Modeling Ebola in Sierra Leone with Stochastic Processes

#40661

February 9, 2015

Contents

1 Introduction

Our paper presents a modeling strategy for the 2014-2015 Ebola outbreak in Western Africa. Specifically, **we model Sierra Leone from the start of its outbreak in May of 2014**. A major factor in any population model is the growth factor. We discovered that the Center for Disease Control [1] used an SIIR model that overestimated the current Ebola growth rate, thus failing to properly model the epidemic. Our goal is to *accurately* model the epidemic, then present a solution for eradicating Ebola.

1.1 Modeling Approach

Prior to the year 2014, there have been Ebola outbreaks, but none near the magnitude of the 2014-2015 epidemic. Due to this non-equivalent history, not much is known about the large-scale spread of Ebola. We make a few simplifying assumptions to fill voids in current knowledge.

1.1.1 Simplifying Assumptions

- **There is a relatively small number of innately immune persons.** Initially,we are allowed to consider the entire population of Sierra Leone as susceptible to infection.
- **The mechanisms for disease spread are known.** The disease is spread primarily through human contact–specifically bodily fluids and blood; we do not model the spread through other animals, nor do we consider the possibility of spread through air, water, food, etc.

We model the Sierra Leone outbreak up until its current state, January of 2015.

1.1.2 Modeling the Current Outbreak

- **Model a region** in Sierra Leone using a typical SEIR model. We model a single district while not allowing the population to travel; we do not model the entire country.
- **Add an agent-based component** by modifying the SEIR model to use agents whose properties are stochastic.

- **Determine the stochastic properties** of the agents. The infectivity and incubation times of agents will be random variables, so their probability distributions must be appropriately chosen.
- **Model the entire country.** Apply the model to all regions to model the entire country. We create a graph with regions/districts as vertices and region borders as edges between vertices. This allows agents to travel between regions.
- **Simulate disease spread.** Using Matlab, we program our model and run thousands of simulations.
- **Collect data** about the current outbreak such as time-series data of Ebola cases and populations of all regions. Use this data to compare to the model results.
- **Adjust model parameters** to best fit the current data. This will allow us to have a model the accurately represents the outbreak. In return, we will be able to more realistically analyze the effects of medicine during January of 2015 of the outbreak.

The last step to this model will be to **introduce the population to Ebola medication**, and adjust the quantities and frequencies of medicine distribution to find an optimal solution to Ebola eradication.

2 Agent-Based SEIR Model

2.1 SEIR Model

For each town, we model the population with an SEIR model. At calendar day t, region j's population is comprised of:

- $S_j(t)$, the number of healthy persons who are **susceptible** to Ebola,
- $E_j(t)$, the number of persons who were **exposed** to Ebola, are now incubating, but are incapable of spreading the disease,
- $I_j(t)$, the number of **infected** persons who are symptomatic and capable of spreading Ebola, and

- $R_j(t)$, the number of **recovered** persons.

The corresponding system for our model is shown in Figure 1.

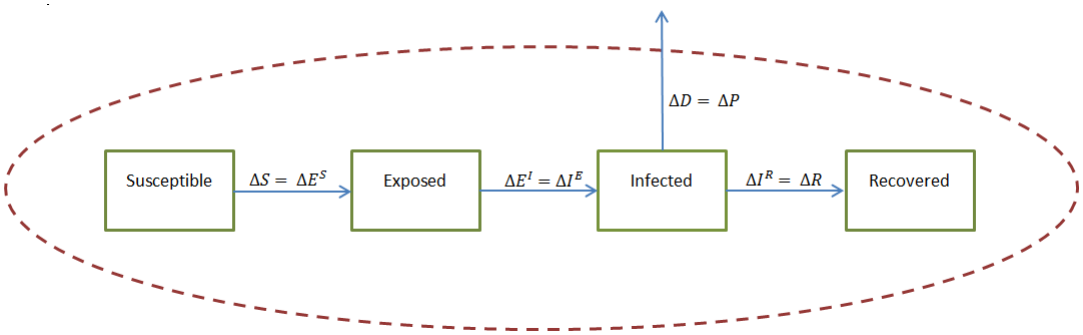

Figure 1: Basic SEIR system for some region. The dashed line represent the region.

