**Stochastic Dynamics of RNA interference:**

**A Single-Molecule Time-based Perspective**

A dissertation presented

by

Baihan Lin

to

Prof. Hong Qian,

Department of Applied Mathematics

in partial fulfillment of the requirements for

AMATH 531: Mathematical Theory of Cellular Dynamics

in the subject of

Applied Mathematics

University of Washington

Seattle, Washington, USA

Dec 2016

**Contents:**

1. **Abstract**  …………………………………………………………………..……………. 3
2. **Key words**  …………………………………………………………………..………..…. 3
3. **Introduction** ……………………………………………………………….…..………..…. 4
4. **Methods**  ……………………………………………………………….…..………..…. 5
   1. Assumptions
   2. Mechanistic Model
   3. Simplified & Realistic Model
   4. Mathematical Modeling
5. **Results** ……………………………………………………………….…..……..…. 12
   1. Variables
   2. Parameters
   3. Data Fitting
6. **Discussion**  ……………………………………………………………….…..………..…. 18
   1. Bifurcation Analysis
   2. Limitations
      1. Long-Range of Incubation Time
      2. Short Natural History
      3. Heterogeneity of Epidemiological Components
      4. Reliability of Data
   3. Future Study
      1. Time-Dependence
      2. Intervention Study
      3. Analogous Stochastic Model
      4. Other Factors
   4. Special
7. **References** ……………………………………………………………….…..………..…. 21
8. **Author Information** …………………………………………………….…..……..…. 22
9. **Acknowledgements**  ………………………………………………….…..……..…..…. 23
10. **Abstract:**
11. **Keywords:**
12. **Introduction**
13. **Methods:**
    1. **Assumptions**

In this paper, I make some assumptions listed below:

1. α (contamination rate per bat per day) and β (transmission rate per person per day) are time-independent and continuous over a time period.
2. The probability of their contacting the disease at time t is proportional to the product of I(t) and S(t).
3. No effects result from different levels of ages, genders, habits, lifestyles, vocations, and other diseases (excluding Ebola Virus Disease).
4. Those who have recovered or died from the disease are forever more immune.
5. Those who have not had the disease are equally susceptible.
6. Those who are infected have the equal incubation period.
7. Everyone in the population has the same probability of contacting with a bat.
8. Infected and non-infected bats are equally distributed in geography.
9. Neglect the outer factors that would have influence on the evaluations from model, such as weather, climate, wind, humidity and geography.
10. Fruit bats carrying the Ebola Virus will not affect their survival rate and reproduction rate comparing to normal fruit bats.
11. Fruit bats carrying the Ebola Virus produce offspring that are healthy.
12. Fruit bats grows exponentially without the constraint of carrying capacity.
13. Fruit bats become contagious by its own, not requiring contact of infected fruit bats (just like a gene mutation).

Clearly these assumptions are more pragmatic than realistic. Some assumptions, such as Assumption 2 and Assumption 5, may be realistic if the population under consideration consists of a group of about the same age and general health level, if there is no inherited immunity, and if the group members mix homogeneously.

For fruit bats, Assumption 12 and 13 are only for simplification purpose, instead of accurate modelling. In the later chapter, more realistic model with logistic growth is also discussed in more details.

* 1. **Mechanistic Model**

Built upon the classic SIR model, I included an exposed group (E), in which people are infected but not yet become infectious and symptomatic. I also decided to include transmission elements from fruit bats, which are the major carriers of Ebola Virus. They are described as the compensating terms of non-infectious (B0) and infectious bats (B1). I included the death term (X) because I thought the high fatality made it significant not to be neglected and the fundamental rate difference of recovered and dead cases made it necessary to be included. A potential importing of Ebola cases is also shown in our proposed model (Figure 1)

**S**

**B0**

**B1**

**E**

**R**

**X**

**I**

α

k

γ0(1-η)

γ1η

ωB1

v

**X0**

**X1**

**O**

**C**

**Figure 1**. Schematic Diagram of Comprehensive SOBBEIRXX Model

* 1. **Simplistic & Realistic Model**

Here due to time limitation and calculation convenience, I simplify the model by considering the importing Ebola cases and the difference between buried and non-buried death as insignificant. These simplifications offer a much more clarified model appropriate for this project and I made two more assumptions that:

1. The importing Ebola cases are relatively insignificant comparing to all other variables.
2. There is no significant difference between the effect onto other variables by buried and non-buried death cases.

These can be justified by considering the relatively low cross-country hospitalizations and modernization of West African rituals. However, these factors are not negligible in a more realistic demand because of the innate complexity of this network. In further study I can include these factors to better investigate the dynamics of the epidemic of Ebola Virus Disease.

**S**

**B0**

**B1**

**E**

**R**

**X**

**I**

α

v

k

γ0(1-η)

γ1η

ωB1

**C**

**Figure 2**. Schematic Diagram of Simplified SBBEIRX Model

That leads to a model consisting of susceptible cases (S), exposed cases (E), infectious cases (I), recovered cases (R), dead cases (X), non-infectious fruit bats (B0) and infectious fruit bats (B1), as shown in the Figure 2.

* 1. **Mathematical Modeling**

As shown in our model, the only way that a person can leave the susceptible group is to become exposed (infected but not yet infectious), either by getting in contact with fruit bats or with infectious humans; the only way that a person can leave the exposed group is to become infectious; the only way that a person can leave the infected group is to recover or die. Based on Assumption 4, once the person enters the recovered or dead group, he/she are forever immune and are no longer susceptible.

From these behaviors, I can come up with the following ODEs:

(1)

(2)

(3)

(4)

(5)

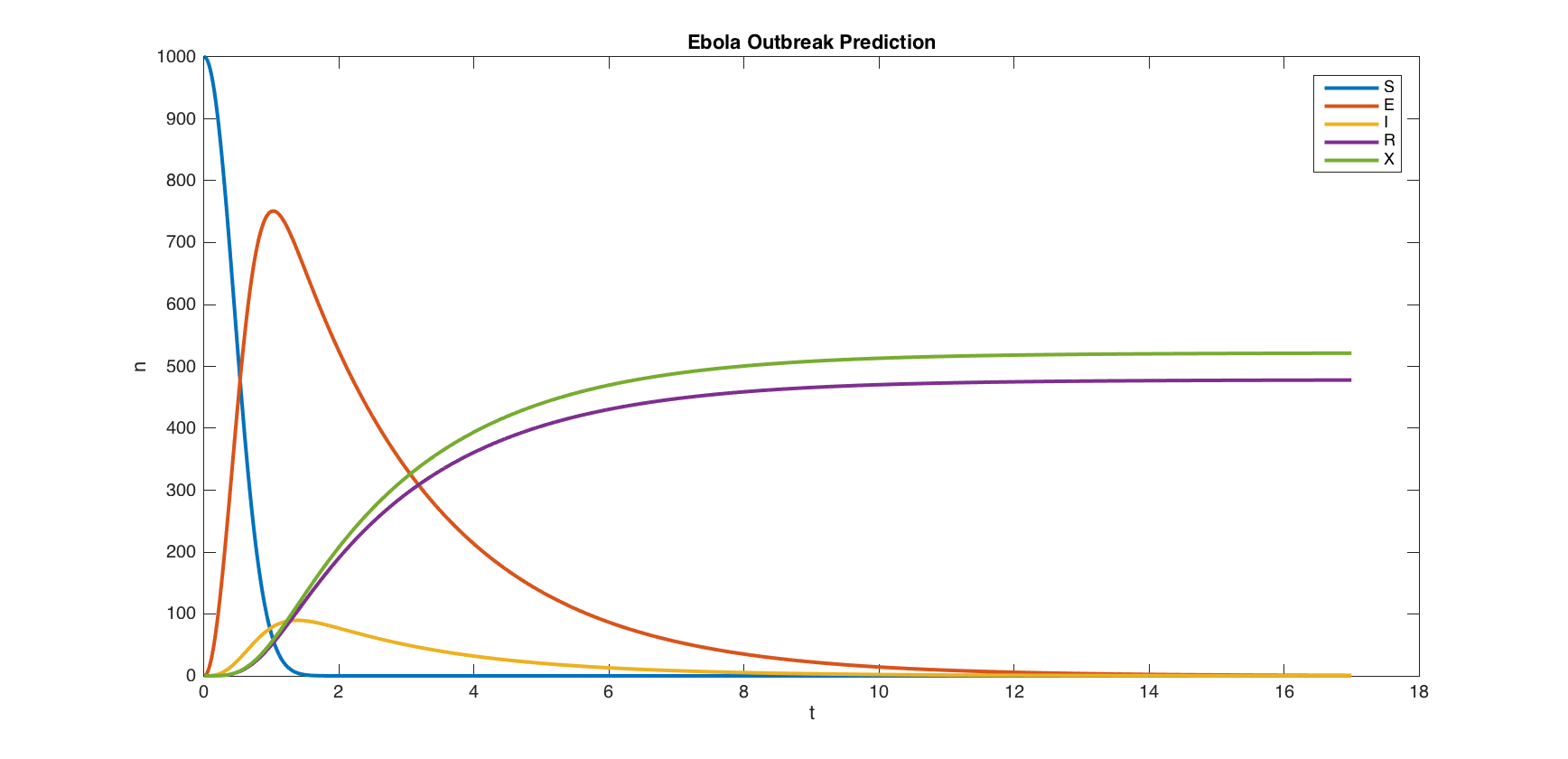
(6)

(7)

(8)

(9)

This ODE system is given as a start point in a simplistic way. As shown in Figure 3, it captures well an interesting dynamic with outcome that no human will be safe from infection after time. More will be discussed later in Discussion.

