# Quantifying the Impact of Contact Tracing on Ebola Spreading

by

Narges Montazeri Shahtori

B.S., Isfahan University of Technology, IRAN, 2015

---

A THESIS

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Faryad Darabi Sahneh

# Copyright

Narges Montazeri Shahtori

2016

# Abstract

Recent experience of Ebola outbreak of 2014 highlighted the importance of immediate response to impede Ebola transmission at its very early stage. To this aim, efficient and effective allocation of limited resources is crucial. Among standard interventions is the practice of following up with physical contacts of individuals diagnosed with Ebola virus disease — known as contact tracing. In an effort to objectively understand the effect of possible contact tracing protocols, we explicitly develop a model of Ebola transmission incorporating contact tracing. Our modeling framework has several features to suit early–stage Ebola transmission: 1) the network model is patient–centric because when number of infected cases are small only the myopic networks of infected individuals matter and the rest of possible social contacts are irrelevant, 2) the Ebola disease model is individual–based and stochastic because at the early stages of spread, random fluctuations are significant and must be captured appropriately, 3) the contact tracing model is parameterizable to analyze the impact of critical aspects of contact tracing protocols.

Notably, we propose an activity driven network approach to contact tracing, and develop a Monte-Carlo method to compute the basic reproductive number of the disease spread in different scenarios. Exhaustive simulation experiments suggest that while contact tracing is important in stopping the Ebola spread, it does not need to be done too urgently. This result is due to rather long incubation period of Ebola disease infection. However, immediate hospitalization of infected cases is crucial and requires the most attention and resource allocation.

Moreover, to investigate the impact of mitigation strategies in the 2014 Ebola outbreak, we consider reported data in Guinea, one the three West Africa countries that had experienced the Ebola virus disease outbreak. We formulate a multivariate sequential Monte Carlo filter that utilizes mechanistic models for Ebola virus propagation to simultaneously estimate the disease progression states and the model parameters according to reported incidence data streams. This

method has the advantage of performing the inference online as the new data becomes available and estimating the evolution of the basic reproductive ratio $R_0(t)$ throughout the Ebola outbreak. Our analysis identifies a peak in the basic reproductive ratio close to the time of Ebola cases reports in Europe and the USA.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgments

Foremost, my sincere appreciation goes to my academic advisor, Dr. Faryad Darabi Sahneh, for his continuous support, encouragement and guidance during my M.S. studies. I would like to especially thank Dr. Caterina Scoglio for her valuable comments and support throughout the course of my M.S. work at the Kansas State University. It was such a pleasure and honor for me to be a part of NetSE group. My warm thanks go to my labmates for their collaboration and friendship. I would like to thank my thesis committee members, Dr. Don Gruenbacheri and also Dr. Caterina Scoglio for all of their guidance through this process.

There are not enough words to thank my family and especially my beloved husband, Hamid, for their unconditional and unending support throughout my work.

I would like to thank Dr. Nora Bello and Joan Saldaña for fruitful discussions and suggestion. Ala Fard for organizing Ebola incidence data for my analysis.

This work has been supported in part by the National Science Foundation under Grant No. SCH:1513639

# Chapter 1

# Introduction

## 1.1 Mathematical Modeling of Disease Spreading

Spreading of an infectious disease is a complex event with many interacting variables. Mathematical model of infectious disease offers a mechanistic framework to study the diffusion of infections within human and animal populations, providing deep insight into epidemics' dynamics[28]. Mathematical models allow us to deduce from current information about spreading of an infectious disease, to anticipate the future and quantify the uncertainty in the predictions. One of the primary tools to analyze severity of infectious disease and predict the disease spreading is the compartmental model. Compartmental model is a mathematical frameworks that can capture the main features of disease spreading such as transmission probabilities, transmission rates and prevalence of new infections[8]. Compartmental models classify the population into different compartments. All individuals in a compartment have common attributes. For instance all the healthy individuals usually are assumed as susceptible. Collaboration of these compartments is often based on transmission process of a virus in a host population. In the simplest model, susceptible–infected–susceptible, *SIS*, population is divided into two states based on the health of people. Healthy individuals are susceptible to the infection of the pathogen, denoted by $S$ and infected individuals are those who are infected by the pathogen, denoted by $I$. Compartmental model could be analyzed through

deterministic version using deferential equations or in stochastic framework. Deterministic compartmental models are assumed that a population is homogenous and the transmission process in a host population is determined by the rules describe in the compartmental model. However, in stochastic version, there is a distribution of possible transmission.

The basic reproductive ratio, $R_0$, is one of the most relevant descriptors that helps public health authorities have a quantitative understanding of the severity of the disease outbreak and its time projection. Therefore, it is crucial to estimate the basic reproductive number and other relevant descriptors early to have a quantitative measurement of severity of the disease outbreak. The most common definition of basic reproductive number is the expected number of secondary infections produced by a typical single infection in an entirely susceptible population[17]. Another definition of basic reproductive number is the expected number of secondary infections over all possible initial infections during their infectious period[18 10]. Thus, $R_0$ is a dimensionless value that represents the average number of additional susceptible people to whom an infected person passes the disease before he/she recovers[26]. The basic reproductive ratio is a threshold condition for epidemics as $R_0 = 1$ separates the increments or decrements of the newly infected[26].

The Ebola virus, commonly known as Ebola, causes a serious illness which is fatal if untreated in most cases[2]. It is transmitted via direct contact with blood, secretions, organs, or other bodily fluids of infected individuals and causes Ebola disease[24]. Ebola cases can transmit the virus to their contacts after becoming symptomatic[11]. The incubation period, the time interval from infection with Ebola virus to the onset of symptoms, is between two to twenty one days[2]. First symptoms of Ebola include the sudden onset of fever, fatigue, muscle pain, headache, and sore throat[2]. Since December 2013, West Africa has experienced the most widespread Ebola disease epidemic in the history with more than $28,000$ reported cases[27]. Secondary infections have been also reported in some European countries and United states[2]. Ebola hemorrhagic fever is considered a highly infectious and lethal disease, raising serious concerns about the public health globally. Although specific pharmaceutical treatment or vaccines are not available for Ebola virus, efforts to stop the spread of Ebola virus included intervention strategies such as surveillance, quarantin-

ing suspected cases, and education of hospital workers in contact with Ebola patients[6]. Contact tracing has proven to be the most successful mitigation strategy to track and suppress Ebola transmission chain, among all intervention strategies. The objective of contact tracing is to identify and monitor individuals who have been exposed to an infectious individual[4]. This procedure allows for isolation of contacts of an infectious individual promptly after he/she becomes symptomatic. In particular, during the early stages of infection, contact tracing seems to be an appropriate approach because Ebola is not a fast growing disease[13]. Side by side of these efforts, mathematical and computational epidemic models were developed and implemented with the aim of predicting newly infected cases as well as evaluating mitigation strategies. In this work, we develop a network–based stochastic modeling framework for Ebola contagious process. The goal of this modeling is to identify local contact network of an infected individual such as household, hospital and, workplace. This framework allows synthesizing different scenarios compatible with Ebola spreading in 2014. To this end, we employ stochastic version of the discrete–time expression of the susceptible, exposed, infected and removed, *SEIR*–based, compartmental model which is compatible with our understanding of Ebola virus epidemiology.

## 1.2 Motivation and Problem Statement

Ultimate goal of most proposed compartmental models for a disease spreading such as Ebola propagation is to predict the future evolution of Ebola virus disease or to estimate $R_0$. These models are mostly suitable for a large host population where the number of cases is high and random effects are diminished at the overall population level. Minimum random effects in a large population could explain why models for an outbreak are mostly reliable in prediction. However, in pre–outbreak stage, when new infections are rare and is likely to have a low prevalence, the accuracy of large–scale mathematical infection model dramatically declines. When numbers of infected cases are small, the network model is patient–centric and only local networks of infected individuals matter. Therefore, before epidemic phase, the localized contacts of infected individuals

influence the future of disease spreading and, this phase is crucial for public health authorities. A proper modeling approach of a disease spreading before epidemic phase needs to consider highly structured contact network of host population and also, assess the inherent uncertainties of data. These challenges make a reliable prediction of local contagious process impossible. However, the proper models could assess effectiveness of contact tracing. The goal of this work is to evaluate risk detection probabilities of contact tracing efforts for Ebola before epidemic phase in potential scenarios and come up with rules and regulations that help public health authorities to suppress those disease spreading which have similar epidemiological characteristic with Ebola. Moreover, to study the efficiency of implementation of contact tracing accompanied by other mitigation strategies in an Ebola virus epidemic, we investigate the effectiveness of implemented mitigation protocols in one of the West Africa countries that had experienced Ebola virus epidemic. In this work, following questions is addressed:

I. Defining quantitative measurements to assess effectiveness of contact tracing.

   We define contact tracing cost as number of detected individuals who have contact with infections but were not infected. Missed–detection probability denotes the probability that a secondary infected individual is not detected before transmitting the virus to others.

II. What is the impact of a delay in implementation of contact tracing on total number of infected individuals at the end of disease evolution?

   We consider two kind of delay:1) Global delay: when there is a delay in starting contact tracing 2) Local delay: when there is a delay in identification of local contact network of an infected individual.

