Comparison of multi-compartmental models for the modeling of the 2014 Ebola epidemic

Iris Lu, Andrew Teng, Yuanbo Wang
Georgia Institute of Technology

I. INTRODUCTION

Ebola virus disease (EVD, or simply Ebola) is a rare and deadly disease that is highly transmissible from person-to-person. It is commonly spread to people from wild animals and is spread to other humans through contact with bodily fluids (saliva, blood, etc.) of someone who is infected. The average fatality rate is 50%, making Ebola one of the most deadly diseases [1].

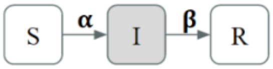
The Ebola virus was first documented in 1976 in Zaire, and there have been several incidences with total cases on the order of hundreds prior to the 2014 epidemic [1]. In 2014, West Africa saw the world's largest and most complex Ebola outbreak ever recorded with 28,639 total cases [1]. Guinea, Liberia and Sierra Leone were the most affected countries and unfortunately also have very weak health systems as well as a lack of human and infrastructural resources. This lack of resources combined with a deadly disease created an epidemic never seen before. In order to make accurate and vital decisions during the outbreak, understanding the transmissibility of the Ebola was necessary in determining how to halt its spread. As data was recorded and the virus was being analyzed in real-time, it was apparent that this was a case where one model does not fit all. Various literature reviews have published finding using different compartments to model the spread of Ebola. Therefore, it is necessary to investigate the various methods for predicting Ebola dynamics and to compare the results against historical data.

II. BACKGROUND

Ebola is a deadly and infectious disease and tools are necessary to predict the next outbreak in order to minimize the size and duration of the outbreak if not completely prevent the outbreak. There are various methods to model the transmission of EVD and can be used to predict the next outbreak. With that knowledge, appropriate intervention efforts can be implemented in the field to curb the extent of the outbreak. Compartment modeling is a mathematical approach that is used to describe transmission among the compartments of a specific system. In the models and approaches discussed below, movement in the model can only stay in the initial state or go forward, without skipping through the states. The total population (N) in these models are represented by the total of individuals in each group. Pending the duration of the epidemic is being studied, this N is assumed to remain constant. Combinations of these models have been developed to achieve various objectives.

2.1 Models

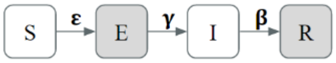
The simplest compartmental model available to represent and study an epidemic would be the susceptible, infectious, and recovered (SIR) model. In order to utilize a SIR model, a few criteria need to be considered. First, the disease in study needs to have a severe outbreak. Everyone who is infected is removed from the population either through recovery or death. Finally, the population must be large, fixed in size, and is confined to a geographic region [2]. Mathematically, the SIR model can be simply denoted with the following three ordinary differential equations, where *dS* represents the change in the susceptible group, *dI* represents the change in the infected group, and *dR* represents the change in the recovered/removed group.



$$\frac{dS}{dt} = -\alpha SI$$
$$\frac{dI}{dt} = \alpha SI - \beta I$$
$$\frac{dR}{dt} = \beta I \quad (1)$$

The infected group can change over time as susceptible individuals become infected and infected individuals recover and therefore move into the recovered group. The parameters α and β are used, where α (*units: time*$^{-1}$*individuals*$^{-2}$) represents the rate of infection, β (*units: time*$^{-1}$*individuals*$^{-1}$) represents the rate of recovery, and $S(t)+I(t)+R(t)=N$. The SIR model is a very common model used to study the spread of Ebola due to its simplistic nature; however, an SIR model may be too simple of a model to use in some situations. In those cases, adding another compartment for those who are exposed may be beneficial, creating a SEIR model.

The SEIR model follows the SIR model but includes an additional compartment, exposed individuals (E). Because of the additional compartment, the system of equations expands to incorporate the dynamic changes caused by the exposed group.



$$\frac{dS}{dt} = -\epsilon SI$$
$$\frac{dE}{dt} = \epsilon SI - \gamma E$$
$$\frac{dI}{dt} = \gamma E - \beta I$$
$$\frac{dR}{dt} = \beta I \qquad (2)$$

The infected group can change over time as susceptible individuals become exposed, those who are exposed become infected, and the infected individuals recover and move into the recovered group. The parameters *ε*, *γ*, and *β* are used, where *ε* (*units: time*$^{-1}$*individuals*$^{-2}$) represents the rate of exposure, *β* (*units: time*$^{-1}$*individual*$^{-1}$) represents the rate of recovery, *γ* (*units: time*$^{-1}$*individuals*$^{-1}$) represents the rate of infection, and N is the sum of the four groups. Simply, 1/*γ* is average duration of incubation and 1/ *β* signifies the average duration of infectiousness [3].

When modeling the spread of an epidemic, it may be useful to scale or limit the number of susceptible individuals. Rachah, *et al.* determined that the number of susceptible individuals is equal to the highest number in the number of cases [4]. This scaled-down approach allows for faster computation while retaining accuracy.

Many intervention efforts have been taken to control and determine epidemic size and duration. Interestingly, the Center for Disease Control and Prevention (CDC) developed a tool, EbolaResponse, to help predict the number of beds necessary in Ebola treatment units (ETUs) to help stop the spread of Ebola and predict the potential number of future cases [5]. They implemented the 'incubating and infectious' group, forming a SIIR model which is essentially a Markov chain model. Data from previous studies were used to create a lognormal probability distribution of being in the incubation state. The SIIR model also assumed that the average infectious state was six days and the risk of transmission was assumed to be equal throughout the six days [5]. A few limitations that can occur and possibly create errors would include the amount of underreporting, movement of individuals between geographic borders, and common burial practices that can expose healthy individuals to the Ebola virus. Since Ebola is a unique disease in the fact that it is still transmissible after the infected individual dies, it may be useful to add a compartment to take this factor into account.

