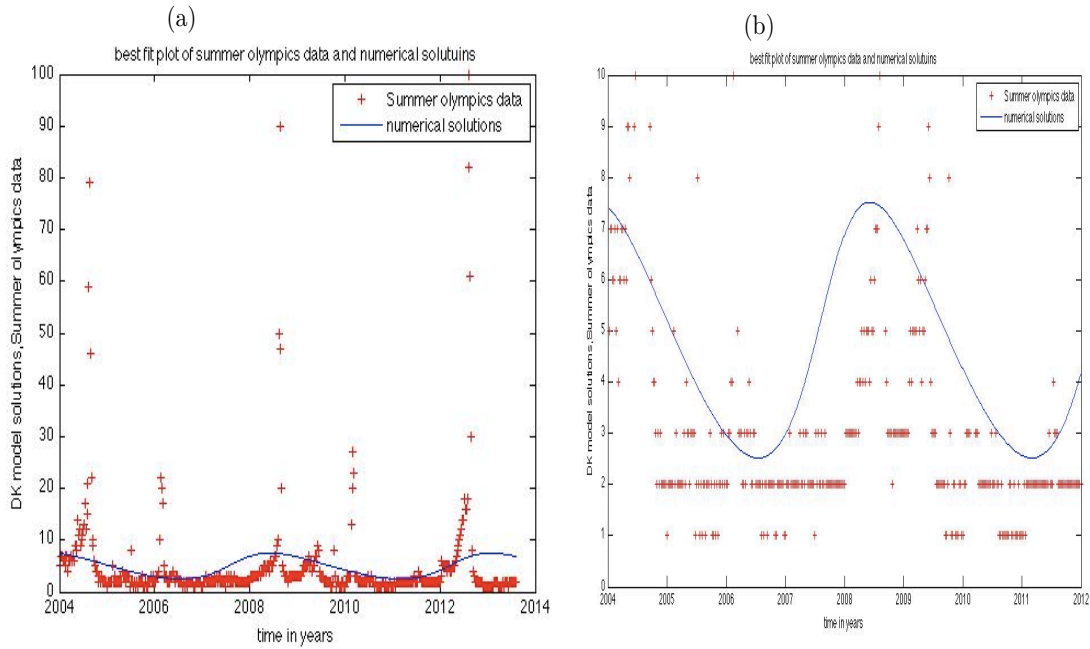


Figure 14: The best Fit Solution for the Summer Olympics Data Sets using a global search optimization algorithm

The results of the estimates and corresponding standard errors using different optimization algorithms for the summer Olympics data sets are compared and summarized by Table 1.

Table 1: Parameter estimates and standard errors obtained by three optimization algorithms: Nelder-Mead<sup>1</sup>; pattern search<sup>2</sup>; global search<sup>3</sup>

Parameter	Estimate <sup>1</sup>	S.E <sup>1</sup>	Estimate <sup>2</sup>	S.E <sup>2</sup>	Estimate <sup>3</sup>	S.E <sup>3</sup>
a	3.662	35.059	3.795	3.785	3.347	4.024
b	4.182	68.610	2.501	15.846	4.736	2.973
$\alpha$	1.720	12.563	1.482	1.257	1.341	1.653

Notice that in general, in Table 1, the Nelder-Mead algorithm (fminsearch) has the largest standard error associated with the estimate of all parameter values than the pattern search optimization algorithm and global search methods. The pattern search optimization algorithm resulted in a higher standard error for the estimate of parameter  $b$  than when the global search optimization method was used. The pattern search method also resulted in the lowest value of standard error for parameter  $\alpha$  compared to the Nelder-Mead algorithm (fminsearch) and global search method. The best fit solution tended to concentrate on the lower values of the data set while maintaining the same period.



Figure 15 shows the best fit solution for the winter Olympics data sets using the Nelder-Mead (fminsearch) optimization algorithm. Figure 15 (a) shows the best fit plot of the data set while 15 (b) is the zoomed-in image of Figure 15 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

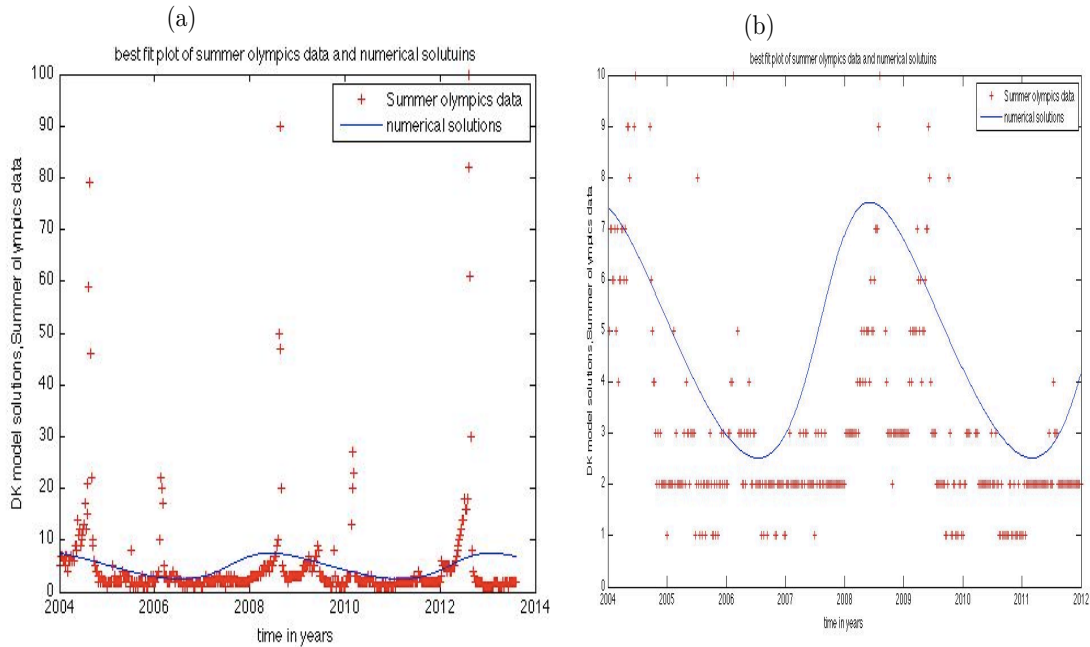


Figure 15: The best Fit Solution for the winter Olympics Data Sets using the Nelder-Mead optimization algorithm (fminsearch)