III. What is the impact of a delay in implementation of contact tracing on $R_0$, contact tracing cost and, missed–detection probability?

IV. When is contact tracing not fully effective? Which mitigation strategies might be more effective?

To answer these questions, we develop a model for early–stage Ebola introduction explicitly incorporating contact tracing. To explore suitable recommendation for different scenarios, we We also, develop a Monte–Carlo method to compute defined quantitative measurements.

V. How effective were mitigation protocols to suppress Ebola propagation in 2014?

To assess effectiveness of mitigation strategies such as contact tracing strategy in a real scenario of Ebola introduction, we consider Guinea, one of West Africa countries that had experienced Ebola outbreak.

## 1.3 Thesis Overview

In this thesis, we implement an *SEIR*–based model for Ebola propagation to estimate the disease states and make inference about the basic reproductive ratio and other quantitative measures. We study the effect of likely contact tracing protocols objectively to find quantitative measures for effectiveness of contact tracing in early–stage of Ebola virus disease spreading. We develop a model for early–stage Ebola transmission explicitly incorporating contact tracing. Then, to investigate the impact of implementation of contact tracing in a real scenario, we use reported data in Guinea, one of the three West Africa countries that had experienced the Ebola outbreak. To this end, we employ a sequential Monte Carlo (*SMC*) filter, an online inference method that allows simultaneous state and parameter estimation with improved accuracy as new streaming data becomes available. The proposed SMC setting allows simultaneous estimation of the number of individuals at different infection stages as well as the parameters of our mechanistic epidemic model, providing posterior distributions of interest. Then, we use the estimated values of the Ebola epidemic model parameters to determine the value of $R_0(t)$. In chapter 2, we introduce a compartmental model for early–stage Ebola transmission incorporating contact tracing to answer problems *I–IV*. In this chapter, we review activity driven network, a heterogeneous network, and then, we implement activity driven network to consider the inherent time–varying nature of transmission process in a host population. In particular, we synthesize Ebola contagion process using the proposed

compartmental model for a temporal network that is obtained with activity driven network generative process. We also, develop a Monte–Carlo method to compute the basic reproductive number of the disease spreading in different scenarios. In chapter 3, we discuss some background on sequential Monte Carlo filter and provide a compartmental model which is compatible with Ebola phenomenological in Guinea to explore answer of question $V$. In particular, we provide a specific setting for SMC filter based on proposed compartmental model that allows to estimate number of new infections and the parameters of the model simultaneously as a distribution. Notably, we estimate the disease states and make inference about the basic reproductive ratio through time.

Finally, chapter 4 summarizes and concludes this dissertation and sets the subjects for future research.

## 1.4 Contribution

Below is the summary of main contribution of this work:

- Developed a stochastic and patient–centric model for early–stage of Ebola propagation incorporating contact tracing (chapter 2) in a heterogeneous network that could capture the inherent time–varying nature of contagion process in a host population.
- Developed a Monte–Carlo method to compute the basic reproductive number (chapter 2) of the disease spread in a heterogeneous network.
- Studied the impact of critical aspect of contact tracing strategy on Ebola disease spreading in a heterogeneous network (chapter 2).
- Developed a sequential Monte–Carlo filtering setting capable of performing the online inference as the new data becomes available (chapter 3).
- Studied the impact of implemented mitigation protocols on the evolution of $R_0$ for Ebola outbreak in Guinea though time(chapter 3).

# Chapter 2

# Quantifying the Impact of Early–Stage Contact Tracing on Controlling Ebola Diffusion

## 2.1 Introduction

Researchers have attempted to study the impact of contact tracing on disease spreading and find a relationship between effectiveness of contact tracing and basic reproductive number of disease. Eames and Keeling propose a formula to correlate the effectiveness of contact tracing and basic reproductive ratio by using detailed pairwise equations for susceptible–infected- removed (*SIR*) based model[13]. Other researchers have studied the impact of contact pattern on efficacy of contact tracing. Previous works study the impact of network structure such as clustering on the effectiveness of contact tracing[28]. For instance, Kiss *et al.* conclude that when in a random network which has a short incubation period, average number of links per nodes increases, effectiveness of contact tracing decreases dramatically[22]. In[23], Klinkenberg *et al.* deduce that knowledge of the initial time of contact tracing and iterative tracing improves this strategy. Researchers also have attempted to analyze the impact of intervention strategies on recent Ebola spreading using

mean–field compartment model, which can be either stochastic or deterministic in nature. Browne *et al.* use deterministic version of compartment model and separate infected individuals into different compartments based on whether they are hospitalized or unreported. They evaluate the impact of relevant epidemiological properties on contact tracing efficiency and present a formula to determine the number of contacts for bringing the effective reproduction below one[4]. Rizzo *et al.* in[31], adopt susceptible–expose–infected–removed (SEIR) compartmental model adding additional compartments related to hospitalized and dead people who had traditional funerals. Then based on this model, they propose a mathematical model based on an activity driven network whose contact network and intervention policies could vary in time. Finally, they implement the model as a predictive tool to imitate the dynamic of of Ebola virus spread in Liberia. Rivers *et al.* use deterministic version of compartmental model to fit time series of reported Ebola cases. They validate the model using least ? square optimization technique and model five scenarios of interventions strategies that could be implemented. Then, the likely impact on the epidemic evolution is examined[30]. Here, we study the effect of likely contact tracing protocols objectively to find quantitative measures for effectiveness of contact tracing. We develop a model for early–stage Ebola transmission explicitly incorporating contact tracing. Our modeling framework has several features to suit early–stage Ebola transmission: 1) the network model is patient–centric because when number of infected cases are small only the myopic networks of infected individuals matter and the rest of possible social contacts are irrelevant, 2) the Ebola disease model is individual–based and stochastic because at the early stages of spread random fluctuations are significant and must be captured appropriately, and 3) the contact tracing model is parameterizable to analyze the impact of critical aspects of contact tracing protocols. The model is built on susceptible–exposed–infected–hospitalized–removed (*SEIHR*) model where susceptible, exposed and infected individuals respectively become monitored–susceptible, monitored–exposed and monitored–infected upon contact tracing. Notably, we propose an activity driven network approach to contact tracing. We implement Activity driven network to consider the inherent time–varying nature of contagion process in a host population, such as variations in connectivity pattern

of contacts. Activity driven network is a heterogeneous network which has a random and memory less process that employs a time invariant function to characterize interactions between nodes, and describes a contact network that evolves over time[29]. At each time step, an activity firing rate is assigned to each node so that a proportion of nodes create new links with other nodes and the contagious process develops over new links. We also, develop a Monte–Carlo method to compute the basic reproductive number of the disease spread in different scenarios. Exhaustive simulation experiments suggest that while contact tracing is important in stopping the Ebola spread, it does not need to be done too urgently. This result is due to the rather long incubation period of Ebola disease infection. However, immediate hospitalization of infected cases is crucial and requires the most attention and resource allocation. The remainder of this chapter is organized as follows. In section 2.2, we propose a compartmental model for Ebola transmission incorporating contact tracing, discuss an overview of activity driven network (*ADN*) and explain activity driven network model based on the proposed compartmental model. Section 2.3 presents the Monte–Carlo method to compute basic reproductive number in a heterogeneous network. In section 2.4, we define true positive and false positive ratios based on the proposed model to plot receiver operating characteristic curve. Section 2.5 summarizes the main results of this article. Section 2.6 concludes the chapter and sets the subjects for future research.

## 2.2 Mathematical Modeling of ebola disease spreading incorporating contact tracing

Spreading of an infectious disease is a complex event with many interacting variables. One of the primary tools to analyze and predict the disease diffusion as well as severity of infectious disease is the compartmental model. Compartmental model is a mathematical frameworks that can capture the main features of disease spreading such as transmission probabilities and transmission rates[8]. In this work, we employthe discrete–time expression of the susceptible, exposed, infected, hospitalized and removed (*SEIHR*) compartmental model. Such model is compatible with the

epidemiology of Ebola disease virus.