The total population P for region j at day t is

$$P_j(t) = S_j(t) + E_j(t) + I_j(t) + R_j(t)$$

with the change in population

$$\Delta P_j(t) = \Delta D_j(t).$$

2.1.1 Susceptible

We assume all individuals start as healthy and susceptible to Ebola, except those currently who are infected at $t = 1$. The change in the number of susceptible persons from day t to day $t + 1$ is

$$\Delta S_j(t + 1) = S_j(t) - S_j(t + 1).$$

Then the number of susceptible individuals for day $t + 1$ is

$$S_j(t + 1) = S_j(t) - \Delta S_j(t + 1).$$

2.1.2 Exposed

When an individual has been exposed to the virus they will appear healthy for a number of days before beginning to show symptoms. During this time they are unable to infect other people.

The number of person who were healthy on day t and exposed on day $t+1$ is

$$\Delta E_j^S(t+1) = \Delta S_j(t+1).$$

The number of persons who were exposed on day t and began showing symptoms on day $t+1$ is

$$\Delta E_j^I(t+1) = \Delta I_j^E(t+1).$$

Then the number of exposed individuals on day $t+1$ is

$$E_j(t+1) = E_j(t) + \Delta E_j^S(t+1) - \Delta E_j^I(t+1).$$

2.1.3 Infected

Once an exposed person begins showing symptoms they will be sick for a number of days before recovering or dying. During this time it is possible for them to expose a healthy person to the virus. If a symptomatic person receives medicine, they are guaranteed to recover instead of dying.
The number of infected persons on day t who recover on day $t+1$ is

$$\Delta I_j^R(t+1) = \Delta R_j(t+1).$$

The number of infected individuals on day t is then

$$I_j(t+1) = I_j(t) + \Delta I_j^E(t+1) - \Delta I_j^R(t+1) - \Delta D_j(t+1).$$

2.1.4 Recovered

Recovered individuals are those who were once incubating or symptomatic. Once a person is recovered, the person will no longer be susceptible to Ebola. The number of recovered persons in region j at day t is

$$R_j(t+1) = R_j(t) + \Delta R_j(t+1).$$

2.1.5 Death

We assume individuals can only die from Ebola. We represent death by changing the region's population.

2.2 Agent Modeling with Stochastic Processes

We use A_j to be the set of all agents in region j. An agent $p_j \in A_j$ is an agent in region j. In this section, we will only be referring to one region, so for readability, we will not include the region subscripts (e.g. we will write A instead of A_j). The following convention is used for an agent p:

- $p^S \in A^S$ is a *susceptible* agent,
- $p^E \in A^E$ is an *exposed* agent,
- $p^I \in A^I$ is an *infected* agent, and
- $p^R \in A^R$ is a *recovered* agent.

Then $A = A^S \cup A^E \cup A^I \cup A^R$ with $A^S \cap A^E \cap A^I \cap A^R = \emptyset$.

2.2.1 Susceptible Agents

When an agent p^S becomes exposed on day $t+1$, then $A^S(t+1) = A^S(t) - p^S$.

2.2.2 Exposed Agents

We use an agent $p^E(t)$ to model an individual who is *exposed* on day t_0 and will become infected. The agent has the following property:

- Incubation time– the number of days until $p^E(t)$ becomes infected is κ.

According to the Center for Disease Control [1], the 2014 Ebola in Liberia and Sierra Leone has κ that best follows the distribution in Figure 1.