Figure 16 shows the best fit solution for the winter Olympics data sets using a pattern search optimization algorithm. Figure 16 (a) shows the best fit plot of the data set while 16 (b) is the zoomed-in image of Figure 16 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

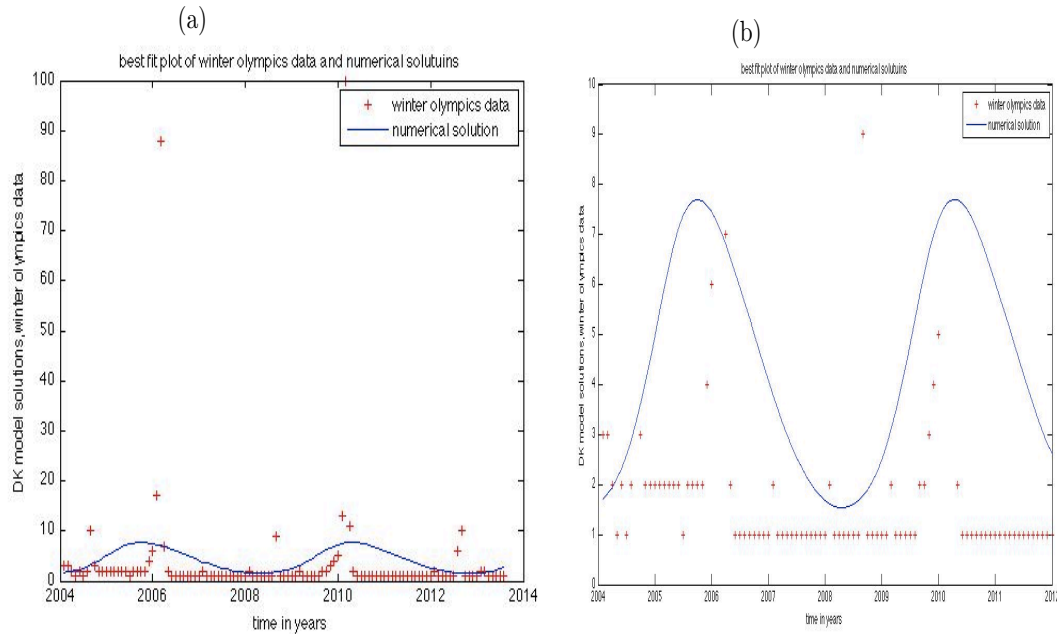


Figure 16: The best Fit Solution for the Winter Olympics Data Sets using a pattern search optimization algorithm

Figure 17 shows the best fit solution for the winter Olympics data sets using a global search optimization algorithm. Figure 17 (a) shows the best fit plot of the data set while 17 (b) is the zoomed-in image of Figure 17 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

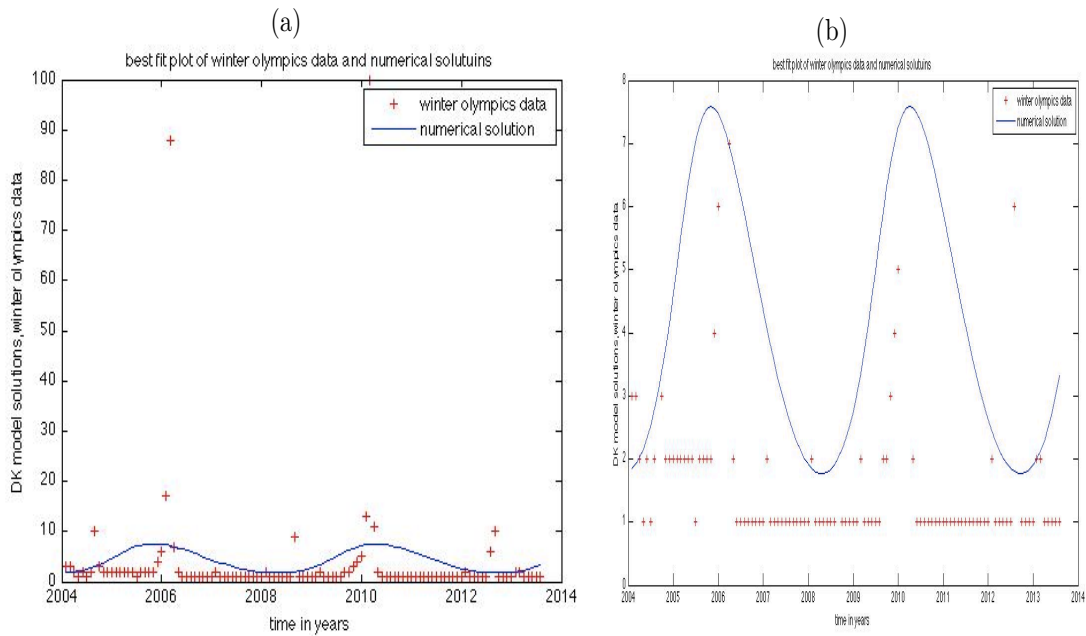


Figure 17: The best Fit Solution for the Winter Olympics Data Sets using a global search optimization algorithm

The results of the estimates and corresponding standard errors using different optimization algorithms for the winter Olympics data sets are compared and summarized by Table 2.