### 2.2.1 Compartmental Model

*SEIHR* model is similar to susceptible, exposed, infected and removed (*SEIR*) model with an additional compartment $H$, where $H$ stands for hospitalized. Each compartment variables denotes a fraction of individuals which belongs to one of the following compartments: susceptible ($S$), exposed ($E$), infected ($I$), hospitalized ($H$) and, removed ($R$). In this model, when a susceptible individual has contact with an infected individual, the susceptible individual goes to $E$ with probability $\beta$. An exposed person before proceeding to compartment $I$, undergoes an incubation period with average length of $\frac{1}{\bar{\lambda}}$ [32]. Then, the infectious individual moves to $H$ compartment with probability $\gamma$ and finally, a hospitalized person enters $R$ state with probability $\delta$. To evaluate the impact of public health control strategies such as contact tracing, for detection of new Ebola patients, we develop a network–based model, adding additional compartments to SEIHR model. We construct our model based on contact tracing implementation guideline published by CDC. In the guideline, it is mentioned that when an infectious individual enters hospital and his/her laboratory results become positive, any persons who has contact with him/her in the last 21 days should be traced [1]. Therefore, in our proposed compartmental model, we assume that when an infectious individual enters $H$ compartment with probability $\gamma$, any person who has been exposed to the infected individual should be identified and followed up for 21 days. We classify transmissions between different epidemiological compartments of the proposed model into two groups: Node–based transmissions and edge–based transmissions. In node–based transmissions, a node moves from one epidemiological compartment to another individually, and, its transmission does not depend on its neighbors' state. Contrary to the nodal transmission, edge–based transmission is only dependent on a node's neighbors' state. Based on these definitions, we introduce the transmission process of the proposed model as follows:

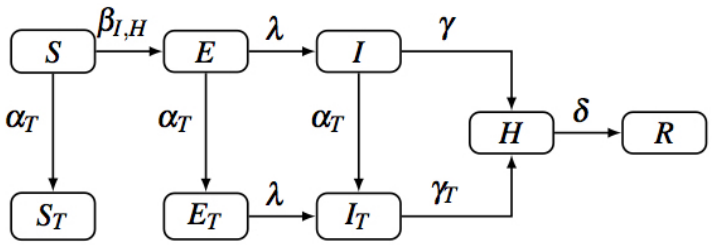


Figure 2.1: Schematic of transmission process

**Edge–based Transmission** When susceptible individuals have contacts with an infectious or hospitalized one, they move to exposed compartments with probability $\beta$.

**Node–based Transmission** Similar to *SEIHR* model, an exposed individual goes through an average incubation period of $(1/\bar{\lambda})$ before moving to infectious compartment ($I$). Infectious individuals progress to hospitalized compartment, $H$, with average delay of $\gamma^{-1}$ and their susceptible, exposed and, infectious contacts moves to $S_T$, $E_T$ and $I_T$ compartments, respectively with an average identification delay period of $1/\bar{\alpha_T}$. Since a portion of the the contacts of the infected individual might be inaccessible or a portion of them might be unwilling to report all their contacts immediately, we assume that there is a delay to identify contacts of an infected individual. An exposed individual who is traced, (hereafter referred to as $E_T$ compartment) undergoes an average incubation period $1/\bar{\lambda}$ days before progressing to infectious compartment where infections are traced ($I_T$). Infectious individual who is traced enters to hospitalized compartment with probability $\gamma_T$ where $\gamma_T > \gamma$. Finally, hospitalized individuals move to $R$ compartments or removed compartment with probability $\delta$.

A schematic of epidemiological stochastic transmission process of the proposed model is depicted in figure 2.1.

Based on the proposed model, we develop a quantitative approach is developed to measure effectiveness of contact tracing implementation. To asses the impact of contact tracing protocols in Ebola disease spreading before epidemic phase, we propose two measurements: Missed–

detection probability and contact tracing cost. **Definition 1:** To asses risk detection capabilities of contact tracing efforts for Ebola virus, we introduce missed–detection probability. Missed–detection probability denotes the probability that a secondary infected individual is not detected before transmitting the virus to others. Based on our model, we propose the missed–detection probability as follows: $\frac{N_{E\to I}}{N_{E\to E_T}+N_{E\to I}}$, where $N_*$ represents total number of individuals who move from one compartments to another one. **Definition 2:** The aim of contact tracing is to detect secondary infections among all of the contacts of an infected person. However, a large proportion of an infectious individual's contacts are susceptible. We define contact tracing cost to estimate the associated cost in different scenarios. Contact tracing cost denotes the number of detected individuals who had contact with infections but were not infected. Based on proposed model, we define contact tracing cost as: $Cost = N_{S\to S_T}$, where $N_{S\to S_T}$ represents total number of susceptible individuals who moves to $S_T$ compartment.

### 2.2.2 Activity Driven Network

Disease contagious process and network structure are two important elements that can have a significant impact on disease spreading[13]. Many intervention strategies such as contact tracing, target strategy and, egocentric strategy aim at controlling contagious process based on interactions between individuals in a social network[25 4]. In particular, contact tracing strategy or identification of individuals who have contact with infections, is fundamentally linked to potential transmission paths in the network[22 13]. The goal of contact tracing is to identify all the potential routes in the network and isolate all new infected individuals, before they become infectious[22]. Here, we implement activity driven network ($ADN$) to capture interactions between nodes in a network over a specific period of time and also, assess effectiveness of contact tracing strategy for a temporal network based on Ebola contagious process. Activity driven network is a random and memoryless process which can capture structural features of a network such as evolution of contact patterns over time[29]. Activity driven network considers an activity firing rate $a_i$ for each node which is the probability of establishing links with other nodes per unit of time[25] . Activity firing rate are

assigned according to a probability distribution $F(a)$ which can describe dynamic of the network and its corresponding structure[25]. Typically $F(a)$ is a heavy tail density function; $F(a) \propto a^{-c}$, where $2 \leq c \leq 3$ and $a \in [\varepsilon, 1]$[29]. At each time step $\Delta t$, an active node generates $m$ links with $m$ other nodes that are selected randomly. The generative network process in an increment time $\Delta t$ is determined as follows[25]:

- At time $t$, network $G_p(N,m)$ has a $N$ disconnected vertices.
- Each node $i$, with probability $p_i = a_i \eta_i \Delta t$ becomes active and generate $m$ links with others nodes. Where $\eta_.$ is a constant scaling factor.
- At time $t + \Delta t$ all the edges in network $G_t$ are removed.

### 2.2.3 ADN For Ebola Contagion Process

We implement activity driven network (*ADN*) to generate a random network at a time step $\Delta t$. At each time step $t$, we simulate Ebola contagion process using the proposed model for a temporal network that is obtained with generative network process of *ADN*. In generative network process, we assume that $\Delta t = 1$ and only those nodes which are in susceptible, infected, exposed or hospitalized compartments can generate $m$ links with other nodes with probability $p$. This algorithm is constructed based on *"CDC emergency guidelines of implementation and management of contact tracing"* for Ebola virus disease[1]. Based on the guideline, any person who has a potential exposure to a susceptible, probable and confirmed Ebola disease virus (EVD) case, should go under observation for 21 days[1]. To implement contact tracing strategy in a temporal network such as *ADN* at time $t$, from $t_k < t$, we capture all the nodes that become neighbors of an infectious node $j$, only if $T_{info} \leq t_k$. Then, we implement contact tracing at time $t_z = T_{CT}$, where, $T_{info} \leq t_k \leq T_{CT}$. Algorithm 1 sets the rules to produce and simulate Ebola virus spreading in a host population for time $1 \leq t \leq T$, where $T$ is the disease evolution time. In this algorithm, we use the concept of stochastic process to simulate Ebola contagious process. Furthermore, using *ADN*'s generative

**Algorithm 1** : ADN for Ebola Contagion Process

**if** $t \leq T$ **then**
  **if** $x_n \in \{S, E, I, H\}$ **then**
    $Sn \leftarrow n$
  **end if**Where $x_.$ represents state of node $n$.
  **Generate a network**: $G_p(N, m)$, *w.r.t* $(p, Sn)$
  **if** $t \geq T_{CT}$ **and** $\exists\, k | x_k \in \{H\}$ **then**
    Update state of neighbors of infected node $k$ at time $t$ and those nodes which had contact with node $k$ from the time it becomes infectious ($I_{Neigh}$), to their respective states with probability $\alpha$ ($S \rightarrow S_T$, $E \rightarrow E_T$, $I \rightarrow I_T$).
  **end if**
  **Contagion process**: **1. Edge–based transmission**: Find susceptible nodes which have contact with infectious nodes and update their state based on edge–based transmission rule.

  **if** $t \geq T_{info}$ **then**
    $I_{Neigh} \leftarrow n_j \mathbf{1}_{\{x_j = I\}}$
  **end if**, where $n_j$ represents node $n$ which has contact with node $j$ and, node $j$ is an infectious node. **2. Node–based transmission**: Update state of other nodes; $i | x_i \neq S$ based on node–based transmission rules.
  Remove nodes which belongs to $I_{Neigh}$ and are traced for 21 days but are not detected. Return these nodes from $S_T$ or $E_T$ compartments to $S$ and $E$ compartments respectively.
**end if**

network process, we could generate temporary links between nodes and we could capture all such links during the disease evolution.

## 2.3 Reproductive Number in Heterogeneous Network

The basic reproductive ratio, $R_0$, is a descriptor in mathematical modeling of infectious disease. It helps public health authorities to assess the risk of an outbreak in emergence of infectious disease[32]. Furthermore, early estimation of the basic reproductive number helps healthcare authorities to control disease outbreak. A general definition of basic reproductive number is the expected number of secondary infections over all possible initial infections during their infectious period[18 10]. Using general definition of reproductive number and characteristic of heterogeneous network, we propose a new definition for $R_0$ in a heterogeneous network. In heterogeneous net-

work, contact patterns tend to have a high variability in prevalence and so, besides high degree nodes, there are some low degree nodes which they may have no contacts with others[34]. Based on this characteristic, to compute $R_0$ in a heterogeneous network, we only consider those initial infected as possible initial infections if and only if they establish links with other nodes. Therefore, we define $R_0$ as follows: **Definition 3**: Basic reproductive number ($R_0$) is the expected number of secondary infectious cases over all initial infections that establish interaction with others during the infectious period. In particular, in activity driven network which is a heterogeneous network, $R_0$ is :

$$R_0 = \frac{\sum_{i=1}^{n} \mathbf{1}_{\{(x_i^t, x_i^0)=(I,S)\}}}{\sum_{j=1}^{n} \mathbf{1}_{\{((x_j^0,\ldots,x_j^t)=I),(i^t \leftrightarrow j^t)\}}} \tag{2.1}$$

where, $n$ is the number of node in heterogeneous network, $x_i^t$ is the state of node $i$ at time step $t \in \{0,\ldots,T\}$ and ($i^t \leftrightarrow j^t$ shows that node $i$ and node $j$ have interaction with each other at time step $t$. If $\chi$ is true, indicator function, $\mathbf{1}_\chi$, returns 1 and zero otherwise .