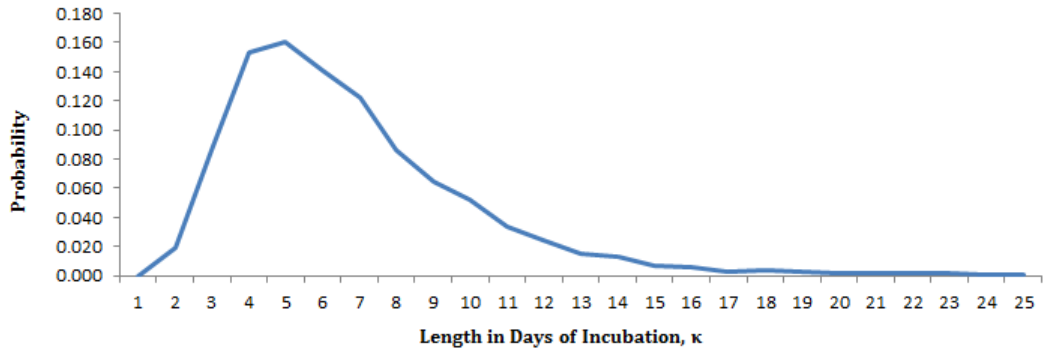

Figure 2: CDC recommended probability distribution function for incubation time, κ (source [1]).

When an agent $p^E(t)$ is created on day t_0, $p^E(t)$ is given κ so that on day $t = t_0 + \kappa$, $p^E(t)$ becomes $p^I(t)$, an infected agent.

2.2.3 Infected Agents

The number of infectious contacts made by an agent during its infected time period is called N; since our model is Stochastic, N is a discrete random variable [3].

For our model, we treat N as a Poisson random variable with $E[N] = \mathcal{R}_0$, the basic reproduction number [3]. Since N is Poisson [2], we write

$$P(N = k) = \frac{e^{-\mathcal{R}_0}\mathcal{R}_0^k}{k!}, \quad k = 0, 1, 2, 3, \ldots .$$

When an agent $p(t)$ becomes infected on day s, it is assigned N, it under goes a *Bernoulli process*

$$\{B(t) : s \leq t < t_d\}$$

where $B(t) \in \{0, 1\}$ is binary. Then $B(t)$ has the probabilities

$$P(B(t) = 1) = \frac{N}{t_d - s} \text{ and } P(B(t) = 0) = 1 - \frac{N}{t_d - s}.$$

Since $p(t)$ will stay infected for $t_d - s$ days, then $E[B(t)] = \frac{N}{t_d - s}$. If $B(t) = 1$, then agent $p(t)$ infects *one* other agent $p_o \in A^S$ at time$t$.

2.2.4 Recovered Agents

Recovered agents do not change their status. Applying any process to a recovered agent is trivial, since the agent will always stay as a recovered agent.

2.3 Markov Process

An agent p undergoes a *Markov process*, where its states can only be susceptible, exposed, infected, recovered, or removed from the total population (dead). We let $X(t)$ represent the current state of p so that

$$\{X(t) : t \in T\}$$

is the Markov process where T is the set of all calendar days, and $X(t)$ is in $\{S, E, I, R, D\}$ the set of all states. Each agent $p \in A$ undergoes the Markov process during day t to determine its state for day $t+1$. Note that process for a susceptible agent $p \in A^S$ is like the "compliment" of the Bernoulli process for an infected agent $p \in A^I$ since $p \in A^S$ has no stochastic properties, but is dependent on the stochastic, infectious property of an agent $p \in A^I$ (i.e. the only way for $p \in A^S$ to change states is too become infected).

Sierra Leone Represented as a Graph and Agent Travel

In our model, the agents have the ability to travel to different parts of the country. This feature was added to more accurately model the spread of disease.

Superimposed Graph

To model something as large and as spread out as a country, it was not safe to assume that the entire population lived in one area where they could all interact with each other, so to accurately model the spread of a disease we need to better represent population spread. To do this, we break a country up into regions. We represent each region as a *node* or *vertex* v in a *graph* G of Sierra Leone. In this graph, two nodes are connected if it is possible to travel directly from one region to another; thus, there is an *edge* E between the two nodes. An example of this is shown in figure 8. In this figure, each

point represents the capital of Sierra Leone's 14 districts/regions. The nodes are connected if their districts share a border.