****

**Figure 3** A figurative outcome predicted by simplistic model.

However, it is clear to see that Assumption 12 and 13 are very unrealistic so a modified version of Equation (8) and (9) should be:

(10)

(11)

in which a logistic growth with carrying capacity is included within bat population, as well as an interaction-based infection pathway in Equation (11).

As can be seen from this system, I can just focus on Equation (1) to (5) because Equation (6) and (7) are just a sum of previous ones and Equations for fruit bats are pragmatically irrelevant.

To deal with this ODEs, I tried linear analysis to find equilibria, but it turns out to be non-linear so I have to use other approaches. In a probability base, applying the Markov Model used in Chemical Reaction, I rewrote the ODE system to be:

(12)

(13)

(14)

(15)

(16)

which can be further modified into:

(17)

(18)

(19)

(20)

(21)

So I decided to use Markov Chain but for a 7-variable system, then it appeared that bat variables should not be presented in the Markov Transition Matrix because they are not a closed system where probability add up to 1, as shown in Table 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **S** | **E** | **I** | **R** | **X** |
| **S** |  |  | 0 | 0 | 0 |
| **E** | 0 |  |  | 0 | 0 |
| **I** | 0 | 0 |  |  |  |
| **R** | 0 | 0 | 0 | 1 | 0 |
| **X** | 0 | 0 | 0 | 1 | 0 |

**Table 1** Markov Transition Matrix of SBBEIRX Model.

To obtain the steady states, I took the eigenvalues:

; ; ; ;

We are only interested in the eigenvector for eigenvalue , because that is the one to satisfy which indicates the unalterness of probabilities in human population after a long series of Monte Carlo random walks. Plugging in all the parameters, I got , which agrees my prediction that . However, in a realistic sense, an outcome of every individual in a population eventually reaching recovery state seems unlikely. Real data simulation in later chapters also supported my prediction. I think the reason the real simulation diverges from this steady state is the introduction of a changing parameters from bat-human population, which was neglected here in Markov Chain Transition Matrix. More will be discussed in Discussion.

1. **Results** 
   1. **Variables**

A detailed description of all the variables in the model is presented here:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Definition** | **Value** | **Source** |
| N | Effective population size | N(0) = 100 | Set |
| S | Number of susceptible people in the population | S(0) = 100 | Set |
| E | Number of infected (but not yet infectious) people in the population | E(0) = 0 | Set |
| I | Number of infectious people in the population | I(0) = 0 | Set |
| R | Number of recovered people in the population | R(0) = 0 | Set |
| X | Number of death in the population | X(0) = 0 | Set |
| X0 | Number of the deceased that is buried | N/A | N/A |
| X1 | Number of the deceased that is not buried | N/A | N/A |
| C | Cumulative number of cases after the onset of symptoms | C(0) = 0 | Set |
| B0 | Number of infectious bats | B0(0) = 99 | Set |
| B1 | Number of non-infectious bats | B1(0) = 1 | Set |
| O | Number of imported Ebola cases | N/A | N/A |

**Table 2** variables, definitions, initial values and their sources

* 1. **Parameters**

For the parameters, as shown in Table 3 below, most of them I provided with actual literature data with sources.

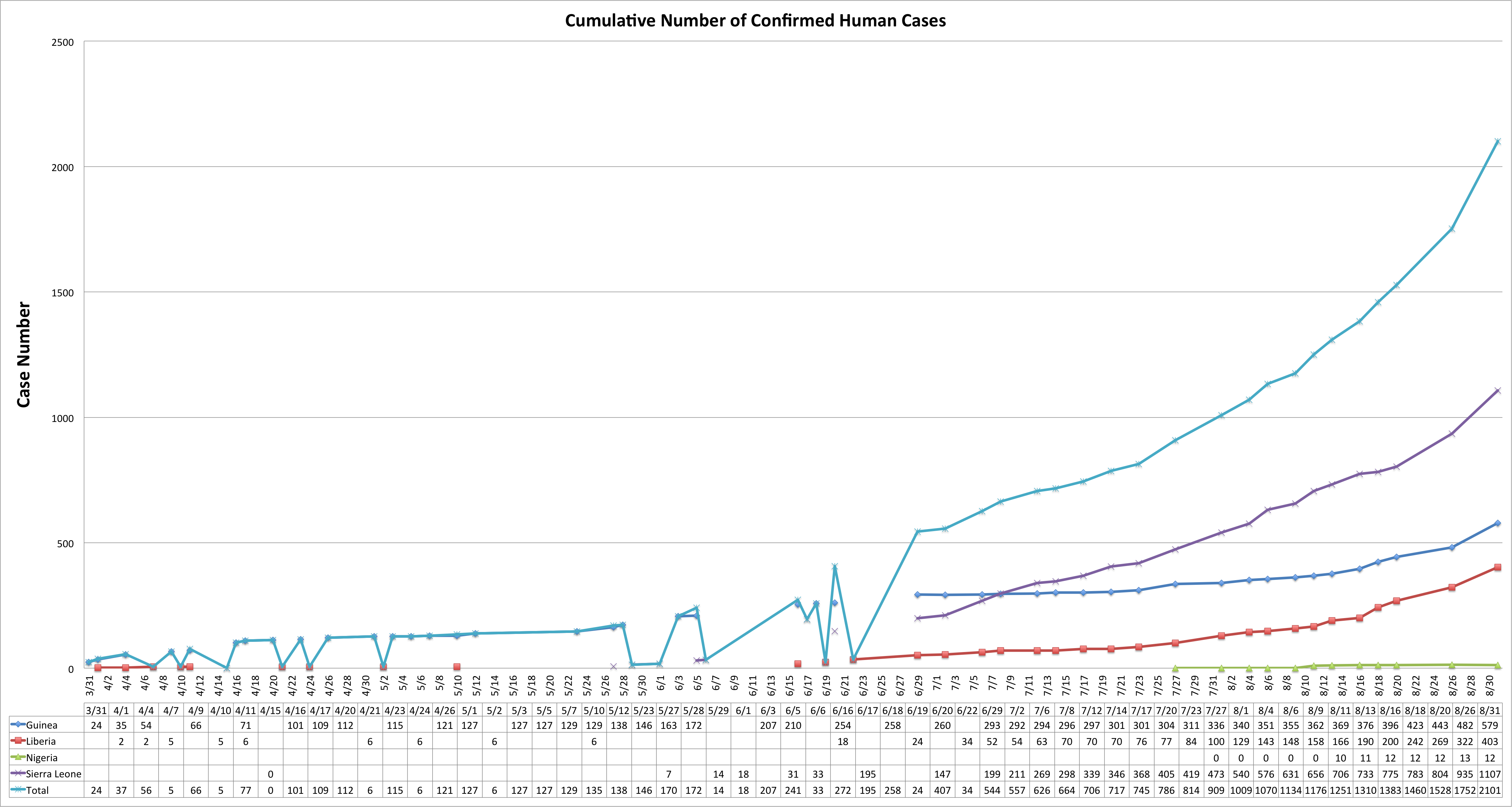
For only three blue-marked parameters, due to limited or distinct sources of data, I decided that, in order to determine the best values for the three blue-marked parameters which no set reliable literatures as guides, I should use Ensemble Kalman Filter, a recursive Monte Carlo-based Bayesian algorithm to search. Another worthwhile possibility to attempt is to apply Asymptotic Variance-Covariance, but I have not been successfully calculated it out yet.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Definition** | **Value** | **Source** |
| α | Time-dependent Contamination Rate per bat per day | 0.5 | By Ensemble Kalman Filter |
| β | Human-Human Transmission Rate per person per day | 0.4 | By Ensemble Kalman Filter |
| ω | Bat-Human Transmission Rate per person per bat per day | 0.1 | By Ensemble Kalman Filter |
|  | Infection Recovery Rate | 4 /month | is the mean infectious period (average time from symptom onset to recovery) (Shaman et al., 2014) |
|  | Infection Death Rate | 3 /month | is the mean fatal period (average time from symptom onset to death) (Shaman et al., 2014) |
|  | Exposure Infection Rate | 4 /month | is the mean incubation period (average time before an exposed person becomes infectious) (Shaman et al., 2014) |
|  | Case Fatality Rate | 0.5 | (WHO, 2016) |
|  | Growth Rate of carrier bat | 1/300 months | is the lifespan of lesser short-nosed fruit bat (Crichton, 2000) |
|  | Carrying Capacity of Bat Population | 108 | K for fruit bat in Ghana, West Africa (Kamins et al., 2011) |