## 2.4 Receiver Operating Characteristic

A receiver operating characteristic or *ROC* curve is a fundamental method to illustrate the performance of a system such as separating true positive results from incorrect positive results in a test or comparing two alternative tasks[16]. In *ROC* curve, we plot sensitivity or true positive ratio ($TPR$) as a function of false negative ratio ($FPR$) or $1 - Specificity$. Sensitivity defines as the fraction of samples that are detected correctly and false positive ratio is the proportion of samples that are identified as positive, incorrectly.

Major aim of contact tracing is to identify secondary infectious individuals before they become infectious in order to halt chain of infection. We assume that positive samples are those secondary infectious individuals who have contact with infections during the disease evolution. Then, true positive in context of contact tracing strategy defines as those secondary infectious individuals who are traced. Similarly, false positive defines as those individuals who are not infected but traced as a

possible secondary infection. Therefore, based on different epidemiological compartments in our proposed model, we define $Sensitivity$ and $1-Specificity$ as follow: **Definition 4:** Sensitivity is the total number of exposed individuals who are traced over total number of exposed individuals during the disease evolution. Then,

$$TPR = Sensitivity = \frac{\sum_{i=1}^{n} x_i^{1:T} = E_T}{\sum_{i=1}^{n} x_i^{1:T} = E} \tag{2.2}$$

**Definition 5:** False positive ratio or $1-Specificity$ defines as total number of susceptible individuals who are traced over total number of susceptible individuals who have contact with infections during the disease evolution. Then,

$$FPR = 1 - Specificity = \frac{\sum_{i=1}^{n} x_i^{1:T} = S_T}{\sum_{i=1}^{n} x_{i,(i\to k)}^{1:T} = S} \tag{2.3}$$

where, $x_i^{1:T}$ is the state of node $i$ from $t = 1$ to $T$ which is the end of the disease evolution, and $(i \to k)$ shows that the susceptible node $i$ has contact with an infectious node $k$ . A point $(p_i, p_j)$ in the $ROC$ shows that with probability $p_i$ infected individuals who have contacts with an infection could be identified as infectious persons and with probability $1-p_j$ susceptible individuals who have contact with an infection could be identified as healthy persons.

## 2.5 Results

To generate a realization for Ebola disease spreading without any immunization strategy, parameters of our proposed model for contagious process are given in table 3.1. We assume that number of initial infected individuals is $I_0 = 2$ and each active node could generate only $m = 7$ links with other nodes, where total number of node is $N = 1000$. Usually when a person is sick, his/her social interactions reduce. Therefore, we assume that constant scaling factor of firing rate for a

hospitalized, infected and susceptible individual is as follows: $\eta_H << \eta_I < \eta_S$. Activity driven network's parameters are shown in table 2.2. To assess effectiveness of contact tracing in early

Table 2.1: Time–invariant parameters of Ebola contagion process

| Parameter | Value |
|---|---|
| transmission probability($\beta$) | 0.11 |
| latency ($\lambda$ ) | 0.095 |
| recovery/remove probability ($\delta$) | 0.1 |
| hospitalization probability in existence of contact tracing ($\gamma_T$) | 0.9 |
| hospitalization probability ($\gamma$ ) | 0.33 |

stage of epidemic for the generated scenario, we assume three different scenarios for contact tracing implementation. The first one is when we implement contact tracing from beginning ($T_{TC} = 1$), second scenario is when contact tracing is started at day $T_{TC} = 9$ and, third one is when contact tracing is implemented at day ($T_{TC} = 22$). In all of these scenarios, we assume that $T_{info} = 1$. To evaluate the effectiveness of contact tracing in a more realistic scenario, we consider five identification delay time, $\alpha^{-1} \in \{1,2,3,5,10,20\}$ in three scenarios of contact tracing implementation.

Table 2.2: Parameters of activity driven network generator

| Parameter | Value |
|---|---|
| $c$ | 2.2 |
| $m$ | 7 |
| $\eta_S$ | 2.2 |
| $\eta_I$ | 1.1 |
| $\eta_H$ | 0.005 |

### 2.5.1 Effectiveness of Contact Tracing

In figure 2.2, we plot, as a function of identification delay $\alpha^{-1}$, the ratio of cases; $\frac{(I+H)_\infty^\alpha}{(I+H)_\infty^0}$, for three scenarios. Where $(I+H)_\infty^0$ is the mean of infected and hospitalized individuals at the

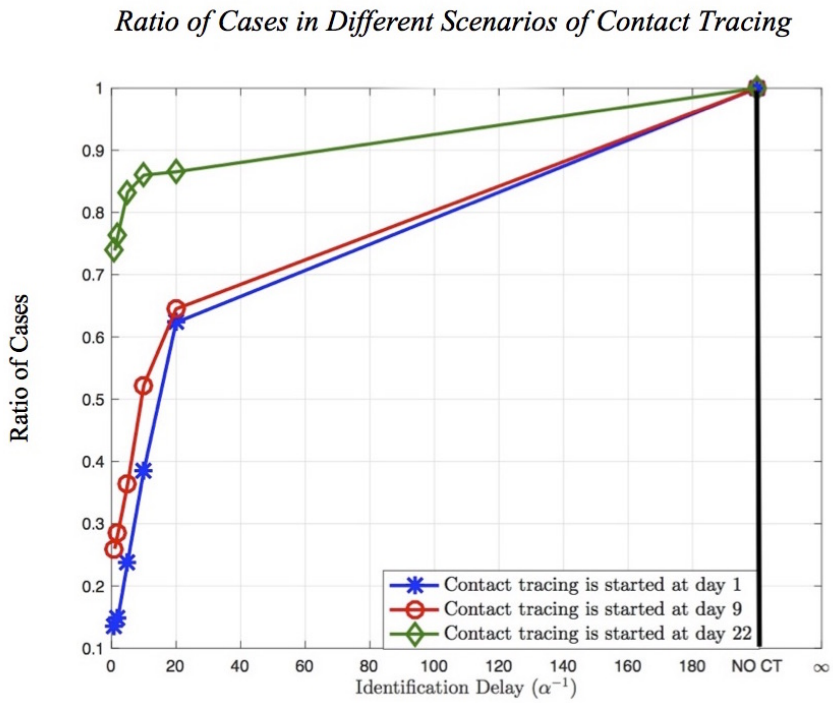


Figure 2.2: Comparison of ratio of cases as a function of $\alpha^{-1}$ .

end of the disease evolution, when contact tracing strategy is not implemented. $(I+H)_{\infty}^{\alpha}$ shows average number of infected and hospitalized individuals at the end of the disease evolution for an identification delay of $\alpha$ and, $NO_{CT}$ shows when contact tracing or other immunization protocols are not implemented. Figure 2.2 clearly shows that, contact tracing is more effective in first and second scenario when identification delay, $\alpha^{-1}$, is less than 10 days. In third scenario, we could still observe a reduction in the ratio $\frac{(I+H)_{\infty}^{\alpha}}{(I+H)_{\infty}^{0}}$, only if contacts of infections could be identified immediately. In figure 2.3, we study the impact of hospitalization period ($\gamma^{-1}$), which is the time it takes to hospitalize an infectious individual, without contact tracing strategy on ratio of cases. It shows the importance of immediate access to hospitals for an infectious individuals. Comparing figure 2.3 and figure 2.2 clearly shows that immediate hospitalization is even more efficient than immediate and accurate contact tracing protocols.

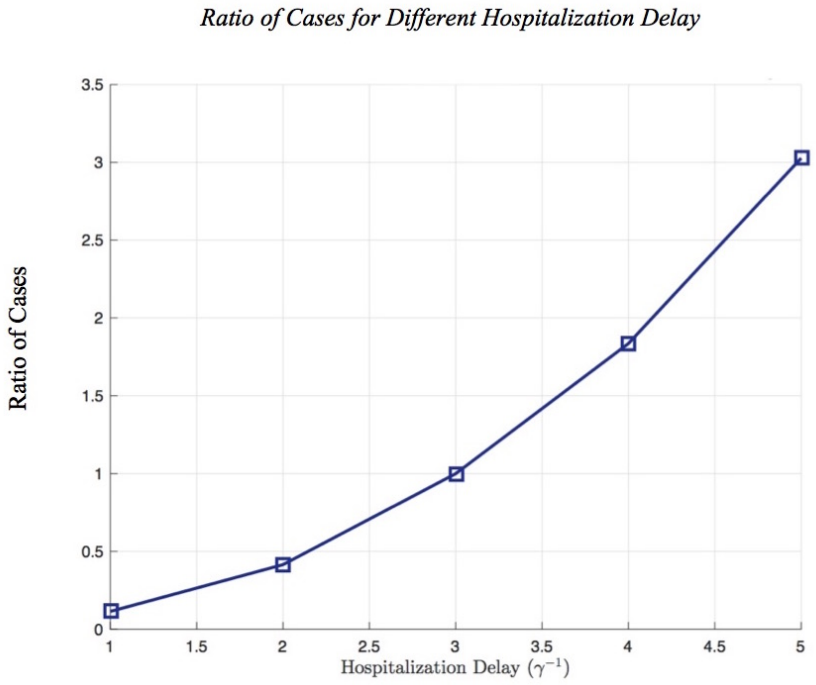


Figure 2.3: Comparison of ratio of cases as a function of $\gamma^{-1}$.