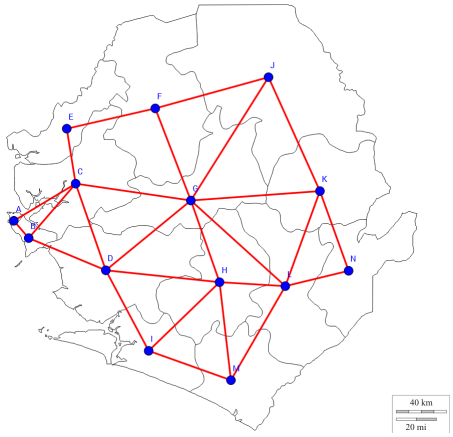

Figure 3: Map of Sierra Leone [5] with superimposed graph $G = (V, E)$.

For computational purposes, it is much easier to represent a graph as an adjacency matrix. The graph in Figure 8 can be represented by the 14×14 adjacency matrix in Figure 4. The adjacency matrix works as follows: in some column $z \in V$, the value for the index at row $y \in V$ is one if there is an edge from z to y and zero otherwise.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n
a	0	1	1	0	0	0	0	0	0	0	0	0	0	0
b	1	0	1	1	0	0	0	0	0	0	0	0	0	0
c	1	1	0	1	1	0	1	0	0	0	0	0	0	0
d	0	1	1	0	0	0	1	1	1	0	0	0	0	0
e	0	0	1	0	0	1	0	0	0	0	0	0	0	0
f	0	0	0	0	1	0	1	0	0	1	0	0	0	0
g	0	0	1	1	0	1	0	1	0	1	0	1	0	0
h	0	0	0	1	0	0	1	0	1	0	0	1	1	0
i	0	0	0	1	0	0	0	1	0	0	0	0	1	0
j	0	0	0	0	0	1	1	0	0	0	1	0	0	0
k	0	0	0	0	0	0	0	0	0	1	0	1	0	1
l	0	0	0	0	0	0	1	1	0	0	1	0	1	1
m	0	0	0	0	0	0	0	1	1	0	0	1	0	0
n	0	0	0	0	0	0	0	0	0	0	1	1	0	0

Figure 4: Adjacency Matrix for Sierra Leone representation

We define the set τ_z to be the set of all adjacent vertices for vertex z. An agent $p_z \in A_z$, where $z \in V$ can travel from region z to region $y \in V$ if $y \in \tau_z$. For example an agent in region a can only move to regions b and c.

All agents, except those who are infected, have the ability to travel. The disease can still still spread with this constraint since an exposed agent can carry the disease to new regions. An agent $p \notin A^I$ undergoes a *Bernoulli process* during day t to determine its region location for day $t+1$

$$\{Y(t) : t \in T\}$$

where $Y(t) \in \{0, 1\}$ is a random variable. The associated probabilities are

$$P(Y(t) = 0) = 1 - \gamma \text{ and } P(Y(t) = 1) = \gamma.$$

The value of γ is decided during simulation. When $Y(t) = 0$, agent p does not travel. When $Y(t) = 1$ the probability that the agent p in region z travels to a region $y \in V$ is

$$P(p(t+1) \in A_y | p(t) \in A_z) = \frac{1}{|\tau_z|}$$

where $|\tau_z|$ is the size or cardinality of the set τ_z. Clearly, all regions that have edges to region z have an equal probability.

3 Medicine Model

To model the distribution of medicine we used the following parameters:

- d, the delay or number of days between the beginning of the outbreak and the first shipment of medicine,
- r, the rate or number of days between shipments, and
- q, the quantity of medicine in one shipment.

3.1 Medicine Assumptions

- **Treatment does not discriminate based on age, race, gender, religion, education, income, or social status.** The distribution of medicine will be random except for if a person has been symptomatic for too long.
- **Medication delivery will be reliable.** There should be no problem delivering medication to large regions. For example, we will assume the medication will not be stolen or held by gangs.
- Medication does not have shelf life.