There are many compartments that can be further integrated into these models. Table A1 covers the various compartments commonly used in epidemiology and how each group is defined. Compartments such as Maternal Immunity (M) and Carrier (C) are not found in models for Ebola, because Ebola is not transmitted in those manners. The most basic compartments are S, I, and R and can be found in every model. Advantages of the simplistic model is the lower requirement of datasets necessary to develop the SIR model. There are different rates of recovery based on infectious level, therefore compartments E, I_1, and I_2 have been used to differentiate the groups in the population. Transmission of Ebola from the handling of dead bodies during the funeral procedures was a large enough contributing factor to transmission that the F compartment was added. Interventions (H/P/etc.) compartments can indicate whether its implementation has or will change the outcome of the model. Unfortunately, as the model becomes more complex, the dataset required grows more extensive in order to develop the model. This complexity can be difficult to gather in areas where infrastructure is not conducive to a standardized method of reporting, especially prior to a major outbreak.

2.2 Parameter Estimation

The parameters used to model Ebola can be calculated based on previous historical or current data. Since each compartmental model can mathematically be represented by a system of ordinary differential equations, the parameters, such as rate of infection and rate of recovery, can be fitted using a least-squares fitting approach [6], [7]. Here, the sum of squared errors between the data and the mathematical solution are minimized. Assumptions in regards to initial conditions and stage progression are made in order to solve the differential equation. In a deterministic model, there is no randomness or variation in the parameters, and therefore, the outcomes remain constant.

Different studies used various parameters to study Ebola. In a study performed by Rachah, *et al*., a simulation of an Ebola outbreak was performed using *Kaurov's estimated parameters* [4]. Kaurov's estimation of parameters are used to estimate the rate of infection and rate of recovery where it is assumed that 95% of the population is susceptible and 5% of the population is infected making α=0.2 and β=0.1.

Using the SEIR model as previously described, Althaus, *et al*. study used maximum likelihood estimates of the parameters. They were found by fitting the model created to the data with the assumption the number of cases and deaths were Poisson distributed [8], [9]. With the additional exposed compartment, it was determined that the model created by Althaus, et al. fit the data well and suggested that control interventions were not successful [8]. Unfortunately, a limitation of this model would be that the model could not account for fluctuations in the number of new cases and that the transmission rate decays rapidly from control measures after the first infectious case.

2.3 Measurement Performance

There are multiple ways to measure model performance. Comparing the predicted data with the actual data would allow for one to assess the accuracy of the measurement. The root mean square error (RMSE) was used in order to assess the quality of the estimators, how well the model fit to the data used to develop the model, and the quality of the predictors, how well the model fits to the data predicted. This metric allows for variability in data length between model during comparisons and therefore providing a level of robustness.

To evaluate how intervention implementation impacted the epidemic, a multivariate uncertainty and sensitivity analysis can be performed. The Legrand group varied specific intervention parameters mentioned earlier and simulated 700 epidemics and computed the mean size of these epidemics. The partial rank correlation coefficients (PRCCs) were computed between each varying parameter and the mean size and quantifies this linear relationship between the intervention and epidemic size [10].

III. METHODS

3.1 Data Collection

In order to properly model the Ebola virus, three datasets were utilized, one from the World Health Organization (WHO), one from the Center for disease control and Prevention (CDC), and the one from Virginia Polytechnic Institute and State University (Virginia Tech). The WHO dataset covers the most recent 2014 Ebola outbreak starting August 29, 2014 while the CDC data starts in March 2014. The WHO and CDC datasets provide us with cumulative case information for different infection status, whereas the Virginia Tech dataset provides data that will allow for the evaluation of the intervention methods implemented in the field in West Africa.

CDC provides cumulative number of infected cases for Guinea, Liberia, and Sierra Leone, while WHO provides epidemiological data covering cumulative confirmed cases in Guinea, Liberia, and Sierra Leone. The data are separated either by age groups, 0-14, 15-44, and 45+ years, or gender [11]. The data can be sorted by country and then divided into the following groups: confirmed, probable, and suspected. The total population size is approximately 15,889 when looking at all three countries. Depending on how the data is subdivided, some of the counts will vary due to cases where the gender is known but the age was not and vice versa, which results in variable total counts.

Caitlin Rivers, a computational epidemiologist at Virginia Tech, has a curated, de-identified, and publicly available Ebola dataset on GitHub [12], [13]. The dataset contains publicly released data from the World Health Organizations as well as the Ministries of Health of the affected countries. It specifically only contains laboratory confirmed, suspected or probable cases of the disease as Rivers claims that is the best representation and estimation of the state of the Ebola epidemic. The data are separated by many variables; however, each country in question has a different number of variables based on the data that was collected. There are many variables that are found in each country's dataset including total deaths, gender (male or female), hospital admissions, and suspected/confirmed/ probable cases.