### 2.5.2 Contact Tracing Performance

To evaluate performance of contact tracing on Ebola contagious process, we employ *ROC* curve. To plot *ROC* curve, we compute $sensitivity$ and $1-specificity$ in each iteration from equations 2.2 and 2.3. Then, we compute average of these two ratios in $10,000$ iterations. Figure 2.4 shows $sensitivity$ as a function of $1-specificity$ for 5 identification delays ($\alpha^{-1}$) in three contact tracing implementation scenarios. In each scenario, we plot $sensitivity$ as a function of $1-specificity$ . We project 3–D *ROC* curve to get 2–D *ROC* curve, as shown in figure 2.5. In the first scenario when $\alpha^{-1}$ increases, the probability to identify secondary infected individuals produced by an infectious individual decreases. In second and third scenarios this reduction is more significant. Missed–detection probability is the number of exposed individuals who are not traced as defined previously. Therefore, missed–detection is equal to $1-Sensitivity$. Figure 2.5 clearly shows increment of identification delay period can increases the missed–detection probability. Furthermore, immediate implementation of contact tracing protocols can help to reduce the probability of miss-

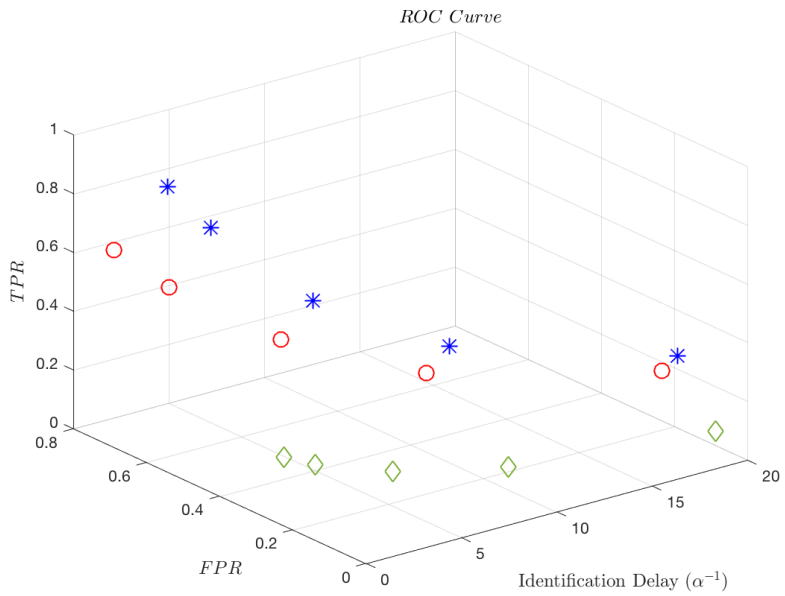


Figure 2.4: 3–D *ROC* curve of contact tracing with 5 identification delay implemented in three different scenarios.

ing secondary infections. Another significant observation is that in all scenarios, we could observe that the probability to identify healthy contacts of infectious individuals ($specificity$) increases, when $\alpha^{-1}$ decreases. So, these healthy individuals are assumed as a possible Ebola disease case, which results in high cost of contact tracing.

### 2.5.3 Basic Reproductive Number

Using equation 2.1, we compute basic reproductive number in an independent simulation. In figure 2.6, we calculate average of $R_0$s from $10,000$ independent simulations. It shows that for the same value of $\alpha$, contact tracing protocol allows a feasible reduction in value of $R_0$ in first scenarios. However, contact tracing strategy does not lead to $R_0 < 1$. In figure 2.7, basic reproductive number is shown as a function of hospitalization delay period $\gamma^{-1}$. It represents that reduction of $\gamma^{-1}$ results in reduction of $R_0$ and for $\gamma^{-1} \leq 2$, basic reproductive number becomes less than one.

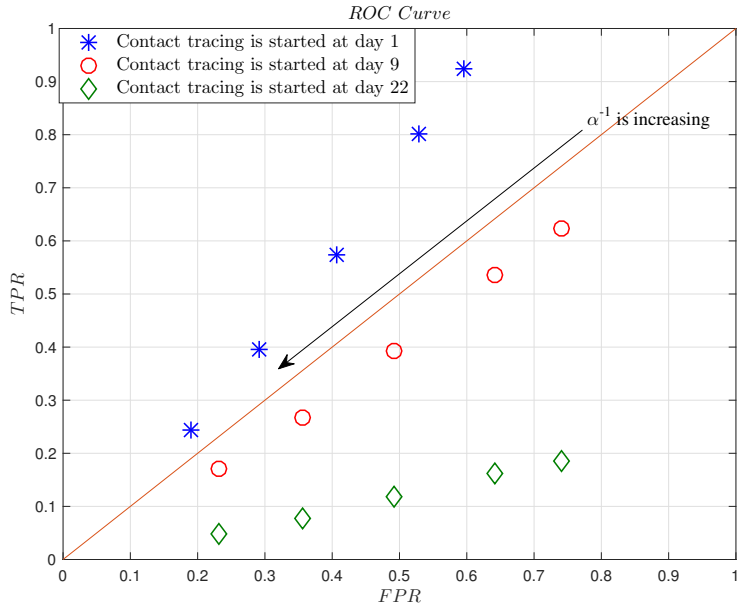


Figure 2.5: 2–D *ROC* curve of contact tracing implementation in three different scenarios.

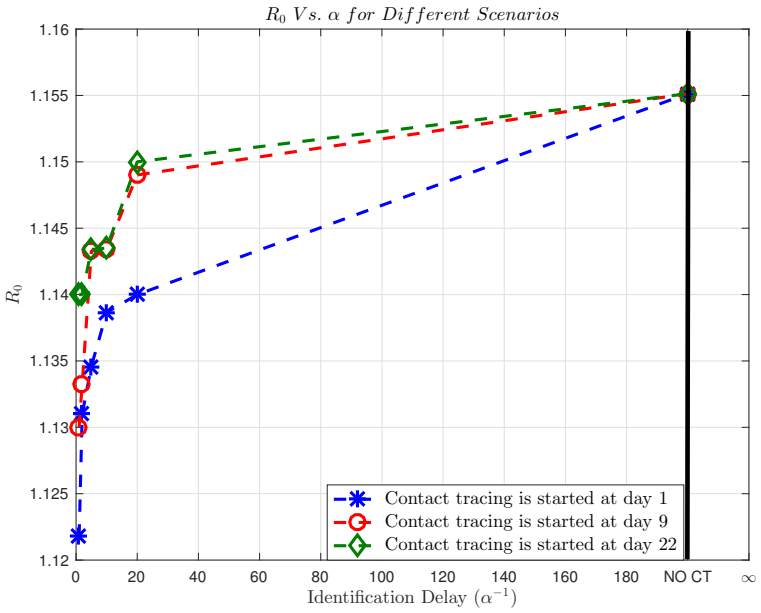


Figure 2.6: $R_0$ as a function of $\alpha^{-1}$ in three scenarios.

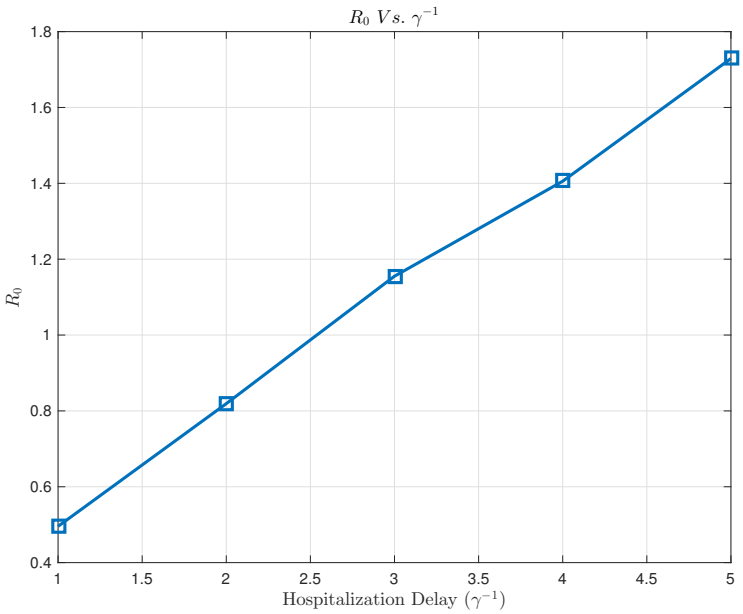


Figure 2.7: $R_0$ as a function of $\gamma^{-1}$.