Medical shipments were initialized in the capital and are then split and distributed to other towns. The number of days between a shipment being initialized and reaching its destination town is equal to the length of the shortest path in the graph between the two vertices. Shipments are split proportionally to the number of current cases in each town.
Once a shipment of medicine arrives in a town it is immediately distributed to the sick population of the town. Priority is given to those who have been sick for the shortest time, since the problem statement states that the medicine works on those people whose disease is not advanced.
When an individual who has been sick for a small number of days receives medicine, they are guaranteed to recovery and their recovery time is shortened. This shortened recovery time is represented by supplying the agent with a new value for t_d,

$$t_d = \min\{t_d, t + t_r\},$$

where t_r is the number of days that the medicine takes to help an individual recover.
During the time before recovery the individual can still infect others, but does so at a much lower rate. For a medicated individual,

$$P(B(t) = 1) = \frac{m}{t_d - s},$$

where m is a parameter representing the basic reproduction number for medicated individuals.

4 Results

When testing our model we were first interested in finding an appropriate function to model $\mathcal{R}_0$. This was found through research as well as trial and error simulations. Next we ran simulations for a variety of medical distribution plans in order to determine the effectiveness of the new medication on an outbreak. Finally, we examined the strengths and weaknesses of our model.

4.1 Finding an appropriate $\mathcal{R}_0$

The first challenge in constructing a model on which to test medicine distribution plans was to determine an appropriate function for $\mathcal{R}_0(t)$. In order to do this we used data for the number of active Ebola cases in Sierra Leone between May 29, 2014 and January 29, 2015 and tested different functions for $\mathcal{R}_0(t)$ in order to find a good fit.
We primarily tested functions of the form

$$\mathcal{R}(t) = \begin{cases} \mathcal{R}_0 & : t < t_0 \\ \frac{\mathcal{R}_1 t_0 - \mathcal{R}_0 t_1 + (\mathcal{R}_0 - \mathcal{R}_1)t}{t_1 - t_0} & : t_0 \leq t < t_1 \\ \mathcal{R}_1 & : t_1 \leq t < t_2 \\ \mathcal{R}_2 & : t_2 \leq t \end{cases}$$

with $\mathcal{R}_0 > \mathcal{R}_1 > \mathcal{R}_2$. Our reasoning for this form came from [4]. The value $\mathcal{R}_0$ is a base infection rate before the nature of the infections is determined. Once it is determined that the Ebola virus is the cause of the infections, the infection rate drops quickly as word spreads and people begin taking

precautions against infection. Within around 10 days, the infection rate levels off and remains at that level for a long time before dropping even further.

Values for t_0 and t_1 were selected first as 20 and 30, respectively. These values were selected because they fit [6]. Next $\mathcal{R}_0$ and $\mathcal{R}_1$ were determined by running simulations using values between 1 and 4 for $\mathcal{R}_0$ and between 0.5 and 1.5 for $\mathcal{R}_1$. For each pair of values tested, we performed 100 simulations. An average simulation for that pair was then found by taking the mean of the active cases for each day. We discovered that $\mathcal{R}_0$ between 1.9 and 2.025 and $\mathcal{R}_1 = 1.1$ gave the best fit to the reported data for Sierra Leone.

We added $\mathcal{R}_2$ and t_2 to the function in order to account for a later decrease in the number of active cases in the Sierra Leone data. The value of t_2 was selected as 100 since the observed change in behavior occurred around day 115 and the need to account for the virus's incubation time. Simulations for $\mathcal{R}_2$ between 0.2 and 0.8 were then run in the same manner as earlier. It was found that $\mathcal{R}_2 = 0.65$ gave the best fit to the later data.