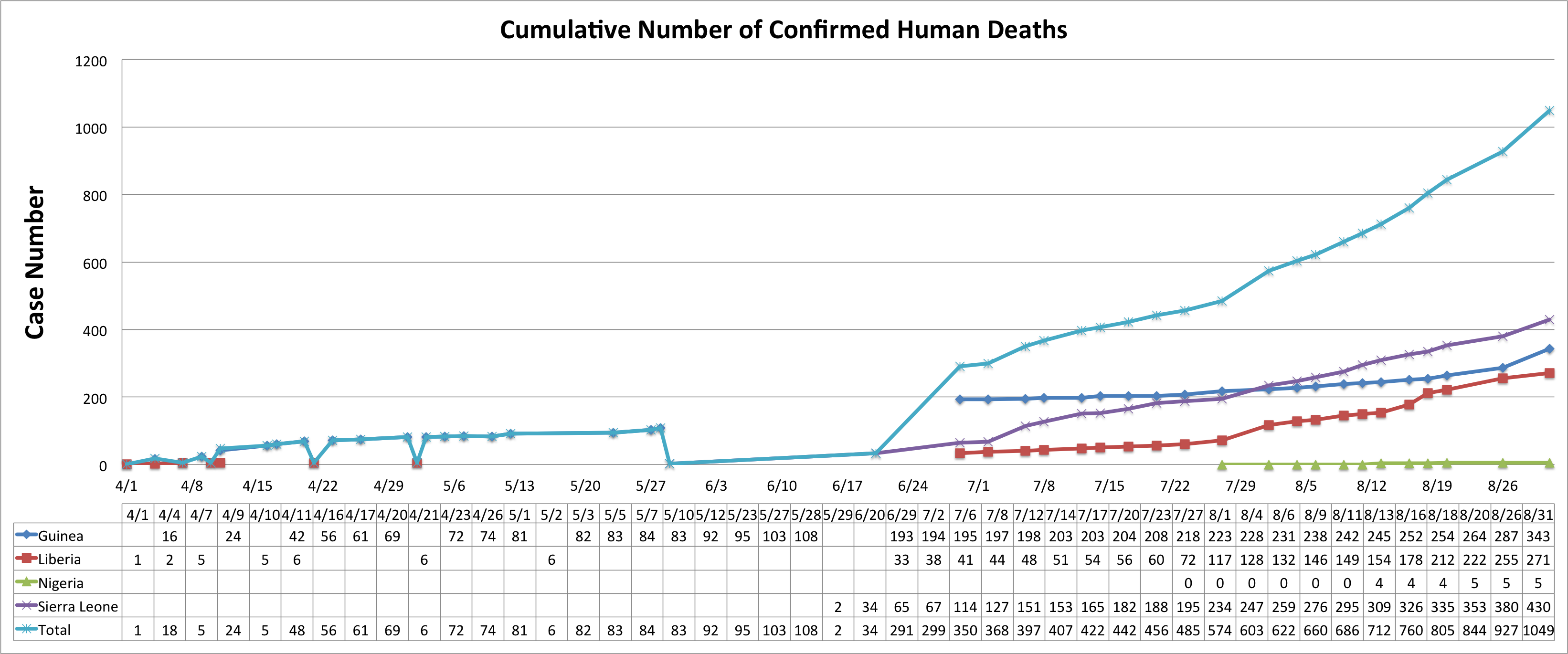
**Table 3** parameters, definitions, values and their source