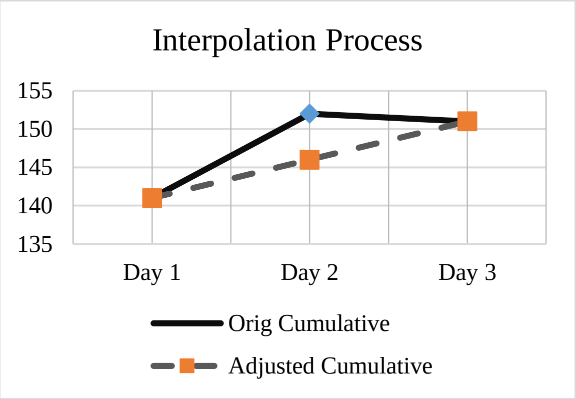


Figure 1: Visualization of data cleaning and linear interpolation process

Cumulative drops were corrected through linear interpolation of the counts between the two closest enclosing report dates. Specifically, if a, b, *and* c *were the counts collected at three incrementing timestamps, where* $a<c$ *and* $b>c$, *we replace* b *with* $b'=(c-a)/2$, *so that* $a<b'<c$. *Missing dates were filled using linear interpolation; if no data were reported between dates* x *and* y, *where* $x<y$ *and* $y-x=n>1$, *then the count for date* $x+k$ *is calculated as* $a+k(b-a)/n$, *where* a *and* b *are the counts collected on* x *and* y, *respectively, and* $x+k<y$.

3.2 Data Cleaning

Once the datasets were collected, they were cleaned first by adjusting cumulative cases to remove negative changes in exposed, infected, and intervened individuals through a linear interpolation process. The interpolation process is the most conservative process available and is illustrated in Figure 1**.** The data were then normalized with the time intervals of the data report and the initial population sizes were scaled. Compartmental models require that the input data have equal time intervals. Data for missing dates were filled in through linear interpolation on the counts from the two closest enclosed reporting dates described below Figure 1. The initial population size for each country of interest were downsized to 2000 individuals. Rachah *et al.* also reduced the initial number of individuals based and it did not affect the model results. The new number of individuals was still able to produce expected results. After performing the dataset cleaning, the curated dataset was then used for training and testing of the three compartmental models of interest. Table 1 lists the details of this finalized dataset.

3.3 Workflow

The cleaned data set are partitioned (explained in 3.6 under cross-validation) into seven parts, where one is left out for external validation, and are loaded into the workspace. The population N is using the same approach as described in Rachah *et al*. The rates of movement for the compartments of interest are calculated based on these inputs.

Next, the initial conditions to develop the compartmental models are defined. The initial data point in the time window of interest is taken. The relevant parameters for a specific model are initialized. The time range of interest is defined. These initial conditions are inputted into a built-in MATLAB function “fmincon” that optimizes the parameters of interest. The parameter values are restricted to the range of [0,1]. Parameter optimization is performed by minimizing the root mean squared error when comparing the model predictions to the dataset held in the cross validation. This optimization was done for the six folds in the cross-validation implemented. The parameters values are calculated based on solving a system of ordinary differential equations that have been defined for each model (Equations 1-3).

Table 1: Data Span, Number of Days Reported, and Number of Weeks Covered by Final Dataset

Country	**Model**	**Data Span**	**Number of Days Reported**	**Number of Weeks after Interpolation**
Liberia	SIR	3/1/2014-2/17/2016	156	99
	SEIR, SEII$_n$R	7/1/2014-10/5/2014	47	14
Sierra Leone	SIR	3/1/2014-2/17/2016	156	91
	SEIR, SEII$_n$R	8/13/2014-10/13/2014	15	4
Guinea	SIR	3/1/2014-2/17/2016	156	103

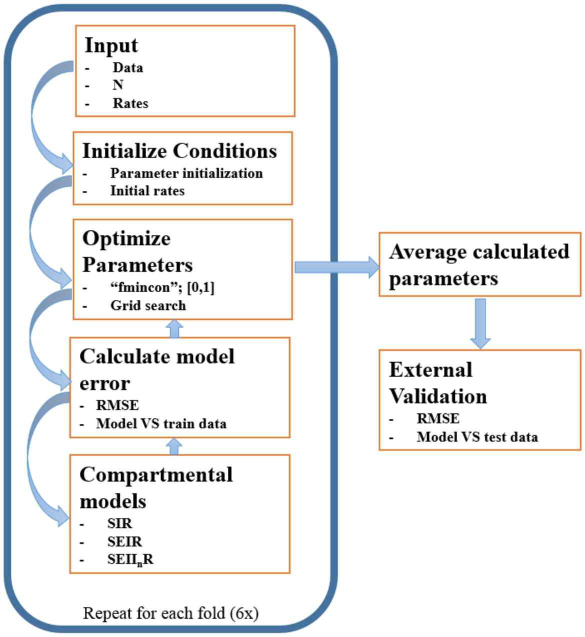


Figure 2: Workflow

Once the six sets of parameters are calculated, the average parameter value was taken and used as the final model. These final parameters are externally validated with the seventh partition, where the RMSE is calculated for each model. This workflow can be visualized in Figure 2.

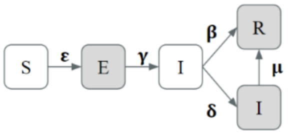
3.4 Model selection

In order to investigate different methods of predicting the dynamics of EVD, three compartmental models will be developed in this study, SIR, SEIR, and SEII$_n$R (susceptible, exposed, infectious, intervention- those at a treatment center, and recovered).

The SIR model will describe how the population transitions from groups *S(t)* to *I(t)* and *I(t)* to *R(t)*. This is the most basic compartmental model and has been chosen to see if it can model and predict Ebola outbreaks and therefore, reduce resources allocated to collecting more information for only a marginally better model.