## 2.6 Conclusion

Our analysis identifies that immediate implementation of contact tracing results in a large reduction in number of infectious individuals at the end of the disease evolution. Moreover, immediate identification of infections helps to reduce total number of infected individuals. However, implementation of only contact tracing protocols does not lead to a significant reduction in basic reproductive number. Immediate identification of secondary infections needs a collaboration between public health authorities and patients. Public health authorities could notify the host population as soon as possible and educate people about the disease. Therefore, an exposed individual who observes the symptoms of disease in himself/herself, can immediately go to hospital. It was shown that immediate hospitalization at early–stages of the disease spread results in significant reduction of $R_0$ and impede the transmission chain of the Ebola disease.

Although immediate and accurate contact tracing strategy could identify infected individuals and their infectious contact properly, it cannot distinguish healthy contacts from exposed contacts. Therefore, it may increase the financial burden on the public health authorities when many people

may need to be traced in such scenario, as happened in west African countries in 2014.

# Chapter 3

# Sequential Monte Carlo Filtering Estimation of Ebola Progression in West Africa

## 3.1 Introduction

Researchers have attempted to analyze the recent Ebola outbreak data, forecast the future of the outbreak, and estimate the basic reproductive ratio. Rivers *et al.*[30] used a deterministic compartmental model to fit time series of reported Ebola cases using least–square optimization, and provided forecasts according to implement Gillespies stochastic simulations. Browne *et al.*[4] proposed a compartment model separating the infectious population into reported/hospitalized or unreported compartments. Afterwards, they investigated the impact of contact tracing on reproductive ratio. Fishman *et al.*[14] utilized incidence decay with exponential adjustment (IDEA) method to compute reproductive ratio. Althaus[3] used an offline optimization algorithm to find parameters of the susceptible-exposed-infected-susceptible (SEIR) epidemic model that fits best to collected Ebola data during a fixed time period. The major shortcoming of such approaches is that they provide an offline inference of an outbreak that is inherently dynamic and parameters of model change dur-

ing disease evolution, so we need to keep tracking parameters when new data become available. Furthermore, since lots of factors such as intervention strategies could affect on parameters, we expect that the basic reproductive ratio changes during the disease evolution. Therefore we need techniques that are able to trace new data as they become available.

A few researchers have studied the evolution of $R_0(t)$ for Ebola progression. Towers *et al.*[37] estimated the basic reproduction ratio, $R_0(t)$, by fitting exponential regression models to small successive time intervals of the Ebola outbreak. Therefore, they obtained an estimate of temporal variations of the growth rate. Their application of regression models ignores the systemic epidemiological information of Ebola progression—as reflected in an SEIR model—and thus are more suitable for exploratory analysis of the incidence data. Furthermore, the scarcity of incidence data during short time intervals impacts the stability of regression model fitting. A more robust analysis should take advantage of the epidemiological knowledge of the dynamical system under study.

In this chapter, we use an SEIR-based model for Ebola propagation to estimate the disease states and make inference about the basic reproductive ratio through time. To this end, we implement a sequential Monte Carlo (SMC) filter, an online inference method that allows simultaneous state and parameter estimation with improved accuracy as new streaming data becomes available. Sequential Monte Carlo filter is particularly powerful for inference about epidemic models which are inherently nonlinear and involve numerous uncertainties. Specifically, our SMC setting allows simultaneous estimation of the number of individuals at different infection stages as well as the parameters of our mechanistic epidemic model, providing posterior distributions of interest. In SMC, the distribution of interest is estimated by a large number of random samples, termed particles, conditioned on the observations. A sampling mechanism propagates these particles[20]. Afterward, we use the estimated values of the Ebola epidemic model parameters to determine the value of $R_0(t)$.

Compared with existing studies on the recent Ebola epidemics in West Africa, our approach has the advantage of performing the inference online as the new data becomes available and es-

timates the evolution of basic reproductive rate $R_0(t)$ of the Ebola outbreak through time. Interestingly, our analysis identifies a peak in the basic reproductive ratio close to the time when cases were reported in Europe and the USA.

The remainder of this chapter is organized as follows. Section 3.2 presents basic tools used in sequential Monte Carlo filter and discusses some background on epidemic modeling. Section 3.3 outlines our modified SEIR model for Ebola and explains the particle setup and our data set. Section 3.4 presents main results of this study. Section 3.5 concludes the chapter by a few suggestions for future research.

## 3.2 Background

### 3.2.1 Epidemic Modeling

Mathematical models of infectious disease offer a mechanistic framework to describe and study the spread of infections within human and animal populations, providing deep insight into their dynamics and suggesting practical strategies to reduce the severity of epidemics[28]. Here, we introduce a brief background discussing the susceptible-exposed-infected-recovered (SEIR) model which is compatible with our understanding of Ebola virus epidemiology.

In the SEIR model, each individual belongs to one of the susceptible, exposed, infected, or removed/recovered compartments. In this model, when a susceptible individual has contacts with an infected one, they enter the exposed compartment (E) at rate $\beta I$. Homogeneity of the population and how people have contact with each others in the host population is represented by a percentage factor $c$. After the incubation period of disease, with average length of $1/\lambda$, they enter the infected compartment (I). Infectious individuals move to the recovered/removed compartment (R) at rate $\gamma$[21]. The basic compartmental SEIR model is[21]

$$\begin{aligned}
\frac{dS}{dt} &= -\beta cSI, \\
\frac{dE}{dt} &= \beta cSI - \lambda E, \\
\frac{dI}{dt} &= \lambda E - \gamma I, \\
\frac{dR}{dt} &= \gamma I.
\end{aligned} \tag{3.1}$$

In this compartmental SEIR, the size of host population is assumed to remain constant throughout the evolution time, i.e., $P = S + E + I + R$, and demographic effects are ignored.

An important mathematical descriptor of epidemics is the basic reproductive ratio. The basic reproductive number is defined as the expected number of secondary individuals produced by a typical single infected individual during its infectious period[18]. Thus, $R_0$ is a dimensionless value that represents the average number of additional susceptible people to whom an infected person passes the disease before he/she recovers[26]. For instance, if an infectious person passes the disease on three others on average during their infectious lifetime, then $R_0 = 3$, indicating that the number of new infectious individuals would increase with each generation, so we can expect to experience an epidemic outbreak. Conversely, if $R_0 < 1$ the disease will die out[26]. Thus, the basic reproductive ratio is a threshold condition for epidemics as $R_0 = 1$ separates the increments or decrements of the newly infected[26]. A more general definition of $R_0$ in mathematical epidemiology is *the average number of expected new infections over all possible infected types during the infectious period of a typical infected individual*[10]. Based on this definition, Diekmann *et al.*[10] proposed the *next generation matrix* method—a powerful technique for finding $R_0$ in complex epidemic models. Applying the next generation matrix technique to the SEIR model (3.1) finds $R_0 = \frac{\beta c}{\gamma}$[18].

### 3.2.2 Sequential Monte Carlo Filter

Sequential Monte Carlo (SMC) — or particle filter — refers to a class of statistical techniques that estimate unknown parameters, namely states in this context, as new streaming noisy observations becomes available[33]. In SMC, we iteratively sample from the posterior distribution of parameters until the parameters converge to stationary values[38]. This iterative sampling is updated using a stream of data, and as such, it enables us to modify our best guesses for the states according to actual observations. In the following, we explain the dynamic state–space model and estimation of posterior PDF briefly for the particle filters algorithm.

**Dynamic state–space model** The state–space model assumes the Markov property, i.e.,

$$Pr(x_k|x_{0:k-1}) = Pr(x_k|x_{k-1}). \tag{3.2}$$

and describes the distribution of the system state in the next step, as well as the observation, given the current state of the system. More rigorously, a state–space model is defined as[12,33]:

$$\begin{aligned} x_k &\sim p(x_k|x_{k-1}, \theta), \\ y_k &\sim p(y_k|x_k, \theta), \end{aligned} \tag{3.3}$$

where $p(\cdot)$ denotes the probability density function (PDF), $y_k$ represents the $k$th observation, $x_k$ represents system states corresponding to the $k$th observation, and $\theta$ represents parameters of the model. In a state–space model, $y_k$ depends only on $x_k$ and $\theta$, and $x_k$ depends solely on $x_{k-1}$ and $\theta$.