Table 2: Parameter estimates and standard errors displayed obtained by three optimization algorithms: Nelder-Mead<sup>1</sup>; pattern search<sup>2</sup>; global search<sup>3</sup>

Parameter	Estimate <sup>1</sup>	S.E <sup>1</sup>	Estimate <sup>2</sup>	S.E <sup>2</sup>	Estimate <sup>3</sup>	S.E <sup>3</sup>
a	3.721	32.415	3.923	3.646	3.916	3.991
b	4.621	59.314	2.681	12.457	4.329	3.014
$\alpha$	1.914	8.631	1.763	1.424	1.741	1.823

We can notice in Table 2 that the largest standard errors for all the parameter estimates resulted with the Nelder-Mead algorithm (fminsearch). Also, the global search optimization method resulted in larger standard errors than pattern search optimization algorithm. We should note that larger standard errors imply larger uncertainties in the estimates and this trend is well shown by how the best fit plot solution appears in Figures 15–17.

Figure 18 shows best fit solution for the U.S.A. presidential elections data sets using the Nelder-Mead (fminsearch) optimization algorithm. Figure 18 (a) shows the best fit plot of the data set while 18 (b) is the zoomed-in image of Figure 18 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

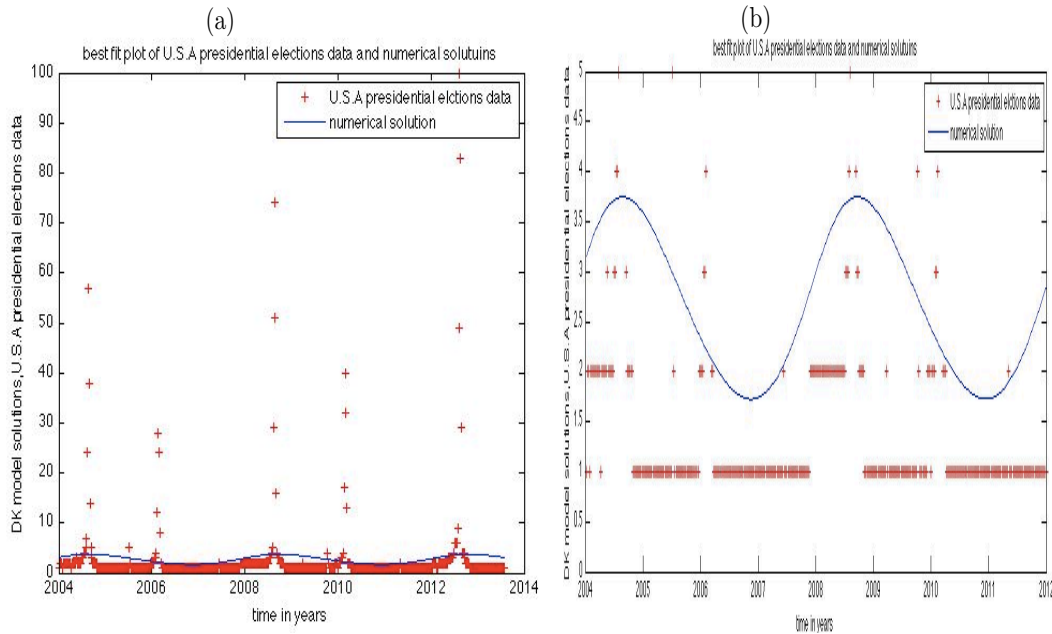


Figure 18: The best Fit solution for the U.S.A Presidential Elections Data Sets using the Nelder-Mead optimization algorithm (fminsearch)

Figure 19 shows the best fit solution for the U.S.A. presidential elections data sets using a pattern search optimization algorithm. Figure 19 (a) shows the best fit plot of the data set while 19 (b) is the zoomed-in image of Figure 19 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