The SEIR model will follow the same flow with the exception of the added exposed compartment because EVD has a long enough incubation period of 2-21 days but often times falling around 8 days [CDC].

The SEII$_n$R model will include an addition intervention group in order to understand how different intervention implementations affected the outcome of the epidemic. More specifically, the intervention of interest is the rate of people infected who go to the hospital. Modulating the availability of hospitalizations would give insight to whether an increase in clinics and hospital centers would be enough to prevent an outbreak and how an organization should allocate their limited resources and funds to best minimize or avoid an outbreak.



$$\begin{aligned}\frac{dS}{dt} &= -\epsilon SI\\ \frac{dE}{dt} &= \epsilon SI - \gamma E\\ \frac{dI}{dt} &= \gamma E - \beta I - \delta I\\ \frac{dI_n}{dt} &= \beta I - \mu I_n\\ \frac{dR}{dt} &= \delta I + \mu I_n\end{aligned} \qquad (3)$$

Assumptions are made for these models such as the population can be grouped into one of the compartments of a specific model. Susceptible individuals coming into contact with infected individuals will be exposed and may be infected at some given probability. For those recovered and survived are assumed to have immunity to the disease, as stated in literature.

The parameters of these models are defined as the following: ε (*units: time*$^{-1}$*individuals*$^{-2}$) is the rate of exposure, β (*units: time*$^{-1}$*individuals*$^{-1}$) is the rate of recovery, γ (*units: time*$^{-1}$*individuals*$^{-1}$) is the rate of infection post-incubation, δ (*units: time*$^{-1}$*individuals*$^{-1}$) is the rate those undergoing the intervention, and μ (*units: time*$^{-1}$*individuals*$^{-1}$) is the rate of recovery post-intervention. Ultimate, the expected outcome will be infection counts.

3.5 Performance and Comparison Metrics

To evaluate the performance of each individual model, the Root Mean Squared Error (RMSE) measure was used. The RMSE of predicted values $\hat{y}_t$ for times *t* of a dependent variable y_t is computed for *n* different predictions as in the equation below:

$$RSME = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \qquad (4)$$

The goal of this project is to explore the model and corresponding parameters that minimize the RMSE.

For each model, cross-validation will be performed by optimizing model parameters using a training set and evaluating accuracies using an external test set.

3.6 Cross-validation

The data are partitioned into a training set, which contains data collected on Mondays through Saturdays, and a test set, which contains data collected on Sundays. We then performed six-fold cross-validation on the training set for each compartmental model:

1) Data Partition. The training set was partitioned into six folds, where each fold contains time series data (infected, exposed, intervention) collected on Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday, respectively. The number of weeks covered by the data for each country and model are listed in Table 2 below.

Table 2: Number of Data Points (Weeks) in Each Fold by Model Type and Country

	Liberia	**Sierra Leone**	**Guinea**
SIR	98	90	102
SEIR	13	3	N/A
SEII$_n$R	13	3	N/A

2) Cross-validation. Leaving out one fold at a time, For each week, the average count was taken over the five folds, each model was trained using the averaged data, and the parameters obtained were used to test the model on the left-out fold to obtain the RMSE. This process was repeated 5 times, leaving out one fold at a time.

3) Model comparison. Parameters obtained from the six-fold cross-validation were averaged to obtain the parameters for modeling the external test set (Sundays' data). The training RMSEs and testing RMSEs were used to compare the performance of the three compartmental models for each country.

3.7 Parameter Estimation

When initializing the parameters and optimizing the parameters using the built-in MATLAB function of "fmincon", a constrained minimization method using the Nelder-Mead algorithm, limits the values of the parameters between zero and one. This constraint is necessary for this model as the parameters describes the rate of forward movement of individuals from one compartment to another. Therefore, negative rates indicating reverse movement of individuals, and numbers greater than 1 indicating more individuals than those in the compartment of interest moved into another are illogical. Initially, a grid search was implemented in order to find the best combination of each parameter value that were constrained between 0 and 1 with a step value of 0.01 in order to better select parameter initialization values. However, this process was computationally expensive and did not yield significantly better results.

3.8 Model Development

A graphical user interface (GUI) was developed in parallel. The GUI allows for users to interactively compare the three compartmental models of interest (SIR, SEIR, SEII$_n$R) using the curated dataset based from CDC, WHO, and Virginia Tech. The GUI optimizes the parameter for each set of ordinary differential equations for all three models. The interface allows the user to select a country to analyze (Liberia, Guinea, or Sierra Leone) and the models will automatically generate and iterate through all the cross-validation folds. The user is given the option to input their own parameters. The model will update accordingly and a RSME value will also be produced for the new model with the inputted parameters. Each model also features a parameter diagram for intuitive understanding of the parameter meanings as well as an error chart. Furthermore, internal and external validation table results are included to validate the results. A screenshot of the GUI can be seen in Figure A1.

IV. RESULTS

4.1 Liberia

When Liberia is chosen as the country of interest in the GUI, parameters for three models were computed along with the RSME for the internal and external validation.

The SIR model resulted in the average rate of infection α = 0.4128 week^{-1}individual^{-2} and the average rate of recovery β = 0.0946 week^{-1}individual^{-1}. This average was calculated based on the optimized parameters for each fold in the internal cross validation. In the external validation of the model, the RSME = 0.0308.