**Estimation of posterior PDF** Given the observation data $y_{1:k}$ up to time $k$, the ultimate goal is to define the posterior distribution, $p(x_k, \theta|y_{1:k})$, which describes the hidden state $x_k$ and parameters $\theta$ of the dynamical system. A sequential filtering uses a recursive formula relating the posterior distribution $p(x_k, \theta|y_{1:k})$ to $p(x_{k-1}, \theta|y_{1:k-1})$. In this way, given a prior distribution $p_0(x_0, \theta)$, one could iteratively find $p(x_k, \theta|y_{1:k})$. According to Bayes' theorem, the posterior probability density

function follows:

$$p(x_k, \theta_k | y_{1:k}) \propto p(y_k | x_k, \theta) p(x_k, \theta | y_{1:k-1}). \tag{3.4}$$

Furthermore, $p(x_k, \theta | y_{1:k-1})$ can be expressed as

$$p(x_k, \theta | y_{1:k-1}) = \int p(x_k | x_{k-1}, \theta) p(x_{k-1}, \theta | y_{1:k-1}) dx_k. \tag{3.5}$$

Instead of performing the integration in (3.5) explicitly, SMC utilizes the Monte Carlo method to approximates the posterior PDF using $J \gg 1$ samples, referred to as *particles*, through[33]:

$$p(x_k, \theta_k | y_{1:k}) \approx \sum_{i=1}^{J} w_k^{(i)} \mathbf{1}_{\{(x_k, \theta_k) = (x_k^{(i)}, \theta_k^{(i)})\}}, \tag{3.6}$$

where $\mathbf{1}_{\{\chi\}}$ is the indicator function returning 1 if $\chi$ is true and zero otherwise, $x_k^{(i)}$ is the state of particle $i$ at step $k$, and $w_k^{(i)}$ is its weight, which are iteratively updated to maximize the likelihood function. The approximation is more accurate if the number of particles is large. Particles' weights are normalized so that $\sum_{i=1}^{J} w_k^{(i)} = 1$.

Among particle filter techniques are the bootstrap filter, auxiliary particle filter, and kernel density particle filter. In this chapter, we use the latter, namely kernel density particle filter, due to its flexibility in modeling of nonlinear processes as explained in Section 3.3.2.

## 3.3 SMC for Ebola Epidemics

In this chapter, we employ the SEIR model as it is compatible with the epidemiology of the Ebola virus. In particular, we use a a discrete-time version since the data reports are per day.

### 3.3.1 Modeling of Ebola

The state variables $S_t$, $E_t$, $I_t$, and $R_t$ denote the fraction of people who are susceptible, exposed, infected and recovered or removed, at time step $t$, respectively, where each step is one day. For

our analysis, we use the discrete–time form of equation (3.1) with stochastic fluctuation and the following assumptions to modify the original SEIR model (3.1).

**Assumption 1:** Since the population is much greater than the number of infected cases of Ebola, we assume $S \simeq 1$. Therefore, the equation for the evolution of $E(t)$ in (1) simplifies to:

$$E_{t+1} = E_t + \beta c_t I_t - \lambda E_t. \tag{3.7}$$

**Assumption 2:** We assume that $c_t$, representing how well the population is mixed, changes due to intervention efforts to prevent the spread of Ebola such as social distancing and quarantining. Specifically, we assume $c_t$ is decreasing at rate $\alpha$, i.e.,

$$c_{t+1} = c_t - \alpha c_t. \tag{3.8}$$

This is a simplified assumption to account for different intervention strategies. We assumed that when the control measured are introduced and information regarding Ebola disease is disseminated, the transmission rate decays exponentially.

According to above assumptions and modifications to the SEIR model (3.1), we propose the following set of stochastic difference equations as our base epidemic model for the Ebola spread:

$$\begin{aligned}
c_{t+1} &= c_t - \alpha c_t + \xi_\alpha, \\
E_{t+1} &= E_t + \beta c_t I_t - \lambda E_t + \xi_\lambda - \xi_\beta, \\
I_{t+1} &= I_t + \lambda E_t - \gamma I_t - \xi_\lambda + \xi_\gamma, \\
R_{t+1} &= R_t + \gamma I_t - \xi_\gamma \\
D_{t+1} &= \varphi R_{t+1} = \varphi R_t + \varphi\gamma I_t - \varphi\xi_\gamma + \xi_\varphi.
\end{aligned} \tag{3.9}$$

In the above equations, $\xi_\chi$ where $\chi \in \{\alpha, \beta, \lambda, \gamma, \varphi\}$ is a random component, with zero mean and variance $\sqrt{\chi}/P$, where $P$ is the population size. The variance of noises are due to stochasticity

of the underlying process[33]. Each of these random components are assumed uncorrelated.

### 3.3.2 Filtering Setup

The kernel density particle filter method utilizes both bootstrap filter and auxiliary particle filter. In bootstrap filter, the probability density is estimated by a set of particles and at each round their weights are computed and those particles with small weights, are eliminated. After each round, the surviving particles produce new particles. The main problem with bootstrap filter is that weights might become too small for many particles and thus adversely affect the estimation accuracy. Auxiliary particle filter uses importance sampling to minimize number of particles with small weights, hence avoiding the degeneracy issue of bootstrap method[12]. The kernel particle filter not only avoids degeneracy, but also estimates the unknown parameters of the model simultaneously[5].

In this chapter, we formulate a filtering method based on the kernel density technique for intermittent observations. This is crucial for our estimation purpose because reports on Ebola cases become available at irregular times. We denote the observation times as $t_1, t_2, \ldots, t_k$ up to time $t$. Therefore, we compute the posterior distribution $p(x_{t_k}, \theta_k^{(i)} | y_{1:k})$ only when a new observation becomes available. We use the actual Ebola cases data in Guinea, one of the three major West Africa countries that experienced the Ebola outbreak, reported by the World Health Organization (WHO)[2].

#### Evolution Setup

According to our Ebola model in (3.9), the $x_t = [c_t, E_t, I_t, R_t, D_t]^T$ follows a normal distribution. Hence, the state–space model is

$$x_{t+1} | x_t, \theta \sim N_\Omega(g(x_t, \theta), Q(\theta)), \tag{3.10}$$

where

$$g(x_t,\theta)=\begin{bmatrix} c_t-\alpha c_t \\ E_t+\beta c_t I_t-\lambda E_t \\ I_t+\lambda E_t-\gamma I_t \\ R_t+\gamma I_t \\ \varphi I_t+\varphi\gamma I_t \end{bmatrix}, \tag{3.11}$$

is the mean and $Q(\theta)$ is the covariance matrix computed according to the description of random components of model (3.9):

$$Q(\theta)=\frac{1}{P^2}\begin{bmatrix} \alpha & 0 & 0 & 0 & 0 \\ 0 & \lambda+\beta & -\lambda & 0 & 0 \\ 0 & -\lambda & \lambda+\gamma & -\gamma & -\gamma\varphi \\ 0 & 0 & -\gamma & \gamma & \gamma\varphi \\ 0 & 0 & -\gamma\varphi & \gamma\varphi & \gamma\varphi^2 \end{bmatrix}. \tag{3.12}$$

Here, $\mathcal{N}_\Omega(\mu,\Sigma)$ represents the truncated normal distribution where $\Omega=\{(c_t,E_t,I_t,R_t,D_t): c_t,E_t,I_t,R_t,D_t\geq 0, E_t+I_t+R_t\leq 1\}$.

**Observation Setup**

Our WHO reports for Ebola in Guinea consists of cumulative cases and death numbers. Therefore, we do not have direct access to the number of 'active' infected population. In other words, the available observations are the number of dead people $D_t$ and the sum of active infected population $I_t$ and the total number of recovered/dead $R_t$, which we assume to have a log–normal distribution[33]. More precisely, we model the observations $Y$ as:

$$Y\sim\mathcal{N}(\mu_Y,\Sigma_Y), \tag{3.13}$$

where

$$Y = \begin{bmatrix} log(y_{I,k}) \\ log(y_{D,k}) \end{bmatrix}, \mu_Y = \begin{bmatrix} b_1(I_t + R_t)^{\zeta_1} \\ b_2(D_t)^{\zeta_2} \end{bmatrix}, \Sigma_Y \sim \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}. \tag{3.14}$$

In matrix $\mu_Y$, $\zeta$, $\sigma$ and $b$ for both observations are typically unknown, but we can predict these values by linear regression[35]. In particular, $b_1$ and $b_2$ are multiplicative constants, and $\zeta_1$ and $\zeta_2$ are power-law exponents which can be calculated based on the significant of dispersed of data points. Furthermore, variances $\sigma_1$ and $\sigma_2 \propto \frac{1}{\sqrt{P}}$[35].

**Kernel density particle filter**

Model (3.13) and (3.14) define the likelihood of observations $y_k$, given $x_{t_k}$ and $\theta$ — $p(y_k|x_{t_k},\theta)$. We apply kernel particle filter, to update and estimate $p(x_{t_{k+1}},\theta|y_{1:k+1})$. At the initial time step $k=1$, weights for all particles are equal to $J^{-1}$, and $\theta_0$ and $x_0$ are generated by random sampling from prior probability density functions $p(\theta_0)$ and $p(x_0)$. Algorithm 2 represents steps to estimate $p(x_{t_{k+1}},\theta|y_{1:k+1})$ when the $k+1$th observation becomes available. Our algorithm is an adaptation of the kernel density particle filter of[33] to irregular observations. We have also included the flowchart of our algorithm in Fig. 3.1 for the sake of the reader's convenience.

In our kernel density SMC Algorithm 2, $\bar{\theta}$ is a weighted sample mean and $V_{k+1}$ is a weighted sample covariance. To control the smoothness of kernel density estimation, it is assumed that $a = 1-h^2$ and $h = 1-(\frac{3\phi-1}{2\phi})^2$. $\phi \in (0,1)$ is a discount factor which reduces the chance of failure in the filter. Readers are encouraged to refer to[33] for more details. Typically, $\phi$ is assumed to be a number between 0.95 and 0.99.