* 1. **Data Fitting**

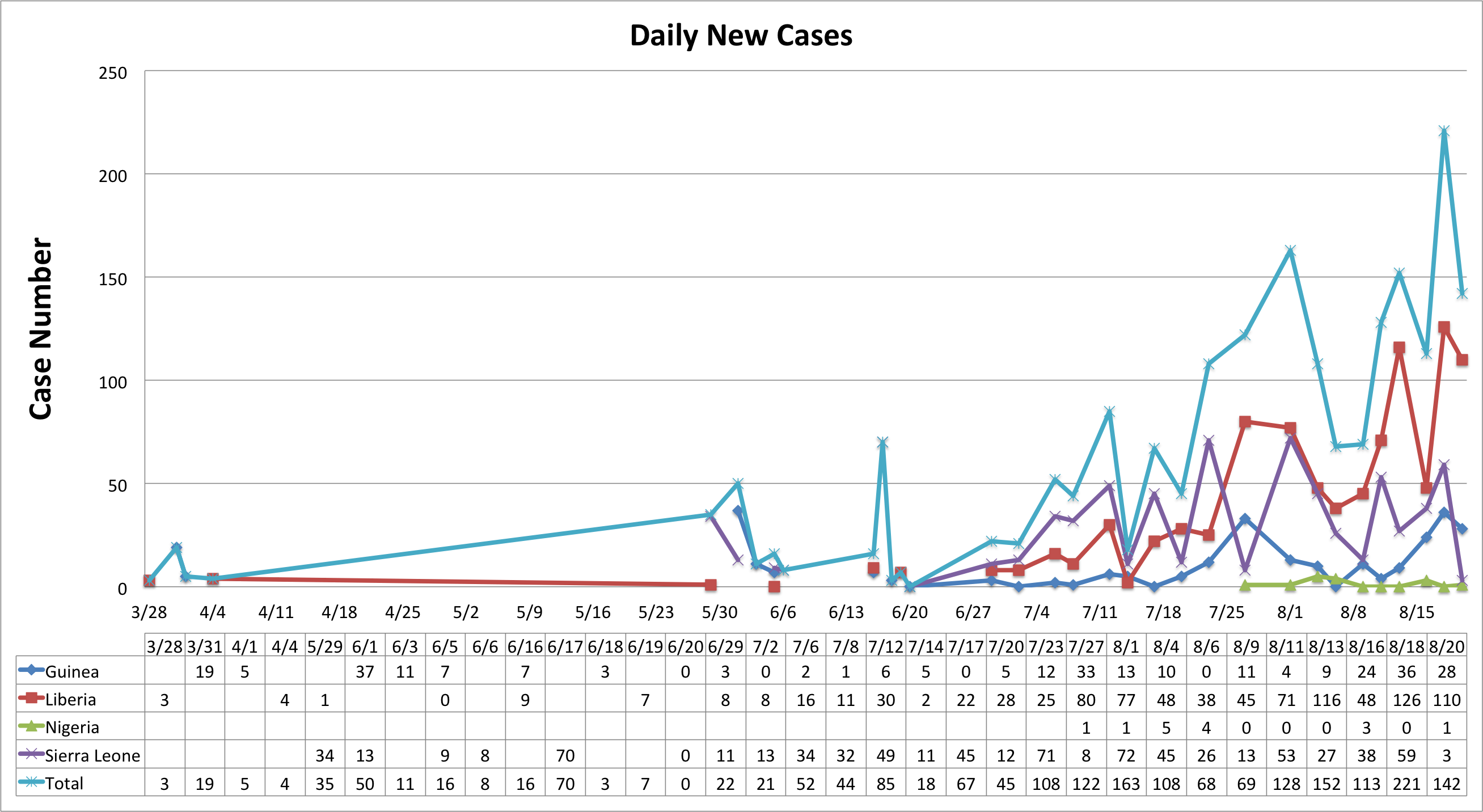
**A**

****

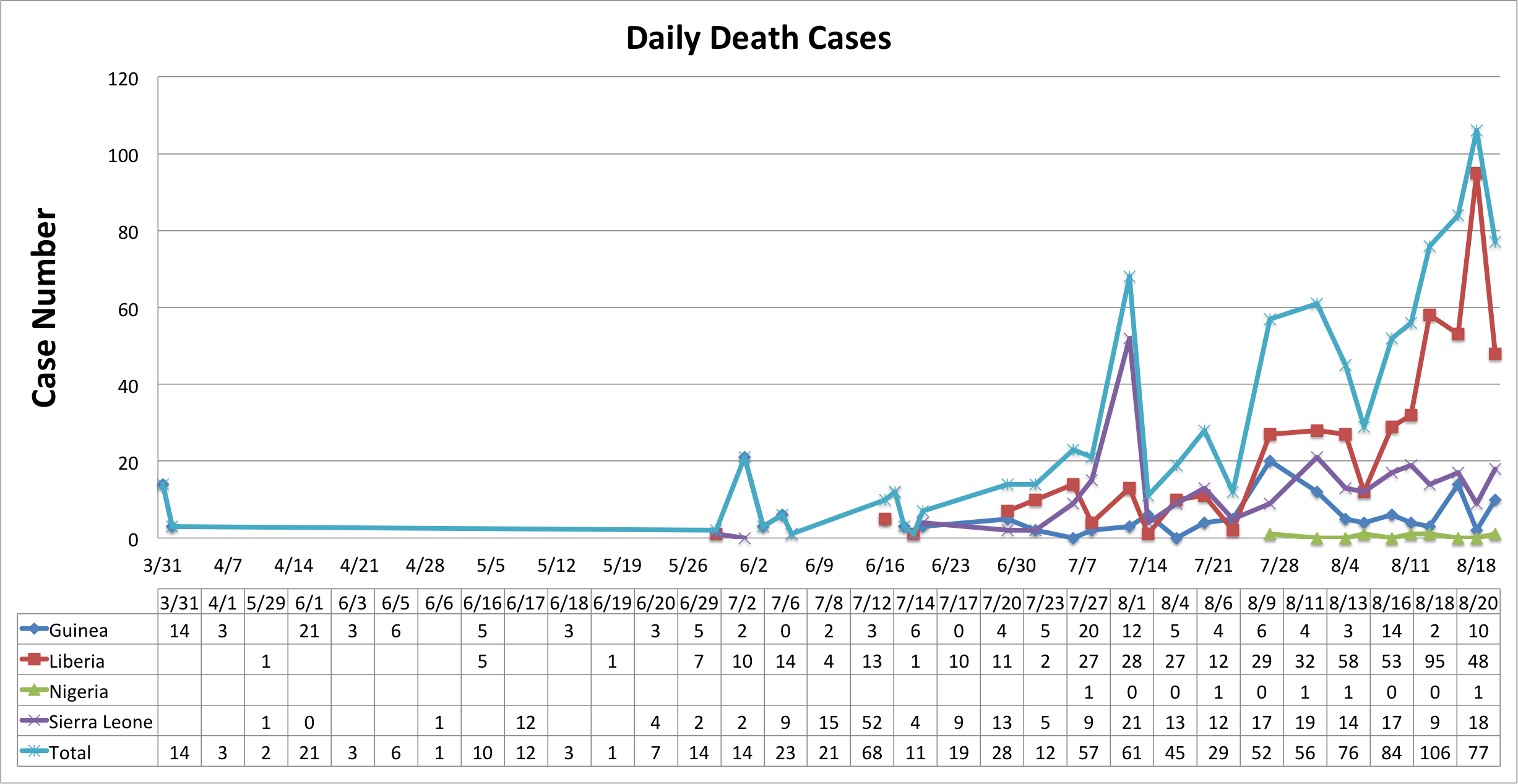
**B**

****

**C**

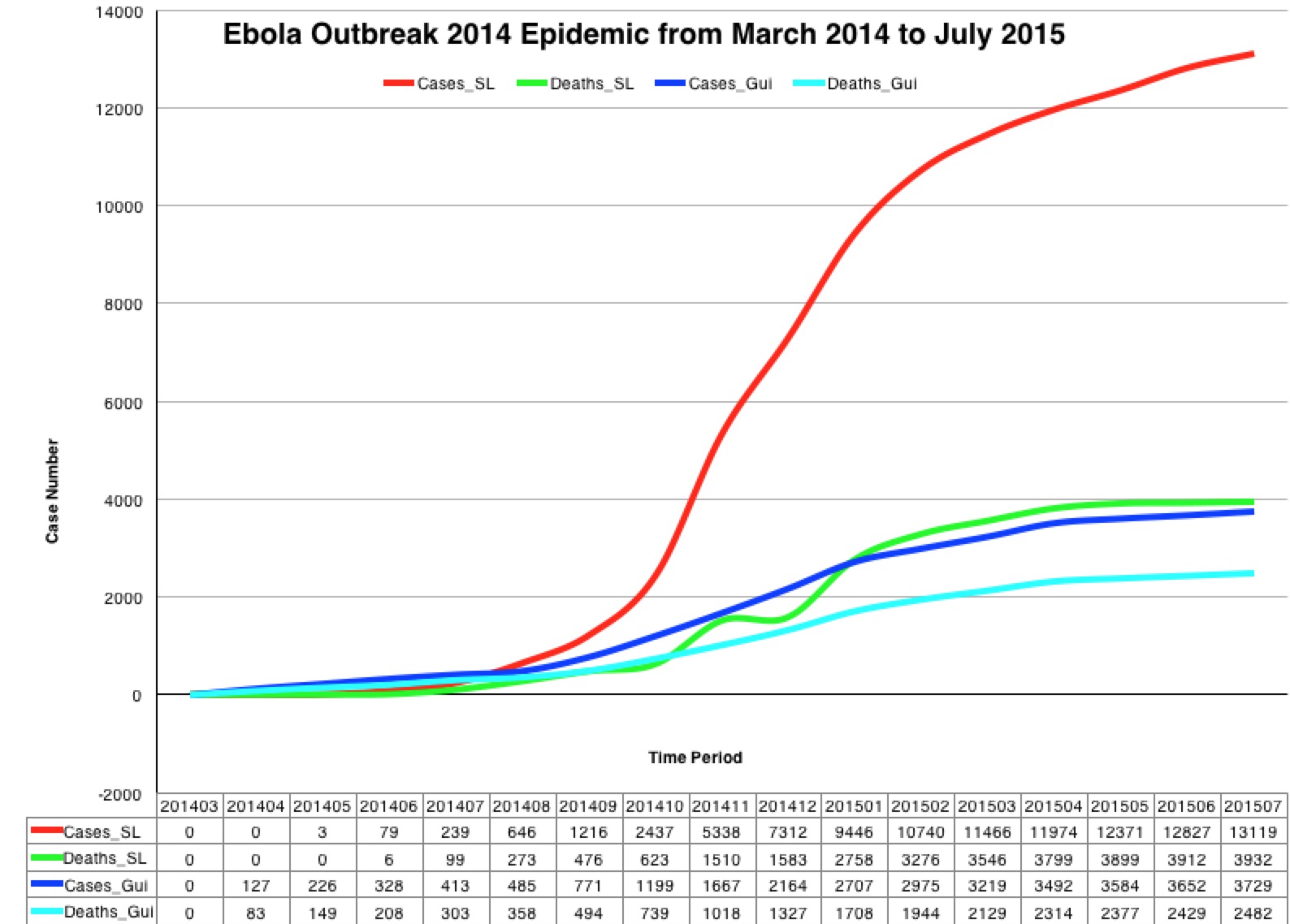
****

**D**

****

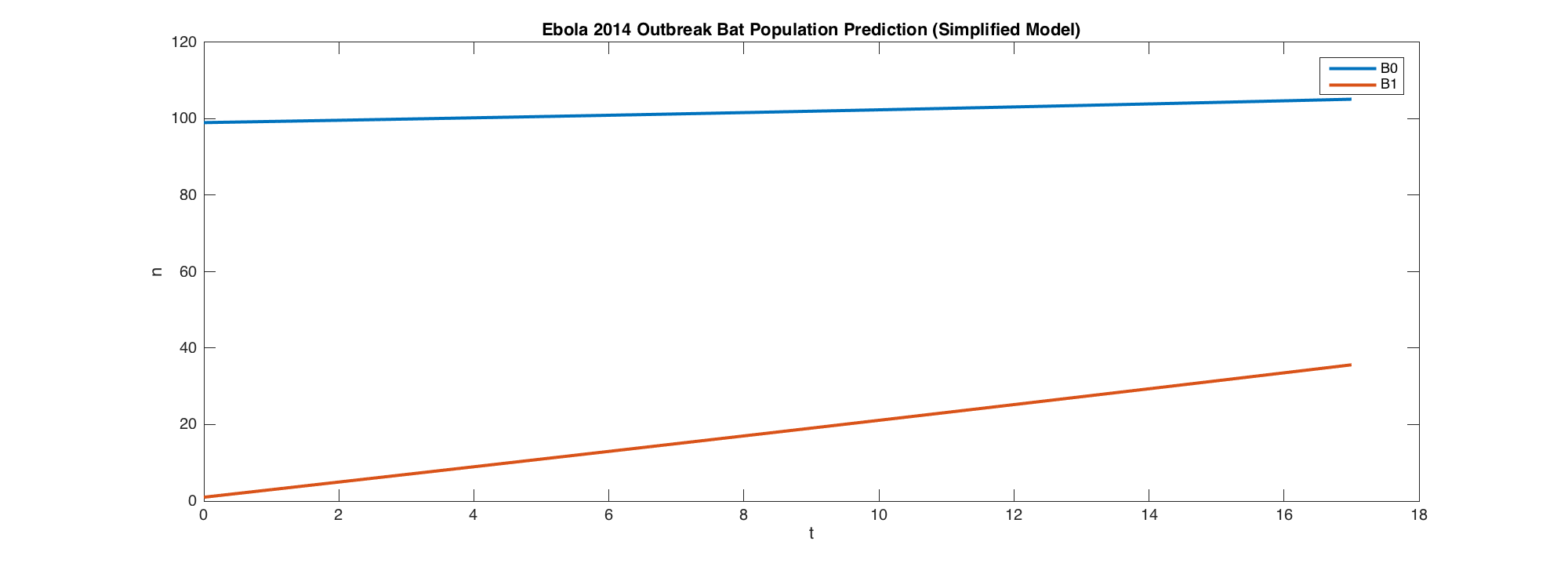
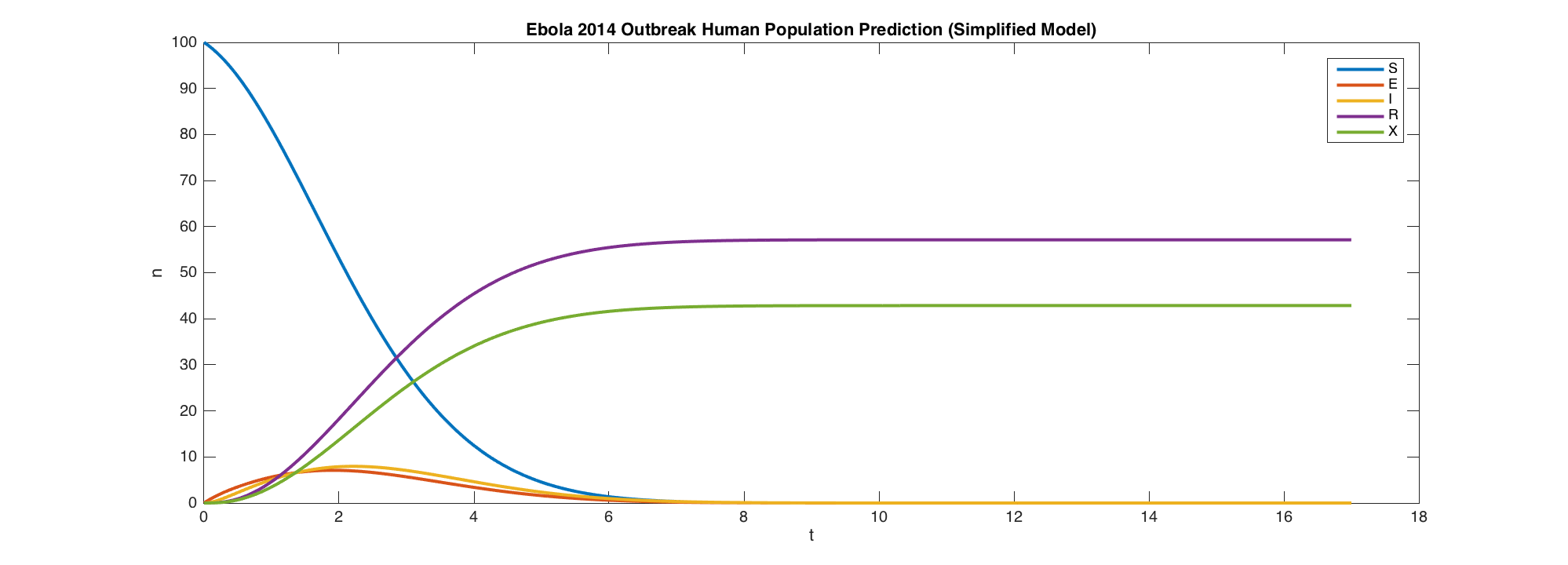
**Figure 4** (A) The cumulative number of confirmed human cases (B) The cumulative number of confirmed human deaths (C) The number of newly confirmed human infection cases (D) The number of newly human deaths. All among Guinea, Liberia, Nigeria and Sierra Leone from Mar 2014 to Aug 2014.

A major goal of this study is to fit the model to current reported data for Sierra Leone, Liberia, and Guinea. I noted that the data available are the cumulative clinical reported cases as shown in Figure 4, that is, the suspected cases, probable cases and confirmed cases according to the definitions given in WHO. This creates extra difficulty for me to identify which fit to best represent the current outbreak. Thus, I recollected the data to be only focusing on outbreak Sierra Leone and Guinea from Mar 2014 to July 2015, by month, for 17 months, as shown in Figure 5.



**Figure 5** the cumulative cases and death in Sierra Leone and Guinea from Mar 2014 to July 2015.

For the simplistic model, as shown in Figure 6A, throughout 17 months, all the infected, susceptible and exposed decreased to zero despite initial rises. This can be explained by the supplemental Figure 6B, where the infected bats are slowly climbing to reach better transmission.