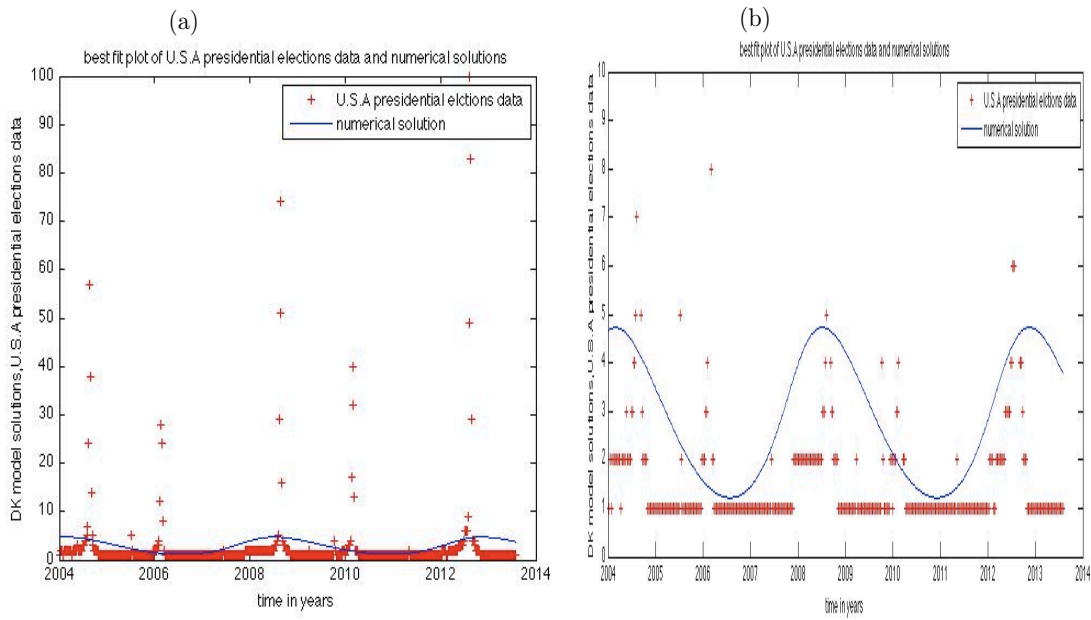


Figure 19: The best Fit Solution for the U.S.A presidential Elections Data Sets using a pattern search optimization algorithm

Figure 20 shows the best fit solution for the U.S.A. presidential elections data sets using a global search optimization algorithm. Figure 20 (a) shows the best fit plot of the data set while 20 (b) is the zoomed-in image of Figure 20 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

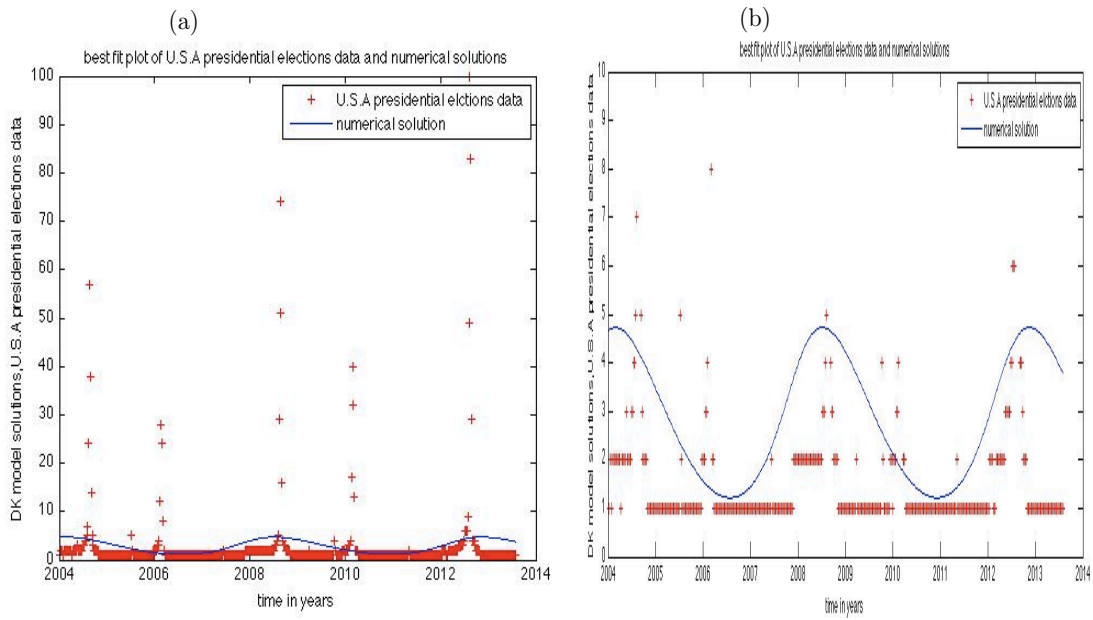


Figure 20: The best Fit Solution for the U.S.A. presidential Elections Data Sets using a global search optimization algorithm

The results of the estimates and corresponding standard errors using different optimization algorithms for the U.S.A presidential elections data sets are compared and summarized by Table 3.

Table 3: Parameter estimates and standard errors obtained by three optimization algorithms: Nelder-Mead<sup>1</sup>; pattern search<sup>2</sup>; global search<sup>3</sup>

Parameter	Estimate <sup>1</sup>	S.E <sup>1</sup>	Estimate <sup>2</sup>	S.E <sup>2</sup>	Estimate <sup>3</sup>	S.E <sup>3</sup>
a	4.111	39.303	4.001	3.219	3.628	3.991
b	3.922	54.612	2.462	16.471	4.521	2.893
$\alpha$	1.801	11.114	1.615	1.390	1.616	1.596

From Table 3 we can see that for the U.S.A. presidential elections data sets using pattern search optimization algorithm yielded larger standard errors for all the parameter estimates than a global optimization. Similarly, the Nelder-Mead algorithm (fminsearch) has the largest standard errors than either the pattern search or the global search. The best fit solution tended to mimic the peak of the data set but at lower levels of data values.



Figure 21 below shows best fit for the U.S.A. elections data sets using the Nelder-Mead (fminsearch) optimization algorithm. Figure 21 (a) shows the best fit plot of the data set while 21 (b) is the zoomed-in image of Figure 21 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