The SEIR model resulted in the average rate of exposure ε = 0.7149 week^{-1}individual^{-2}, the average rate of infection γ = 0.6381 week^{-1}individual^{-1}, and the average rate of recovery β = 0.1583 week^{-1} individual^{-1}. This average was calculated based on the optimized parameters for each fold in the internal cross validation. In the external validation of the model, the RSME = 0.0305.

The SEII$_n$R model resulted in the average rate of exposure ε = 0.7935 week^{-1}individual^{-2}, the average rate of infection γ = 0.689 week^{-1}individual^{-1}, the average rate of recovery β = 0.0999 week^{-1}individual^{-1}, the average rate of hospitalization δ = 0.1008, and the average rate of recovery from hospitalizations μ = 0.4814 week^{-1}individual^{-1}. This average was calculated based on the optimized parameters for each fold in the internal cross validation. In the external validation of the model, the RSME = 0.0315.

4.2 Sierra Leone

When Sierra Leone is chosen as the country of interest in the GUI, parameters for three models were computed along with the RSME for the internal and external validation.

The SIR model resulted in the average rate of infection α = 0.5961 week^{-1}individual^{-2} and the average rate of recovery β = 0.2096 week^{-1}individual^{-1}. This average was calculated based on the optimized parameters for each fold in the internal cross validation. In the external validation of the model, the RSME = 0.0016.

The SEIR model resulted in the average rate of exposure ε = 0.846 week^{-1}individual^{-2}, the average rate of infection γ = 0.8174 week^{-1}individual^{-1}, and the average rate of recovery β = 0.0534 week^{-1} individual^{-1}. This average was calculated based on the optimized parameters for each fold in the internal cross validation. In the external validation of the model, the RSME = 0.0083.

The SEII$_n$R model resulted in the average rate of exposure ε = 0.9966 week^{-1}individual^{-2}, the average rate of infection γ = 0.9956 week^{-1}individual^{-1}, the average rate of recovery β = 6.18e-4 week^{-1} individual^{-1}, the average rate of hospitalization δ = 6.18e-04, and the average rate of recovery from hospitalizations μ = 0.5269 week^{-1}individual^{-1}. This average was calculated based on the optimized parameters for each fold in the internal cross validation. In the external validation of the model, the RSME = 0.0059.

4.3 Guinea

When Guinea is chosen as the country of interest in the GUI, parameters for the available models were computed along with the RSME for the internal and external validation.

The SIR model resulted in the average rate of infection α = 0.5332 week^{-1}individual^{-2} and the average rate of recovery β=0.3542 week^{-1} individual^{-1}. This average was calculated based on the optimized parameters for each fold in the internal cross validation. In the external validation of the model, the RSME = 0.0132.

The SEIR and SEII$_n$R models were not computed because of the lack of a dataset to develop the model.

4.4 Summary

Summary of the average parameters calculated after internal cross validation are found in Tables 3-5 below.

Table 3: SIR average parameters

Country	**Alpha**	**Beta**
Liberia	0.4128	0.0946
Guinea	0.5332	0.3542
Sierra Leone	0.5961	0.2096

Table 4: SEIR average parameters

Country	**Epsilon**	**Gamma**	**Beta**
Liberia	0.7149	0.6381	0.1583
Sierra Leone	0.8460	0.8174	0.0534

Table 5: SEII$_n$R average parameters

Country	**Epsilon**	**Gamma**	**Beta**	**Delta**	**Mu**
Liberia	0.7935	0.6890	0.0999	0.4814	0.0315
Sierra Leone	0.9966	0.9956	6.2e-4	6.2e-4	.0059

Summary of the average training RMSEs for each model and country and the corresponding external validation RMSEs are listed in Table 6. Specific parameter values for each country are described in the sections 4.1-4.3.

Table 6: Summary of Training and Test RMSEs.

Country	Model	Average Training RMSE	External Test RMSE
Liberia	SIR	0.0392	0.0308
	SEIR	0.0392	0.0305
	$SEII_nR$	0.0402	0.0315
Sierra Leone	SIR	0.0126	0.0016
	SEIR	0.0154	0.0083
	$SEII_nR$	0.0112	0.0059
Guinea	SIR	0.0170	0.0132

V. DISCUSSION

Estimated model parameters were compared with the Spatial Temporal Epidemiological Modeler (STEM) Tool [15]. This tool outputs ranges for the parameters estimated for the SIR and SEIR models. The SIR and SEIR parameter estimations fell within these outputted ranges, which further validated our results.

The dataset was partitioned into training and testing sets based on days of the week due to the time-dependent nature of the compartmental models. This method is essential for building accurate models although it may introduce bias as data collection may be regularly scheduled and therefore affect our training set.

When comparing the three compartmental models with each other, it is apparent that the SIR model overall performed the best. The SEIR model performed marginally better than the SIR model under the Liberia case; however, it is not significant. Intuitively, it may seem that the SIR model would perform the worst as it has the fewest compartments; however, the models are highly dependent on the quality of the inputted data. In this case, the exposed data was not ideal and made the SIR model the overall stronger performer. Furthermore, this highlights the fact that the SIR model is a solid model in predicting the flow of infectious diseases. Researchers out in the field may have limited resources in data collection and may not be able to gather enough data for an exposed compartment. The results indicate that the most basic compartmental model is sufficient for modeling and predicting the spread of Ebola, which can lower field and data collection costs. In order for the SEIR and $SEII_nR$ to have better model performance, more data would be required.