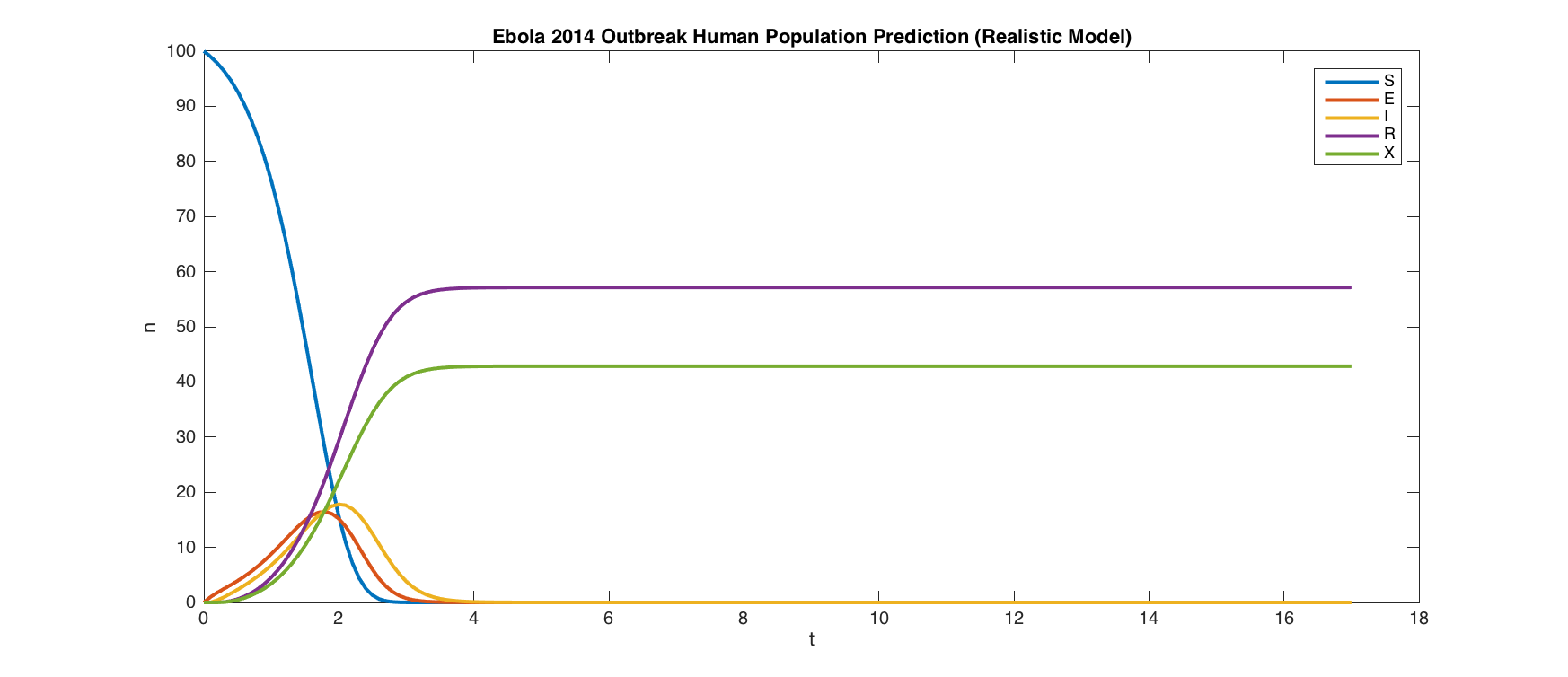
**B**

**A**

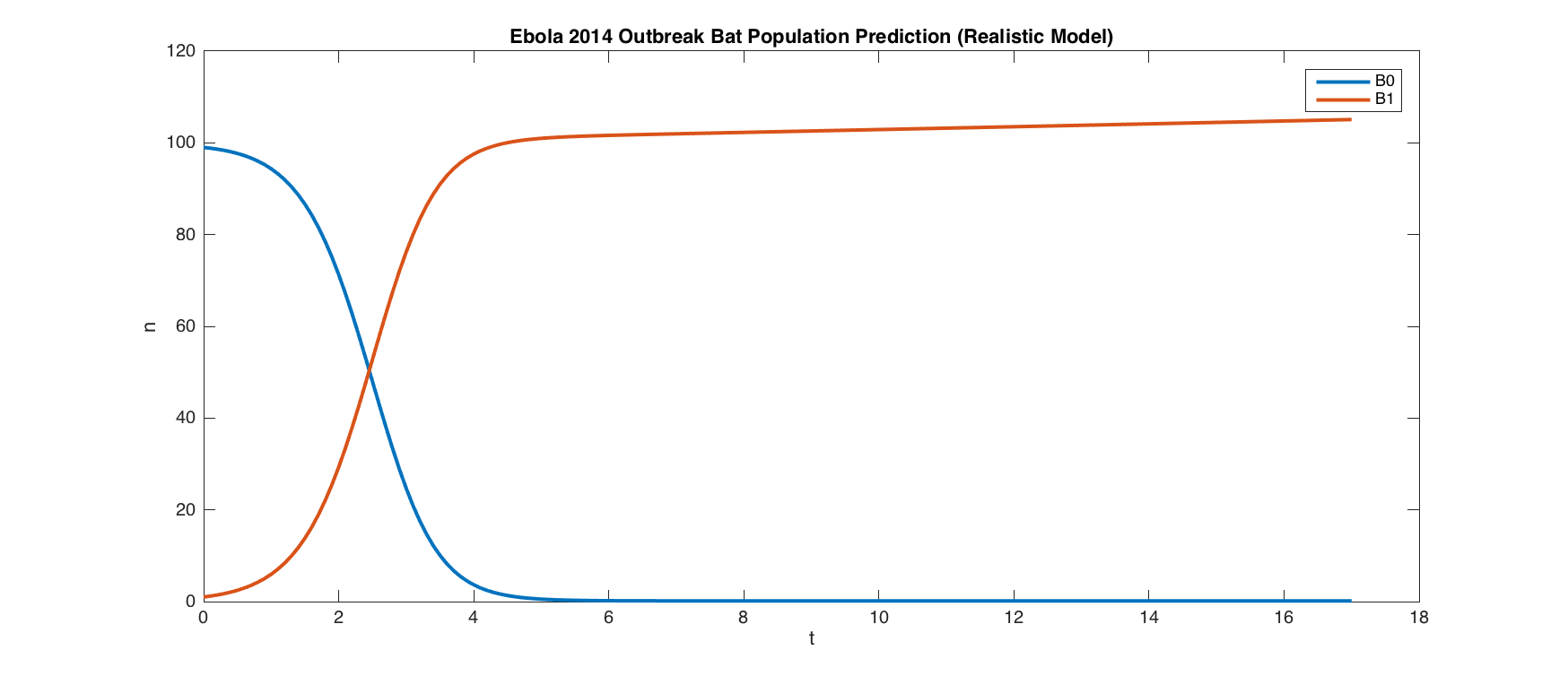
**Figure 6** Epidemic prediction of Ebola Outbreak 2014 by simplistic model in a 100 percentage scale (A) human population (B) fruit bat population.

For the realistic model, as shown in Figure 7B, the bat population grows logistically with a S-shaped curve. As shown in Figure 7A, after Ebola outbreaks, as time goes on, the number susceptible people (S) will dramatically decrease within the first month and eventually almost no susceptible and infected people exist. In another word, every one in the population will have been infected after a long period of time, with only immune recovered as well as dead people left. That distinguishes our model from most of other models proposed currently.

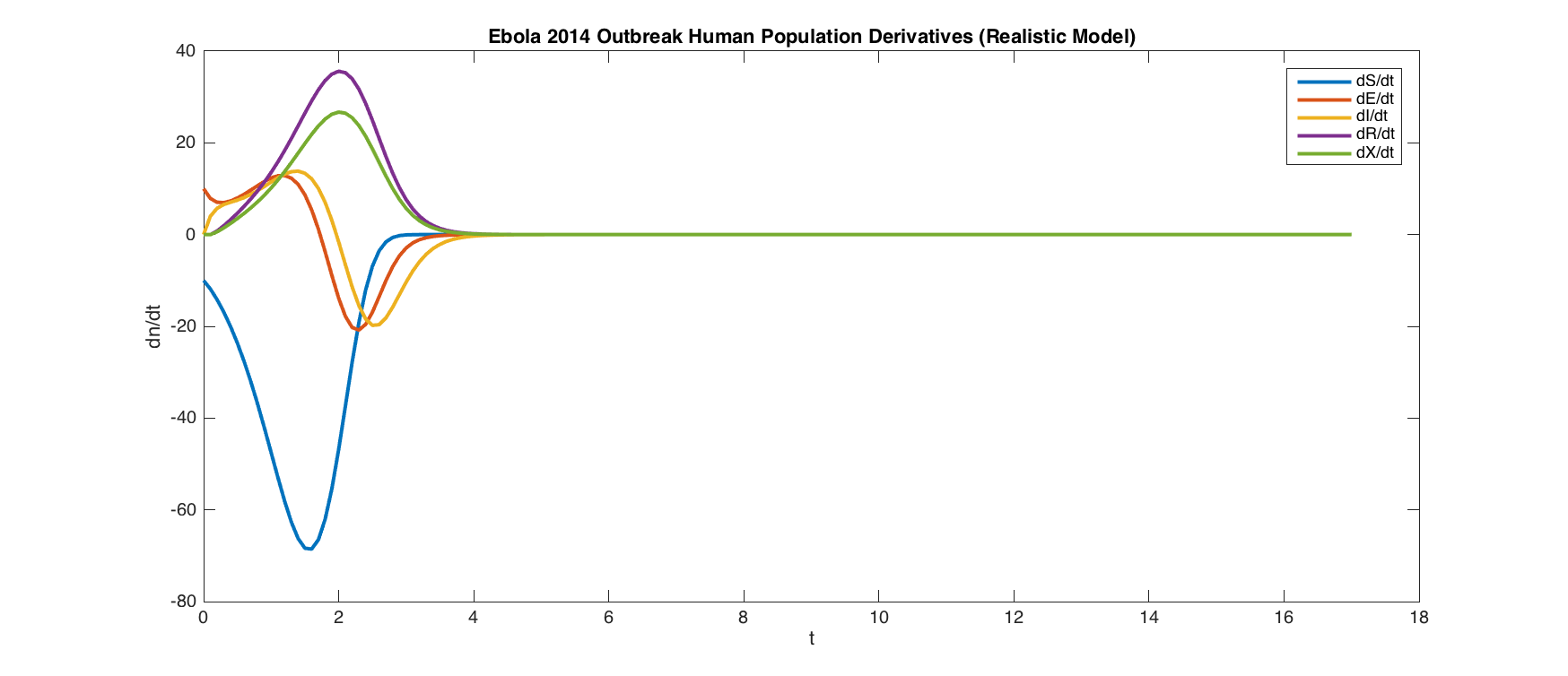
However, the number of infected (not yet infectious) people (E) will shoot up within the first month because people have not enough protective measures when the epidemic spreads out at very beginning. Then after operation of protective measures, gradually the number of infected (not yet infectious) people will decrease. Number of deaths (X) and recovered (R) from the infection will increase at first 4 months and then maintain stable as the disease has been controlled. Due to the immediate acts of protection, the number of infected people (I) will not have a sharply increase within the first month and will decay to almost zero as time goes on, just as I predicted.