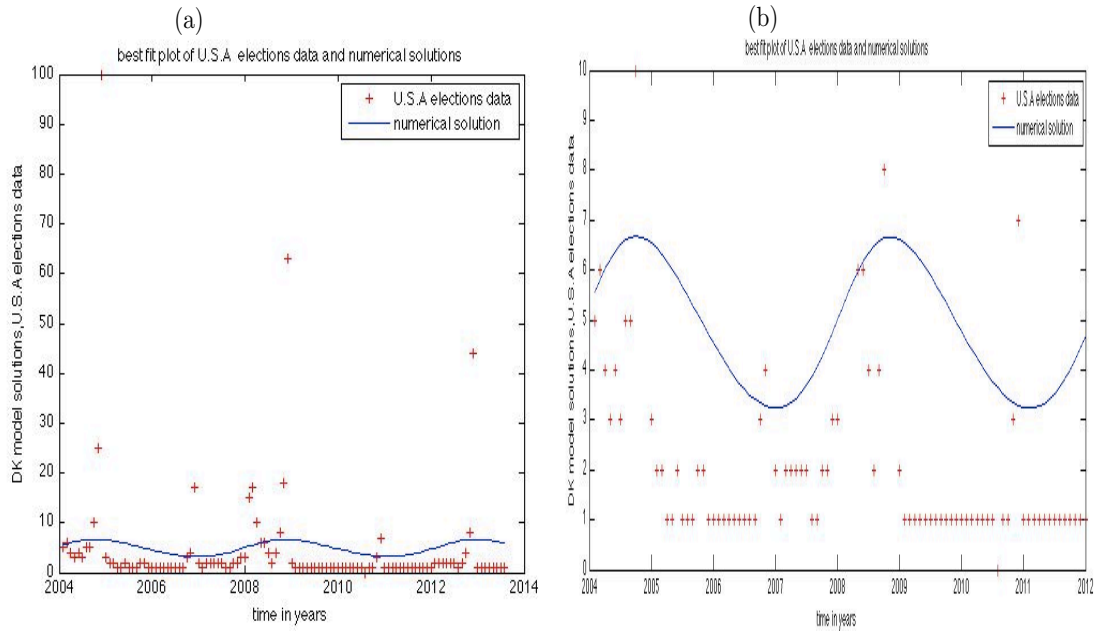


Figure 21: The best Fit solution for the U.S.A Elections Data Sets using the Nelder-Mead optimization algorithm (fminsearch)

Figure 22 shows the best fit solution for the U.S.A. elections data sets using a pattern search optimization algorithm. Figure 22 (a) shows the best fit plot of the data sets while 22 (b) is the zoomed-in image of Figure 22 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

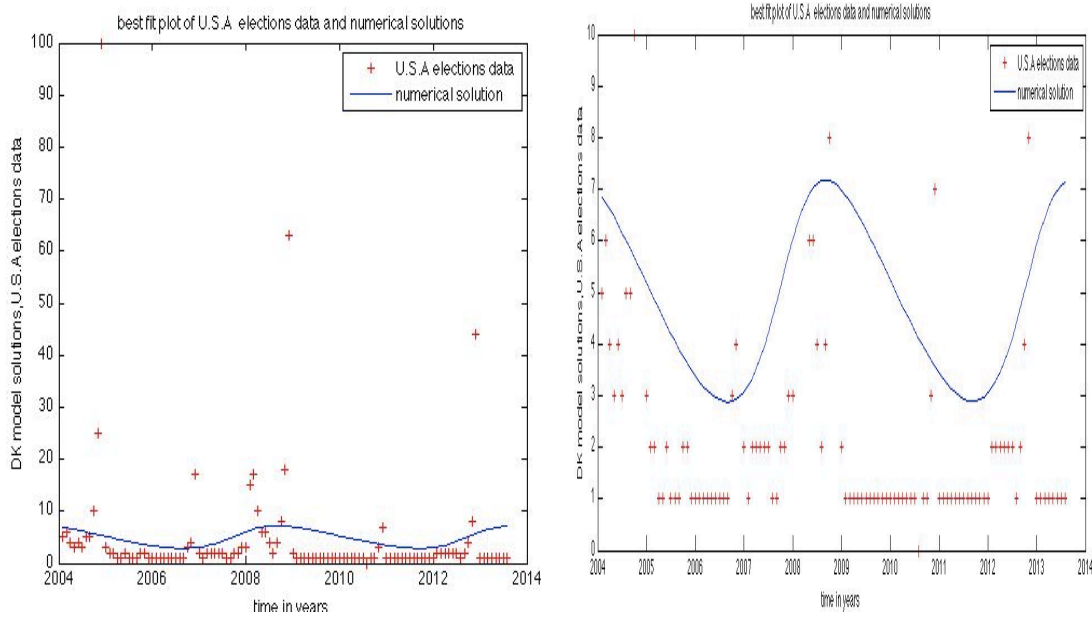


Figure 22: The best Fit Solution for the U.S.A Elections Data Sets using a pattern search optimization algorithm

Figure 23 shows the best fit solution for the U.S.A. elections data sets using a global search optimization algorithm. Figure 23 (a) shows the best fit plot of the data set while 23 (b) is the zoomed-in image of Figure 23 (a). The fixed parameter values are  $\beta = 1$  and  $K = 5$ . Initial conditions  $u(0) = 5, v(0) = 2$ .