When comparing the parameters amongst the countries, Sierra Leone consistently had the highest infection rates than Liberia in every model, which parallels with the raw cumulative data where Sierra Leone has the steepest curve indicating higher infection rates.

Looking at the three compartmental models for a specific country, the infection rate $\alpha = 0.4128$ in the SIR model is similar to the corresponding infection rates in the SEIR and $SEII_nR$ models ($\varepsilon * \gamma \approx \alpha$) where $\varepsilon * \gamma = 0.4562$ and $\varepsilon * \gamma = 0.5467$. The infection rates for each of the models are similar confirming the calculated parameter values.

For Sierra Leone, the SEIR and $SEII_nR$ model resulted in extreme parameter values (closer to either 0 or 1. This result is likely due to the limited amount of data available to develop the model. The number of data points available and implemented was three. Therefore, more information is necessary to build a model that better reflects the infection rate of EVD in this country.

VI. FUTURE WORK

Modeling of infectious diseases is vital to understanding how disease spreads from person-to-person. Defining errors differently can yield possibly different results and should be experimented with. Since epidemiology proposes a multitude of compartments, further additions of various compartments can be tested. Ebola and other diseases do not have geographic limits. Being able to apply these models to other countries in West Africa can also help understand the spread and predict the dynamics of Ebola.

VII. REFERENCES

[1] WHO, “Ebola virus disease,” 2016. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs103/en/#.

[2] C. Bekoe, “The SIR Model and the 2014 Ebola Virus Disease Outbreak in,” *Int. J. Appl. Sci.*, vol. 6, no. 2, pp. 11–24, 2015.

[3] G. Zaman, Y. H. Kang, and I. H. Jung, “Optimal treatment of an SIR epidemic model with time delay,” *BioSystems*, vol. 98, no. 1, pp. 43–50, 2009.

[4] A. Rachah and D. F. M. Torres, “Mathematical modelling, simulation, and optimal control of the 2014 ebola outbreak in West Africa,” *Discret. Dyn. Nat. Soc.*, vol. 2015, 2015.

[5] Martin I. Meltzer, C. Y. Atkins, S. Santibanez, B. Knust, B. W. Petersen, E. D. Ervin, S. T. Nichol, I. K. Damon, and M. L. Washington, “Estimating the Future Number of Cases in the Ebola Epidemic — Liberia and Sierra Leone , 2014 – 2015,” *Morb. Mortal. Wkly. Rep.*, vol. 63, no. 3, pp. 1–5, 2014.

[6] M. C. Eisenberg, J. N. S. Eisenberg, J. P. D’Silva, E. V. Wells, S. Cherng, Y.-H. Kao, and R. Meza, “Modeling surveillance and interventions in the 2014 Ebola epidemic,” no. December, pp. 1–30, 2015.

[7] Z.-Q. Xia, S.-F. Wang, S.-L. Li, L.-Y. Huang, W.-Y. Zhang, G.-Q. Sun, Z.-T. Gai, and Z. Jin, “Modeling the transmission dynamics of Ebola virus disease in Liberia.,” *Sci. Rep.*, vol. 5, p. 13857, 2015.

[8] C. L. Althaus, “Estimating the Reproductive Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa,” *PLOS Curr. Outbreaks*, 2014.

[9] P. E. Lekone and B. F. Finkenstädt, “Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study,” *Biometrics*, vol. 62, no. 4, pp. 1170–1177, 2006.

[10] J. Legrand, R. F. Grais, P. Y. Boelle, A. J. Valleron, and A. Flahault, “Understanding the dynamics of Ebola epidemics,” *Epidemiol. Infect.*, vol. 135, no. 4, pp. 610–21, 2007.

[11] World Health Organization, “Ebola situation reports: archive.” [Online]. Available: http://www.who.int/csr/disease/ebola/situation-reports/archive/en/.

[12] C. M. Rivers, E. T. Lofgren, M. Marathe, S. Eubank, and B. L. Lewis, “Modeling the Impact of Interventions on an Epidemic of Ebola in Sierra Leone and Liberia,” *Plos Curr. outbreaks*, 2014.