****

**A**

****

**B**

****

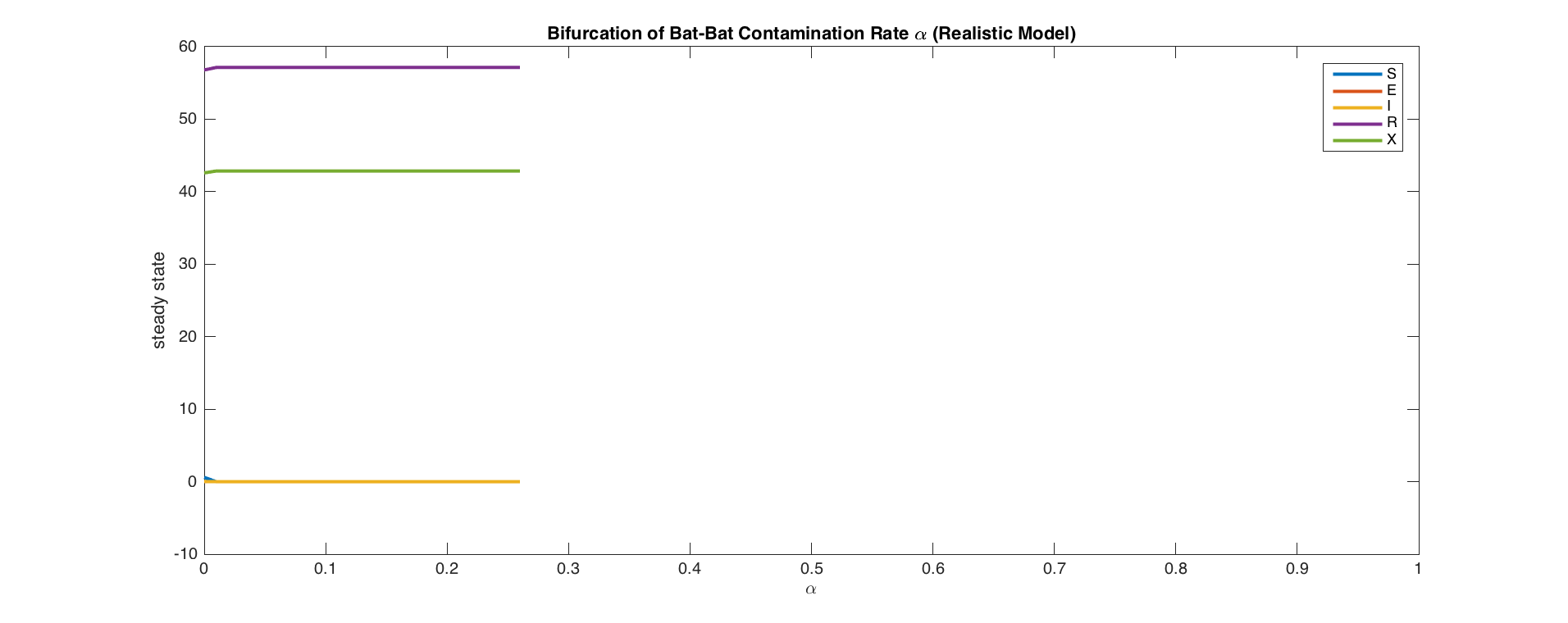
**C**

**Figure 7** Epidemic prediction of Ebola Outbreak 2014 by realistic model in percentage (A) human population (B) fruit bat population (C) differential function.

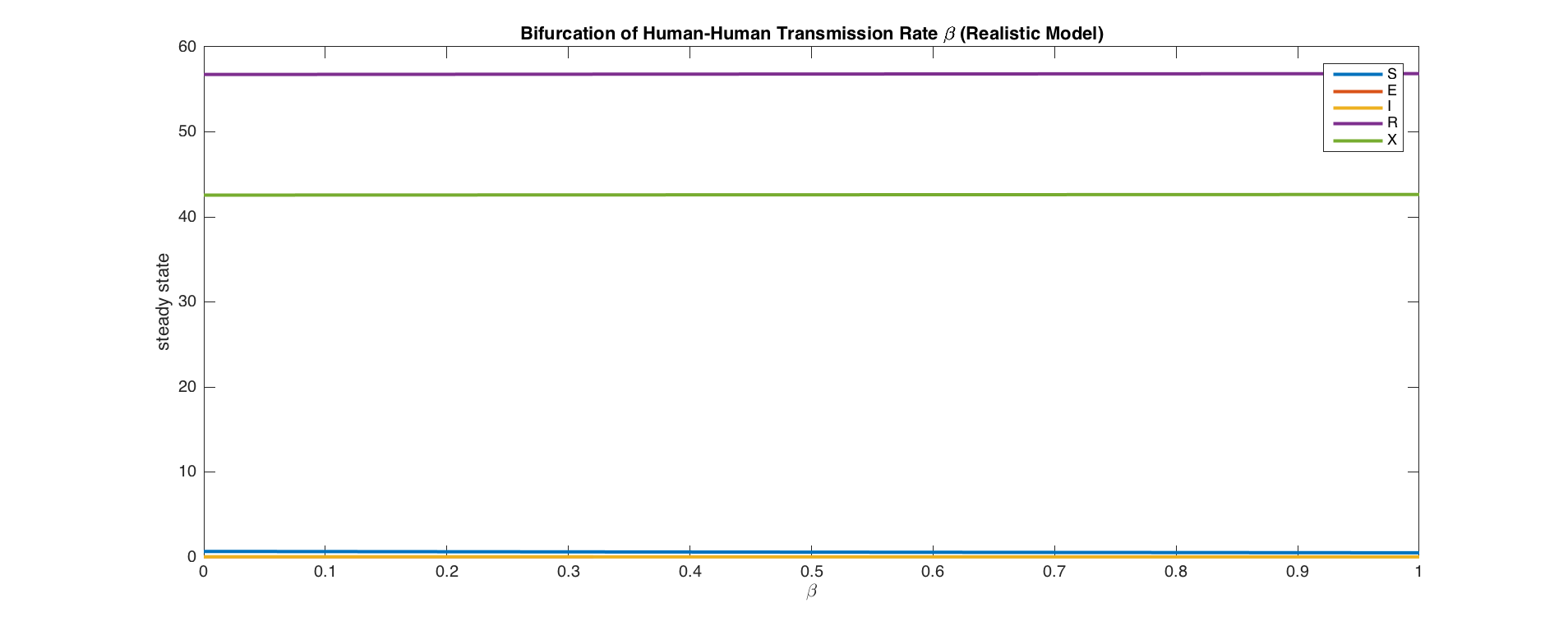
1. **Discussion** 
   1. **Bifurcation Analysis**

As discussed previously in our realistic model, Time-dependent Contamination Rate (α), Human-Human Transmission Rate (β), and Bat-Human Transmission Rate (ω) can be subjective to different conditions. I am interested in how changing these crucial parameters change the stability of my models.