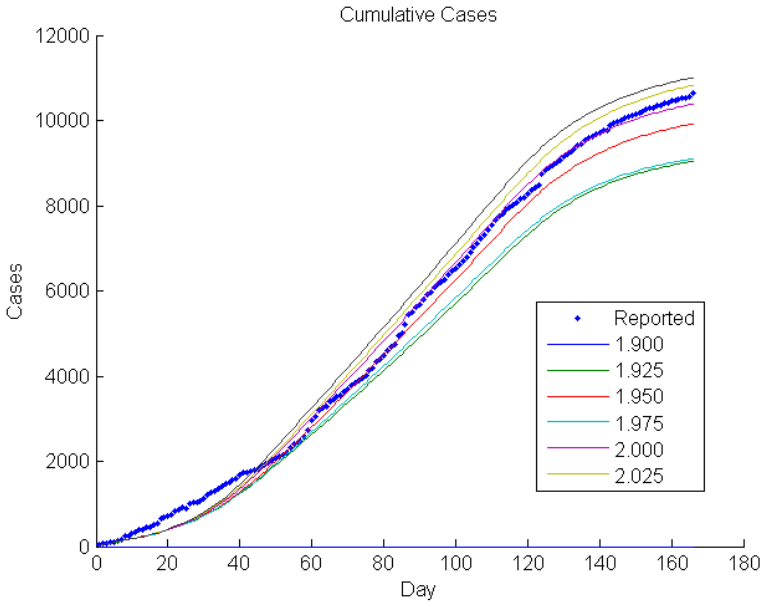

Figure 5: Comparison of reported data from Sierra Leone with simulation results for various values of $\mathcal{R}_0$. In each case $\mathcal{R}_1 = 1.1$ and $\mathcal{R}_2 = 0.65$.

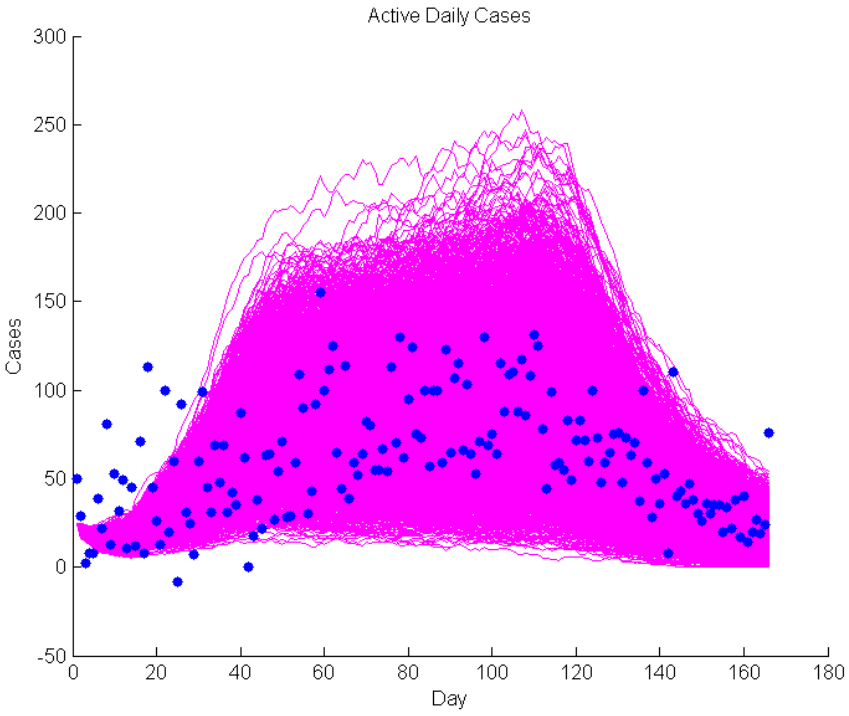

Figure 6: Comparison of active reported cases from Sierra Leone with simulation results for $\mathcal{R}_0 = 1.975$, $\mathcal{R}_1 = 1.1$ and $\mathcal{R}_2 = 0.65$.

In Figure 5, we see that the the model predicts that the number of cumulative cases approaches a horizontal asymptote, which is consistent with the data. However, the data shows linear growth in the first 40 days, where our simulation shows exponential growth.

In Figure 6, we see that the number of active cases on a given day can be bounded by our simulations for a large majority of the reported numbers of cases. The inconsistency of the reported daily for total daily cases is the main cause of error for this figure.