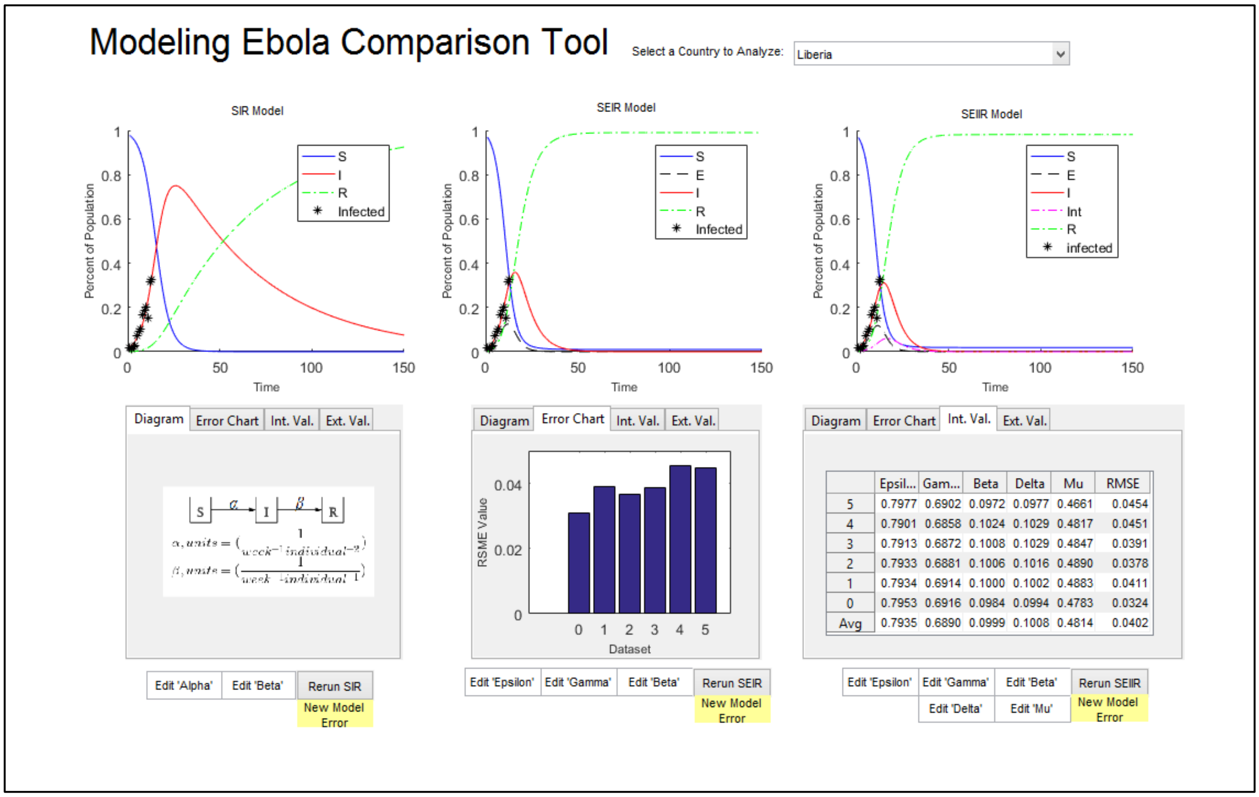
[13] C. M. Rivers, “Data for the 2014 ebola outbreak in West Africa.” [Online]. Available: https://github.com/cmrivers/ebola.

Appendix

Table A1: Various Compartments in Epidemiology

Compartments	**Definition**
Susceptible (S)	Those who are at risk of disease
Exposed (E)	Those who have been exposed to the disease
Infectious (I)	Those who have contracted and are infected with the disease
Recovered (R)	Those who are no longer infected or dead and buried
Maternal Immunity (M)	Those who are born immune to disease from maternal antibodies
Funeral (F)	Those who have died from the disease, but not buried
Carrier (C)	Those who naturally carry the disease
Infectious stage 1 (I_1)	Those who are in first stage of infection
Infectious stage 2 (I_2)	Those who are in second stage of infection
Intervention (H/P/etc)	Those who have been hospitalized/received treatment

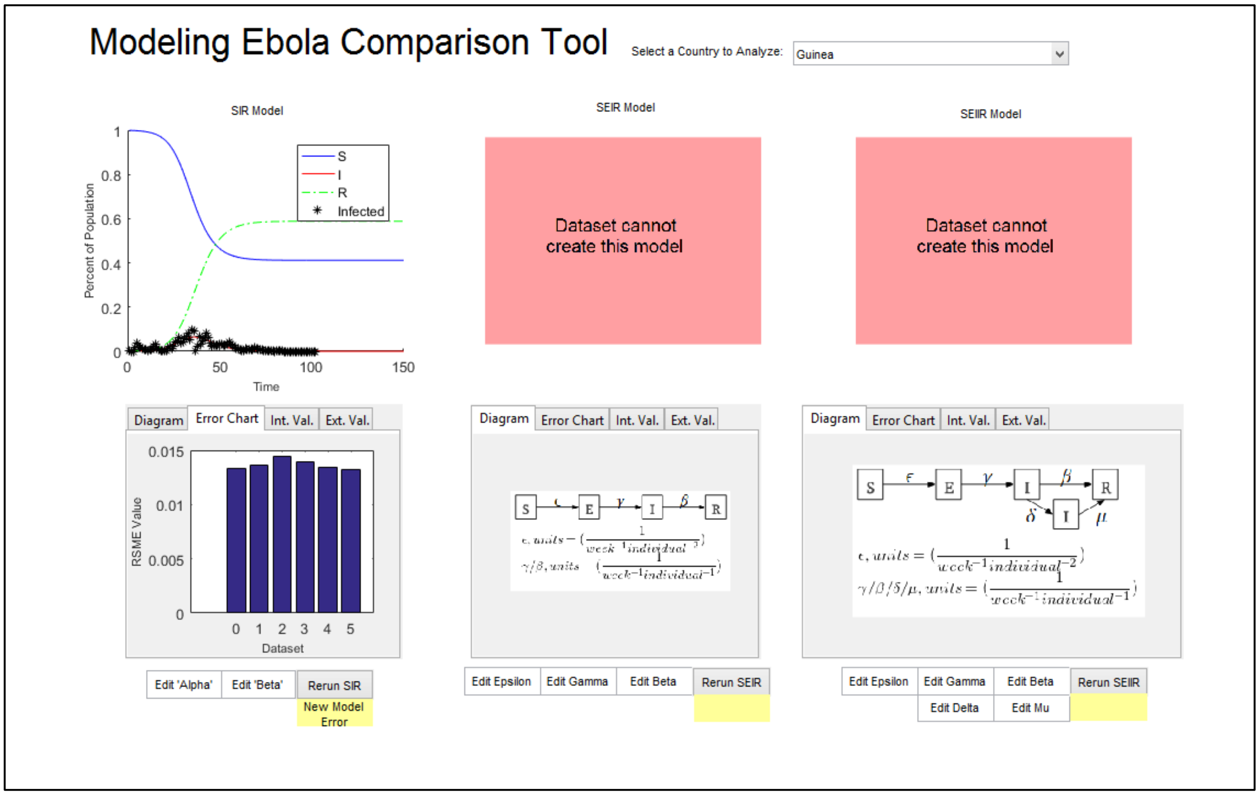
Figure A1: Graphical User Interface (GUI) Screenshot for Each Country of Interest