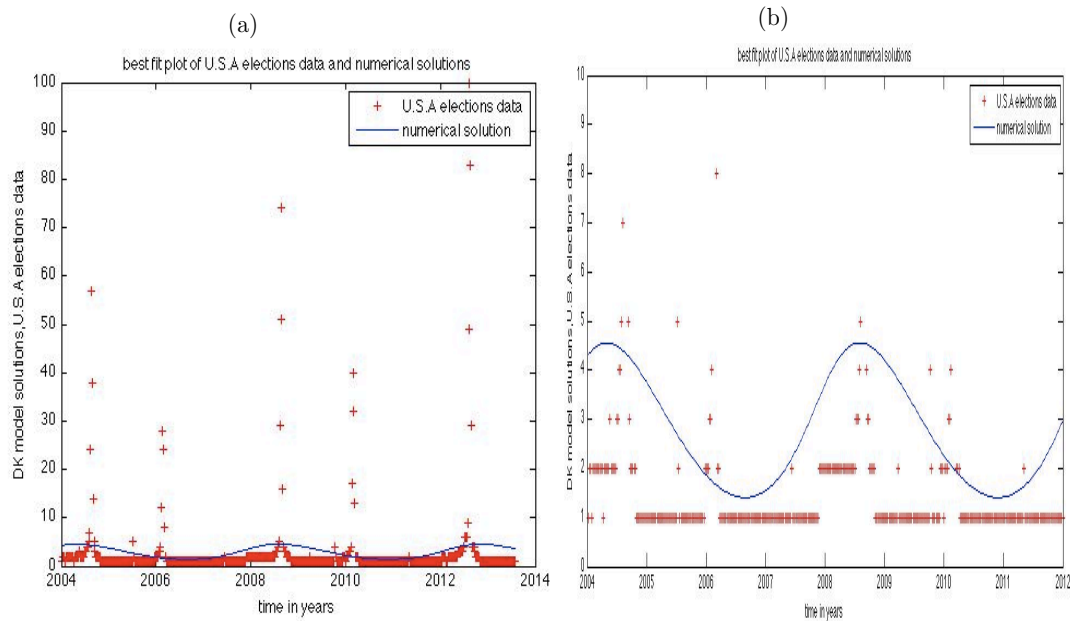


Figure 23: The best Fit Solution for the U.S.A. Elections Data Sets using a global search optimization algorithm

We should note that, in Figures 21–23 representing U.S.A elections data sets, the best Fit Solution appears different from those in Figures 18–20 because the U.S.A presidential elections data sets have a period of four years while U.S.A elections data sets have a period slightly different from four years.

The results of the estimates and corresponding standard errors from the different optimization algorithms for the U.S.A elections data sets are compared and summarized by Table 4 below.

Table 4: Parameter estimates and standard errors obtained by three optimization algorithms: Nelder-Mead<sup>1</sup>; pattern search<sup>2</sup>; global search<sup>3</sup>

Parameter	Estimate <sup>1</sup>	S.E <sup>1</sup>	Estimate <sup>2</sup>	S.E <sup>2</sup>	Estimate <sup>3</sup>	S.E <sup>3</sup>
a	3.541	31.474	3.874	3.525	3.811	4.128
b	4.071	49.730	2.596	13.792	4.267	2.793
$\alpha$	1.693	10.204	1.600	1.236	1.976	1.525

In Table 4 the Nelder-Mead algorithm (fminsearch) has the largest standard errors for the estimates of all the parameter values compared to a pattern search optimization algorithm. The global search optimization method resulted in the the least standard error for the estimate of the parameter  $\alpha$  compared to the Nelder-Mead algorithm (fminsearch) optimization algorithm. The best fit solutions tended to capture the lower levels of the data set values. We can also see that the estimates for  $\alpha$  had a larger uncertainty when we used a global search compared to pattern search. The Nelder-Mead (fminsearch) method only converges for convex problems, of which our model is not, and only guarantees decreasing values otherwise. Without convexity, Nelder-Mead (fminsearch) method tends to get in “far from correct” answers; gradient speedup is sometimes used, but requires smooth objective. Therefore, in general we would not expect reliable results for the estimates.

## 4 SENSITIVITY ANALYSIS

Sensitivity analysis is a systematic study of how sensitive numerical solutions of the model are to any slight changes in the parameter values [1]. The sensitivities of a system provide information on how states of the system respond to changes in parameters. We can use sensitivities to identify time intervals where the system is most sensitive to such changes [3]. Given that sensitivities are used to calculate uncertainties in parameter estimates, a look at the sensitivity functions is an indicator of time intervals at which data points have less or more information for the parameter estimation process [4].

Traditional sensitivity functions are used in mathematical modeling to investigate variations in the output of a model resulting from variations in the parameters and the initial conditions. Consider the following model parameters:  $a$  is the intrinsic growth rate,  $b$  is the halting rate,  $\beta$  is the contact rate, and  $K$  is the carrying capacity. If  $v \equiv v(a)$ , then the ratio  $\frac{\Delta v}{\Delta a}$  gives the average rate of change of spreaders  $v$  with respect to  $a$ . Consider  $v \equiv v(t, a)$ , so that

$$\frac{\partial v}{\partial a} = \lim_{h \rightarrow 0} \frac{v(t, a + h) - v(t, a)}{h} \equiv \frac{\partial v}{\partial a}(t, \theta).$$

Similarly for the parameter  $b$ , consider  $v \equiv v(t, b)$ , so that

$$\frac{\partial v}{\partial b} = \lim_{h \rightarrow 0} \frac{v(t, b + h) - v(t, b)}{h} \equiv \frac{\partial v}{\partial b}(t, \theta).$$

The curves  $\frac{\partial v}{\partial a}(t, \theta)$  and  $\frac{\partial v}{\partial b}(t, \theta)$  are called traditional sensitivity functions and when displayed versus time, they offer information on monotonicity of  $v$  with respect to the parameters  $\theta$ . The sensitivities are defined as  $\frac{\partial v}{\partial a}(t_j, \theta)$  and  $\frac{\partial v}{\partial b}(t_j, \theta)$ . Numerical

values of  $\frac{\partial v}{\partial a}(t_j, \theta)$  and  $\frac{\partial v}{\partial b}(t_j, \theta)$  are obtained by solving sensitivity equations (25)–(26)

$$\frac{dx}{dt} = f(t, x(t, \theta), \theta) \quad (25)$$

$$\frac{d}{dt} \frac{\partial x}{\partial \theta} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial \theta} \quad (26)$$

Note that  $\theta$  denotes the parameter values in the system of equations (13) and (14).