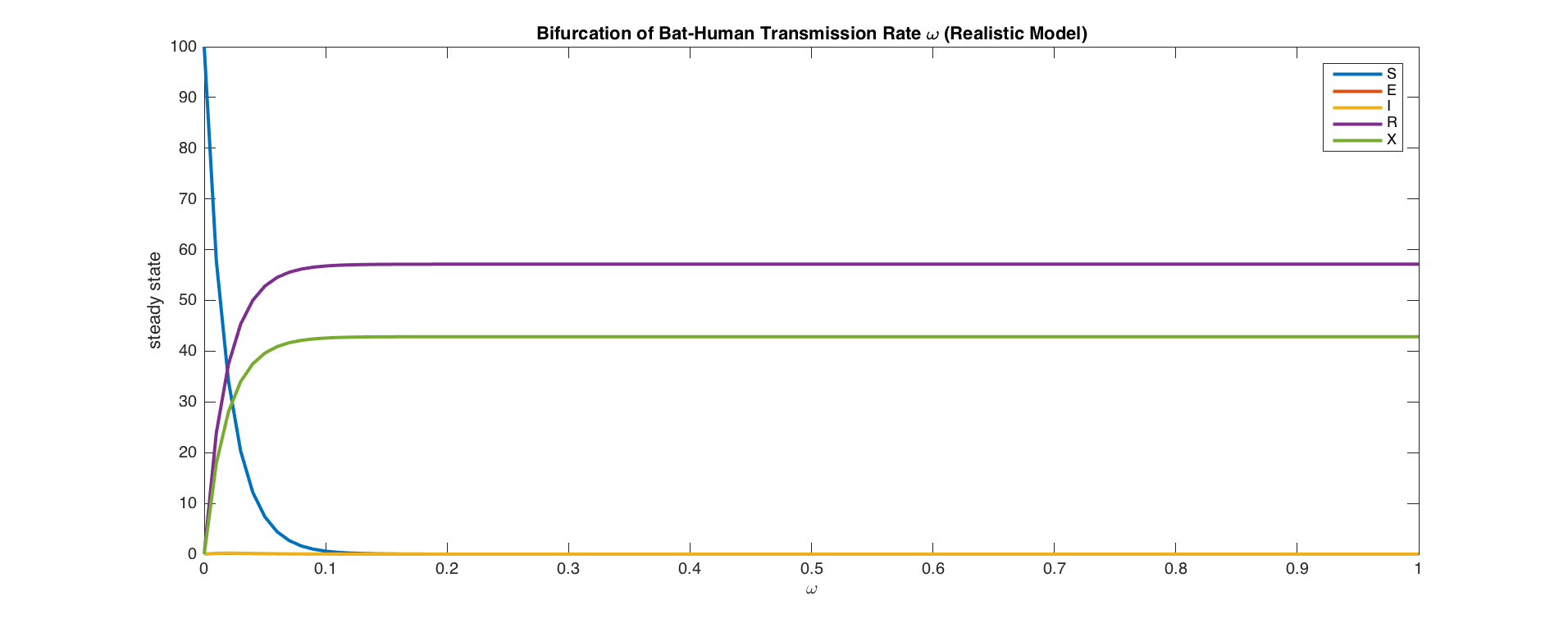
**A**



**B**



**C**

****

**Figure 8** Bifurcation analysis of (A) Contamination Rate per bat per day (α), (B) Human-Human Transmission Rate per person per day (β), and (C) Bat-Human Transmission Rate per person per bat per day (ω)

**Figure 8** bifurcation of (A) Contamination Rate (α), (B) Human-Human Transmission Rate (β), and (C) Bat-Human Transmission Rate (ω)

As shown in Figure 8, the α have a critical transition around ~0.24, while both β and ω remains stable after 0.1, thus implying the robustness of my models resilient to the fluctuation of β and ω.

* 1. **Limitations**

Given the likely unknown changing biases in the data, unknown observational error, and limited number and type of observations, the model is not as well constrained as it could be.

* + 1. **Long-Range of Incubation Time**

The uncontrolled spreading occurs from the time the person affected starts to feel sick and manifests enough symptoms to be disabled. This can not only cause miscount on the “real” numbers for calculation, but also offer challenges for health workers who have to do contact tracing.

* + 1. **Short Natural History**

Ebola has a high fatality rate, so patients not likely to be contagious for very long. Instead, they will quickly end up immune or dead. Thus the interaction is closely related to direct contact rate.

* + 1. **Heterogeneity of Epidemiological Components**

Assumption 3 is relatively unrealistic considering the direct contact rate difference in Western versus third world, the different cultural and social contexts, as well as the different governmental policies. For example, the burial ritual in some West African countries is thought to be significant factor in Ebola transmission.

* + 1. **Reliability of Data**

The timely report of multiple countries is hard. Thus, small fluctuations that may simply reflect an increase or decrease in surveillance or a reappraisal of older data. This cautious attitude toward lower numbers particularly applies to a reported drop in new cases in Liberia in August 2014, which is unlikely to be genuine and more likely reflects a deterioration in the ability of overwhelmed responders to record accurate epidemiological data.

* 1. **Future Study**
     1. **Time-Dependence**

Here, I assumed α and β as a constant. However, it can be time-dependent. So to determine α(t) and β(t), I can do backward calculation based on data.

* + 1. **Intervention Study**

Here, I assumed β(t) as continuous. However, with intervention, whether is it possible for us to assume a radioactive decay is an exciting attempt. (e.g. β(t) = β0 when t < τ, β(t) = β1 + (β0- β1) e-q(t-τ) when t >= τ.

* + 1. **Analogous Stochastic Model**

Here I use Ordinary Differential Equation and Markov Chain. But it seems I can also use Bayesian Difference Equation (with 7 variables). I may also find basic reproduction number more easily that way.

* + 1. **Other Factors**

If I reject the Assumption 3, for example, I can propose health care workers are different from other subpopulation. And different age groups can also offer interesting datasets to explore.

* 1. **Special**

What distinguishes our model from other models in Ebola Virus Disease outbreaks is that I included this external force from outside (in our case, the fruit bat infection). In most proposed models, the system is closed with self-regulation. However, in our system, the introduction of the fruit bats actually predicts an eventual infection of the entire population, which is very interesting and worth studying.

1. **References**
2. Crichton, E., P. Krutzsch. *Reproductive Biology of Bats*. San Diego, CA: Academic Press (2000).
3. Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W. & Hyman, J. M. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *J Theor Biol* 229, 119-126, doi:10.1016/j.jtbi. 2004.03.006 (2004).
4. J. H. Kuhn, L. E. Dodd, V. Wahl-Jensen, S. R. Radoshitzky, S. Bavari, P. B. Jahrling, Evaluation of perceived threat differences posed by filovirus variants. *Biosecur. Bioterror*. 9, 361–371 (2011).
5. Kamins, A. O. et al. Uncovering the fruit bat bushmeat commodity chain and the true extent of fruit bat hunting in Ghana, West Africa. *Biol Conserv* 144, 3000-3008, doi:10.1016/j.biocon.2011.09.003 (2011).
6. Fisman, D. N., Hauck, T. S., Tuite, A. R. & Greer, A. L. An IDEA for short term outbreak projection: nearcasting using the basic reproduction number. *PLoS One* 8, e83622, doi:10.1371/journal.pone.0083622 (2013).
7. WHO, <http://www.who.int/csr/disease/ebola/situation-reports/en/> (2014).
8. Chowell, G. & Nishiura, H. Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC Med* 12, 196, doi:10.1186/s12916-014-0196-0 (2014).
9. Lewnard, J. A. et al. Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis. *Lancet Infect Dis* 14, 1189-1195, doi:10.1016/S1473-3099(14)70995-8 (2014).
10. Nishiura, H. & Chowell, G. Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014. *Euro Surveill* 19 (2014).
11. Shaman, J., Yang, W. & Kandula, S. Inference and forecast of the current west african ebola outbreak in Guinea, sierra leone and liberia. *PLoS Curr* 6, doi:10.1371/currents.outbreaks.3408774290b1a0f2dd7cae877c8b8ff6 (2014).
12. Chowell, G. & Nishiura, H. Characterizing the transmission dynamics and control of ebola virus disease. *PLoS Biol* 13, e1002057, doi:10.1371/journal.pbio.1002057 (2015).
13. WHO, <http://www.who.int/mediacentre/factsheets/fs103/en/> (2016).
14. **Author Information**

For supporting MATLAB Codes, please refer to Supplementary Information Attached. All code for the reproduction of the reported results can be downloaded from my GitHub Page: <https://github.com/doerlbh/Epidemic>.

1. **Acknowledgements**

Thank Prof. Qian for giving us insightful lectures about the dynamic field of mathematical biology and setting challenging questions for me to explore!

Thank Dr. Felix Ye for giving me encouragement and help, and supporting such a great class full of excitement!

Thank my friends in AMATH 4/523 who come along with me in great energy!

Thank University of Washington for giving me the platform to scientifically explore problems and subjects!

I will continue the voyage of exploring the infinite realm of mathematical biology in my academic career fearlessly.

Baihan Lin

March 12 2016