4.2 Medical Results

We simulated outbreaks with a wide variety of values for d, r, q, and m. In almost every case, the resulting infection followed a very similar pattern to the simulations involving no medicine. Simulations involving medicine

typically did not reach as high of a peak in the number of active cases and experienced a steeper and earlier decline in the number of active cases towards the end of the outbreak.
Even in simulations where medicine was available immediately or extremely quickly, the initial increasing trend was observed. The peak in active cases was significantly lower than in other simulations. The fact that the outbreak was not immediately suppressed is likely due to the high rate of infection early on in an outbreak. Infected agents were able to infect others during the short time in which $\mathcal{R}_0$ was high, leading to a spike in the number of cases once the exposed persons began showing symptoms.

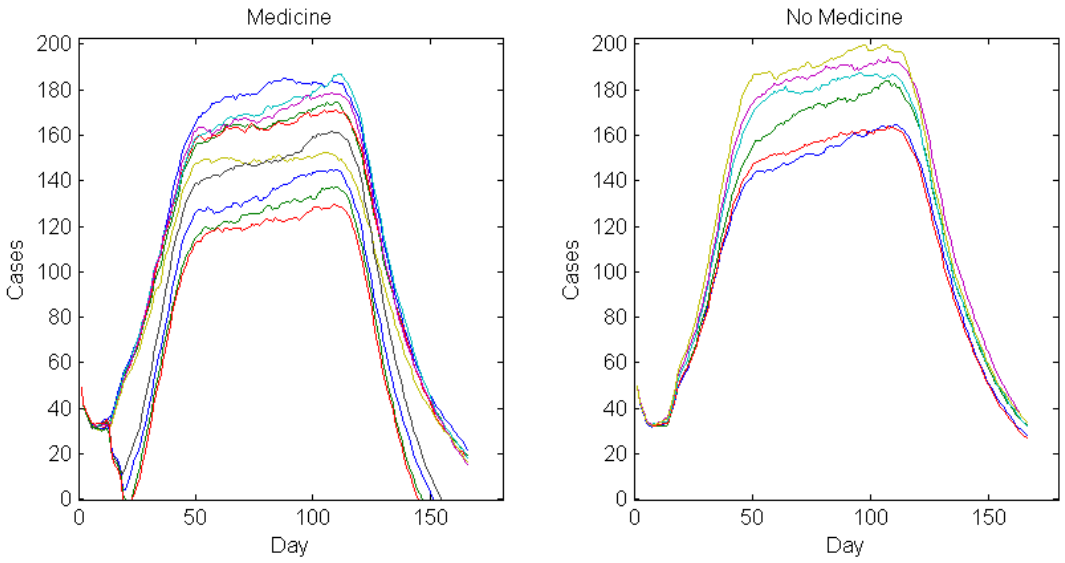

Figure 7: Comparison of the pattern of simulations with and without medicine included.