For the model defined by equations (9) and (10) we have that

$$\frac{df}{dx} = \begin{bmatrix} a - \frac{2au}{K} & -\beta v \\ \beta v & \beta u + \frac{b}{(v+1)^2} \end{bmatrix}$$

and

$$\frac{df}{d\theta} = \begin{bmatrix} u - \frac{u^2}{K} & 0 & -uv & \frac{au^2}{K^2} \\ 0 & \frac{v}{v+1} & uv & 0 \end{bmatrix}$$

So far we have considered the parameter estimation and seen the standard errors for all the four different data sets. We also used different optimization algorithms. It is therefore necessary to carry out sensitivity analysis. Sensitivity analysis will help us to know the level of uncertainty in our estimations associated with this model. In this thesis, let us consider the model given by the system of equations (9)–(10). We had the parameters  $\beta$  and  $K$  fixed. So we need to determine how small changes in  $a$  and  $b$  affects the behavior of the model. In Figures 24(a)–(d), each graph displays the traditional sensitivity functions of  $v$ . We can see that as time increases it seems that  $v$  is more sensitive to  $b$  than to  $a$  and its always positive. Also we note that the curve representing the parameter  $b$  has a larger amplitude compared to the curve representing the parameter  $a$ , implying that the function  $v$  is more sensitive to  $b$  than it s to  $a$ .

By looking at the sensitivity function curves, we see how the different parameters behave and this will provide information that is vital for the model. Figure 24 (a), 24 (b), 24 (c) and 24 (d) displays traditional sensitivity functions using the parameter estimates associated with the summer Olympics, the winter Olympics, the U.S.A. elections and the U.S.A. presidential elections data sets respectively. The green curves correspond to parameter  $a$  while the blue curves correspond to parameter  $b$ . The initial conditions of the model were  $u_0 = 5$  and  $v_0 = 2$ . The figures (a) is for the summer Olympics data set, (b) is for the winter Olympics data set, (c) is U.S.A. elections data set and (d) is for presidential elections data set. We should note the following in the behavior of the function  $v$  with respect to the parameter values of  $a$  and  $b$ :

- $\frac{\partial v}{\partial b} > 0$  implies that, between the years 2004 and 2008,  $v$  is an increasing function of  $b$ .
- $\frac{\partial v}{\partial b} < 0$  implies that, between the years 2008 and 2010,  $v$  is a decreasing function of  $b$ .
- $\frac{\partial v}{\partial a} > 0$  implies that, between the years 2012 and 2014,  $v$  is an increasing function of  $a$ .
- $\frac{\partial v}{\partial a} < 0$  implies that, between the years 2010 and 2012,  $v$  is a decreasing function of  $a$ .

Notice that extreme positivity and extreme negativity are observed in the same intervals listed above for the parameters  $b$  and  $a$  respectively.

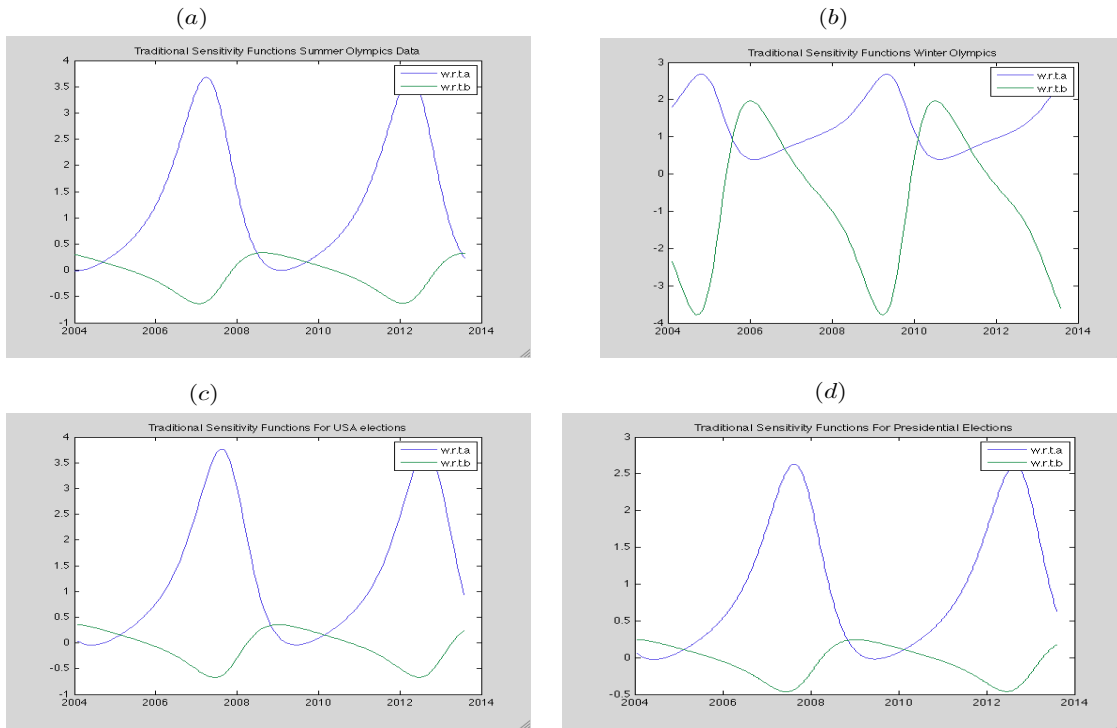


Figure 24: Sensitivity curves

From the sensitivity curves shown by Figures 24 (a)–(d), there is evidence of huge sensitivities every four years. This is mainly because it is after every four years that either the Olympics data sets or the Elections data sets discussed in this thesis are held. Therefore the “feast” behavior of the data is manifested every four years while the “famine” behavior of the data manifests itself in the other years when neither Olympics nor Elections take place.