Modeling Ebola Comparison Tool

Select a Country to Analyze: Liberia

	Epsil...	Gam...	Beta	Delta	Mu	RMSE
5	0.7977	0.6902	0.0972	0.0977	0.4661	0.0454
4	0.7901	0.6858	0.1024	0.1029	0.4817	0.0451
3	0.7913	0.6872	0.1008	0.1029	0.4847	0.0391
2	0.7933	0.6881	0.1006	0.1016	0.4890	0.0378
1	0.7934	0.6914	0.1000	0.1002	0.4883	0.0411
0	0.7953	0.6916	0.0984	0.0994	0.4783	0.0324
Avg	0.7935	0.6890	0.0999	0.1008	0.4814	0.0402

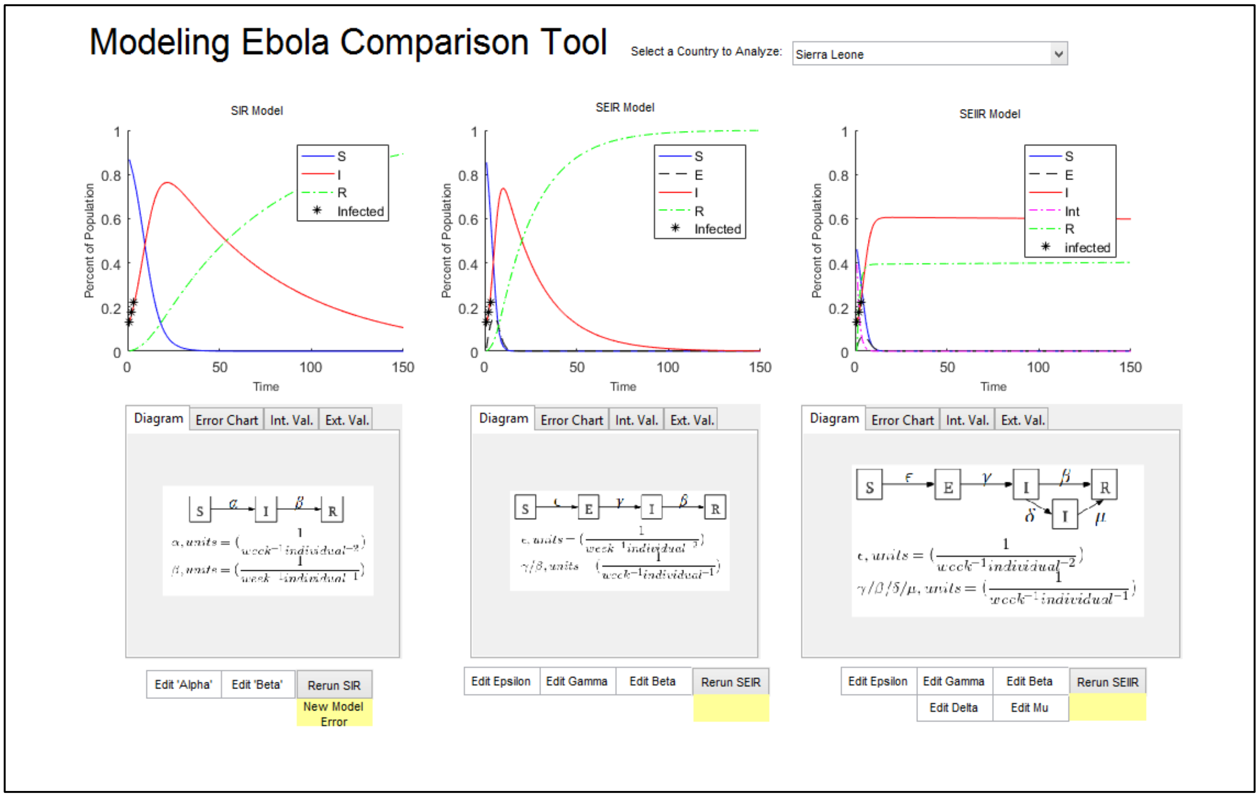


Modeling Ebola Comparison Tool

Select a Country to Analyze: Guinea

Dataset cannot create this model

Dataset cannot create this model



Modeling Ebola Comparison Tool

Select a Country to Analyze: Sierra Leone

$\alpha, units = (\frac{1}{week^{-1} individual^{-2}})$

$\beta, units = (\frac{1}{week^{-1} individual^{-1}})$

$\epsilon, units = (\frac{1}{week^{-1} individual^{-2}})$

$\gamma/\beta, units = (\frac{1}{week^{-1} individual^{-1}})$

$\epsilon, units = (\frac{1}{week^{-1} individual^{-2}})$

$\gamma/\beta/\delta/\mu, units = (\frac{1}{week^{-1} individual^{-1}})$

New Model Error