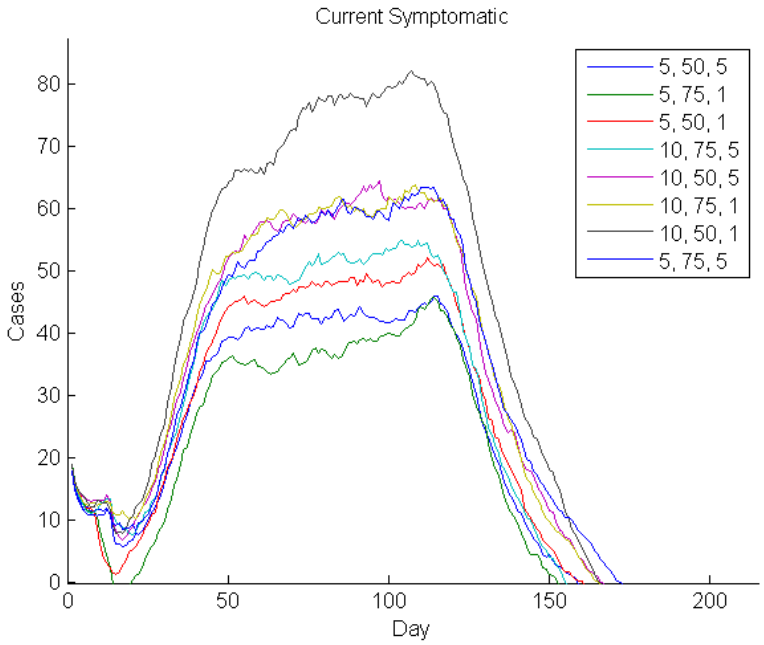

Figure 8: Simulations where medicine was available very quickly and in large quantities. The legend is r, q, d for the corresponding simulation.

4.3 Strengths

Ebola outbreaks are thought to have stochastic nature [4], so our model seems appropriate. Our model also agrees with our data, as shown in all the Figures 5-8.

4.4 Weaknesses

An obvious weakness of this model is our method for determining agent travel. We did not have any data about the likelihood of an arbitrary person from Sierra Leone to travel, thus, our methods are most likely wrong.

5 Future Work

In future work, we would like to test different probability distributions on our infected agents. We only used a Poisson variable for number of infectious

contacts, when the distribution could be better fit by another distribution. We would also like to test our model on all three of West Africa countries (Sierra Leone, Liberia, and Guinea) where the outbreak began. Finding better, more accurate data for the regions of Sierra Leone to compare with our model would be something of interest, since our model has the ability to count the number of outbreak in each region.

5.1 Experimenting with other Eradication Techniques

Although we were given that the world health organization is close to a cure/vaccine, and that we may even rely entirely on these medicines to eradicate Ebola, our model provides one vary interesting a realistic technique: quarantine. If the Ebola cases are known to be isolated in specific regions, we can enforce a "no travel" rule to these places and the disease could be contained. Since our model has a graph with each region as a vertex, we can find the *cut-edges* of a sub-graph, which would be the area to enforce the "no travel" rule. Of course, we would have to know the travel map for the country, and update our graph to contain these edges.

6 Conclusions

We conclude that medicine will eradicate Ebola in Sierra Leone 20% more quickly than without medicine. Additionally, medicine will reduce the peak number of cases by 50%. For these results, there were more than 25 units of medicine per shipment, with one shipment every 10 day.

Still, our results show that Ebola will be eradicated with out medicine. Our analysis shows that the reproduction rate, $\mathcal{R}_0$, is function of time and is *monotonically decreasing*. We believe this is due to Ebola awareness, and the awareness's effects on health workers' and civilians' precautions. These results also suggest that Ebola can eradicate itself, but not as quickly and with less certainty when compared to introducing Sierra Leone with medicine.

7 References

[1] M.I. Meltzer *et al.*,"Estimating the Future Number of Cases in the Ebola Epidemic–Liberia and Sierra Leone, 2014-2015,"*MMWR*, vol. 63, no. 3, Sep. 2014

[2] S. Ghahramani, "Special Discrete Distributions," in *Fundamentals of Probability with Stochastic Processes*, 3rd ed. Upper Saddle River, New Jersey, USA: Pearson, 2005, ch. 5, sec. 2, pp. 202-203

[3] P. Yan, "Distribution Theory, Stochastic Processes and Infectious Disease Modelling," in *Mathematical Epidemiology*, vol. 1945, Berlin: Springer, 2008, ch. 10, pp.236

[4] A. Camacho *et al.*, "Potential for large outbreaks of Ebola virus disease," in *Epidemics*, vol. 9, London: Elsevier, Dec. 2014, pp. 70-78

[5] Sierra Leone Districts. 2015, [Online]. Available: http://d-maps.com/m/africa/sierra/sierra34.gif. Accessed: Feb. 8, 2015.

[6] OCHA ROWCA, "Sub-national time series data on Ebola cases and deaths in Guinea, Liberia, Sierra Leone, Nigeria, Senegal and Mali since March 2014," [Online]. Available: https://data.hdx.rwlabs.org/dataset/rowca-ebola-cases. Accessed Feb. 9, 2015.