## 5 CONCLUDING REMARKS

In this thesis we have discussed the rumor model developed by Daley-Kendall [4]. The results of this thesis have shown it is possible to move from a stable equilibrium to a stable limit cycle by introduction of a factor like the halting rate used in the DK model. In comparing the SIR [2] model and the DK model, it must be noted that we only assumed that a naive person requires only one spreader in order to become a spreader. In the future, one may consider a possibility of a naive person requiring more than one spreader in order to be converted into a spreader.

Another important result in this thesis is finding conditions for the existence of stable limit cycles. Computing the equilibrium points and setting conditions on the trace make it possible to construct the trapping region. We fixed some parameter values i.e.,  $\beta = 1$  and  $k = 5$  in order to determine the region at which the system undergoes a stable limit cycle. This together with the trace curve enabled us to determine the exact point within the region that is considered as a repeller. Finally, we proved that a stable limit cycle exists.

From the analysis of the extended Daley-Kendall model [4], we learned from this thesis that mathematical models can be used to mimic patterns observed in real data sets. However, sometimes we still end up with large standard errors. In our case we had to consider data sets with the same period as the numerical solutions of our model. The results showed different estimates and with different values of uncertainties for each method of optimization used. We should take note that our interest is to have least uncertainties (smaller standard errors) as possible since they are better than larger standard errors (more uncertainties). In the extension of Daley-Kendall [4].

We used the Ordinary Least Squares method [1] and three different optimization algorithms to fit the model to different datasets from google trends. The results showed varied levels of uncertainties (from the values of standard errors seen in Tables 1–4).

It is therefore necessary to use the best approach that will give the best fit and least uncertainties as possible. From the results of *OLS* estimations we observe that the global search resulted in the best estimates and least uncertainties while the Nelder-Mead search (Fminsearch) gave large uncertainties. In the rumor model we investigated, several assumptions were made on errors used in the statistical model for observation process. These assumptions, together with the input parameter values, may have an effect in the final results. This may have been due to violations of some of the statistical assumptions; for example the errors  $\varepsilon_i$  may not have had a mean of zero. Whatever method or approach that one uses, the final result from this model solution will determine if the model is a good representation of the system being investigated. It is therefore important to carry out model validation although this may not give the best results especially when the data is not real as in the Daley-Kendall [4] model where it is not easy to get a data set for rumor.

In general another important result we have noted is that sensitivity analysis by varying one or two or all parameters simultaneously is important for any model. Sensitivity analysis can provide an understanding of the Daley-Kendall [4] model such as how its output was affected by specific assumptions we made and parameters used. Note that in our model, the traditional sensitivity equations denoted by  $\frac{\partial v}{\partial a}(t)$  and  $\frac{\partial v}{\partial b}(t)$  curves shows how  $v$  behaves with respect to changes in the values of parameters

$a$  and  $b$ . This may help in understanding the contribution of the various sources of uncertainty to the model output, uncertainty and its performance in general. As in this our case, we refer to the Daley-Kendall [4] model results. This will provide information on whether the model can be reliable in solving real life problems. We had fixed the value of  $\beta$  and  $K$ , so we only had to analyze how values of  $a$  and  $b$  can change the appearance of the curves representing the behavior of the system. We have seen that not all parameters cause a significant change in behavior of the model 13–14. Note that our estimates may not be 100 percent reliable due to the violated assumptions and also method of estimation used (Ordinary Least Squares method).

## BIBLIOGRAPHY

- [1] H. T. Banks and H. T. Tran, “Mathematical and Experimental Modeling of Physical and Biological Process”, Taylor and Francis Group, New York, 2009.
- [2] F. Brauer and C. Castillo Chavez, “Mathematical Models in Population Biology and Epidemiology”, Springer, New York, 2010.
- [3] A.Capaldi, S.Behrend, B.Berman, J.Smith, J.Wright and A.Lloyd, *Parameter Estimation and Uncertainty Quantification for an Epidemic Model*, Math. Biosci. Eng., (2012); 9(3): 553-576.
- [4] D. J. Daley and D. G. Kendall, *Stochastic Rumours*, J.Inst.Maths Applics., **1** (1964), 42–55.
- [5] H. W. Hethcote, *The Mathematics of Infectious Diseases*, SIAM Rev., **42**, (2000), 599–653.
- [6] The Mathworks, Inc., Natick, MATLAB and Statistics Toolbox Release 2012b, Massachusetts, United States.
- [7] Hirsch M. W, Smale S and Devaney R. L, “Differential Equations, Dynamic Systems and an Introduction to Chaos,” Springer, 2nd edition, New York, 2004.
- [8] W. O. Kermack and A. G. McKendrick, *Contributions to The Mathematical Theory of Epidemics, part 1*, Proc. Roy. Soc. London Ser. A, 115 (1927), 700–721.
- [9] H. S. Strogatz “NonLinear Dynamics and Chaos”, Westview Press, New York, 2000.

VITA

GEOPHREY ODERO

Education: B.S. Mathematics and Chemistry,  
The University of Nairobi  
Nairobi, Kenya, 2004.

M.S. Mathematical Sciences,  
East Tennessee State University,  
Johnson City, Tennessee 2013

Professional Experience: Graduate Teaching Assistant,  
East Tennessee State University,  
Johnson City, Tennessee, 2011-2013