### 3.3.3 Data Explanation

Since approximately December 2013, West Africa has been affected by this virus. However, The World Health Organization (WHO) declared the epidemic to be a public health emergency of in-

**Algorithm 2** Kernel Density Sequential Monte Carlo Filter for Intermittent Observations

1: **Compute** $m_{k+1}^{(i)}$**:**
$m_{k+1}^{(i)} = a\theta_k^{(i)} + (1-a)\bar{\theta}_k$;

2: **Compute first moment of** $x_{t_{k+1}}^{(i)}$**:**
$\mu_{k+1}^{(i)} = E(x_{t_{k+1}}|\theta_k^{(i)}, x_{t_k}^{(i)})$, for all $i \in \{1,2,..,J\}$;

3: **Compute auxiliary weights and normalize them:**
$g_{k+1}^{(i)} = w_k^{(i)} p(y_{k+1}|\mu_{k+1}^{(i)}, m_{k+1}^{(i)})$ , $g_{k+1}^{(i)} = \frac{g_{k+1}^{(i)}}{\sum_{l=1}^{J} g_{k+1}^{(l)}}$;

4: **Sampling:** Select an index $j$ randomly with its corresponding weight $\{g_{k+1}^{(1)}, ..., g_{k+1}^{(J)}\}$;

5: **Reproduce the parameters:** $\theta_{k+1}^{(i)} \sim \mathcal{N}_\omega(m_{k+1}^{(j)}, V_{k+1}^{\theta})$, where $\mathcal{N}_\omega(\mu, \sigma)$ is a truncated normal distribution;

6: **Sample the** $x_{t_{k+1}}^{(i)}$**:** $x_{t_{k+1}}^{(i)} \sim p(x_{t_{k+1}}^{(i)}|\theta_{k+1}^{(i)}, x_{t_k}^{(j)})$, for all $i \in \{1,2,..,J\}$.

7: **Recompute first moment of** $x_{t_{k+1}}^{(i)}$**:**
$\mu_{k+1}^{(i)} = E(x_{t_{k+1}}^{(i)}|\theta_k^{(i)}, x_{t_k}^{(j)})$, for all $i \in \{1,2,..,J\}$.

8: **Compute weights and normalize them again:**
$w_{k+1}^{(i)} = \frac{p(y_{k+1}|x_{t_{k+1}}^{(i)}, \theta_{k+1}^{(i)})}{p(y_{k+1}|\mu_{k+1}^{(j)}, m_{k+1}^{(j)})}$ , $w_{k+1}^{(i)} = \frac{w_{k+1}^{(i)}}{\sum_{l=1}^{J} w_{k+1}^{(l)}}$.

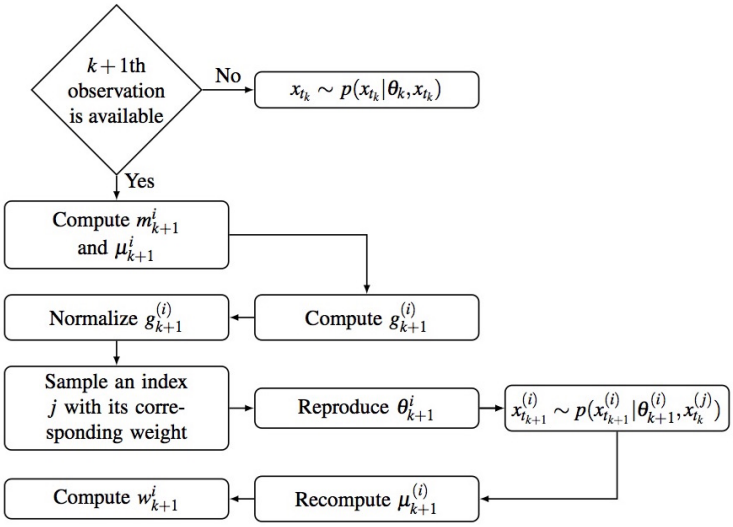


Figure 3.1: Flowchart of kernel density SMC filter for intermitten obsevations as described in Algorithm 2.

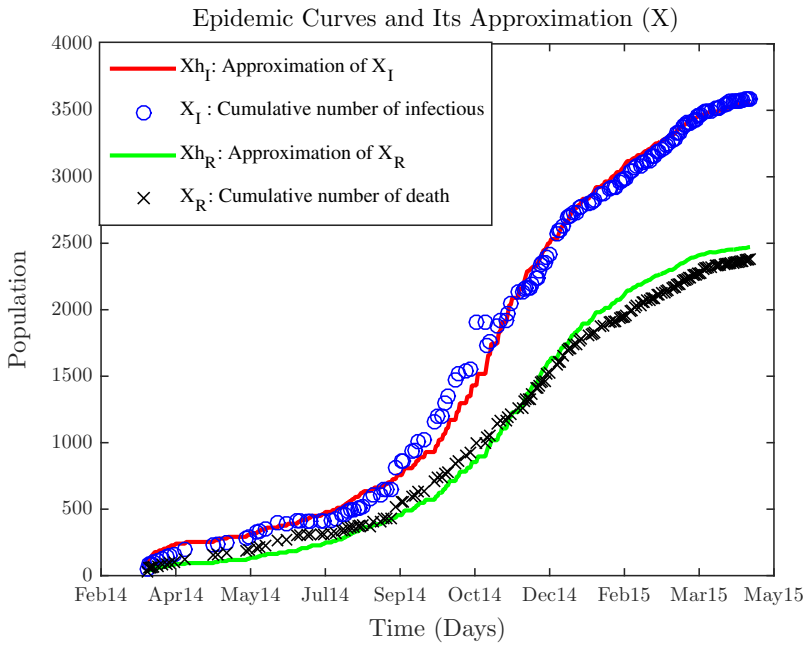


Figure 3.2: Cumulative cases and death data in Guinea reported by WHO[2], and their estimation by the SMC filter in Algorithm 2.

ternational concern on August 8, 2014[2]. We analyze the cumulative case and death counts in Guinea, one of the three West Africa Countries that experienced the Ebola outbreak. Cumulative cases are classified into three categories: confirmed, probable, suspected cases. Similarly, we have three cases for death counts. Confirmed cases are those individuals who are diagnosed by polymerase chain reaction (PCR) method. On the other hand, suspected and probable cases denote those individuals that have symptoms of Ebola but it is not confirmed if they are actually infected[27]. We should mention that the cases were reported at irregular intervals. This data has been collected from reports of WHO available at http://www.healthmap.org/ebola/.

## 3.4 Results

We estimate the states of Ebola propagation from March 23, 2014 to April 30, 2015 having data reported only in $T = 170$ days. Guinea population in 2015 was estimated to be around 12,500,000, however, the population in danger to be infected by the Ebola virus was estimated to be roughly

about $P = 1,000,000$.

We estimate parameters by sampling from a log-normal distribution using $J = 5,000$, number of particles. The sensitivity to changes, discount factor, is chosen as $\phi = 0.95$ and initial weights are all set equal to $w_0 = J^{-1}$. To run the particle filter, the initial state, $x_0$, and initial parameters, $\theta_0$, are generated randomly. Specification of initial priors distributions are reported in Table 3.1 and 3.2.

Table 3.1: Priors specification for parameters

| Parameters | Priors |
|---|---|
| mitigation rate ($\alpha$) | $\mathscr{U}(0.0059, 0.00593)$ |
| transmission rate ($\beta$ ) | $\mathscr{U}(0.259, 0.379)$ |
| latency rate ($\lambda$) | Beta$(78, 577)$ |
| recovery/remove rate ($\gamma$) | Beta$(21, 246)$ |
| fatality rate ($\varphi$) | Beta$(37, 15)$ |

Based on collected data and expert opinion, some measurements for $\gamma$, $\lambda$, and $\varphi$ are available. Therefore, we specify beta distributions for each of parameters, using *Beta buster* and *BetaSlicer*[9,19]. Based on collected information, the average incubation period, $\lambda^{-1}$ is less than 21 days with 95% confidence interval and mean around 8 days[2]. Fatality rate, $\varphi$, is less than 80% with 95% confidence interval and mode 71%[2,7,36]. The average duration of illness onset to death or recovery, $\gamma^{-1}$, is around 12 days[7,36]. Since we do not have enough information about transmission and mitigation rates, $\beta$ and $c$, we assume uniform distributions.

For observation constant in SMC filter, we assume that $b_1 = 0.88$ and $b_2 = 0.54$ and standard deviation, $\Sigma_Y$ is a two by two diagonal matrix whose diagonal elements are $\sigma_1 = 0.00125$ and

Table 3.2: Priors specification for states

| States | Priors |
|---|---|
| $c$ | $\mathscr{U}(0.36, 0.40)$ |
| $E$ | $\mathscr{U}(0.000128, 0.000141)$ |
| $I$ | $\mathscr{U}(0.000050, 0.000061)$ |
| $R$ | $\mathscr{U}(0.000042, 0.000058)$ |
| $D$ | $\mathscr{U}(0.000029, 0.000030)$ |

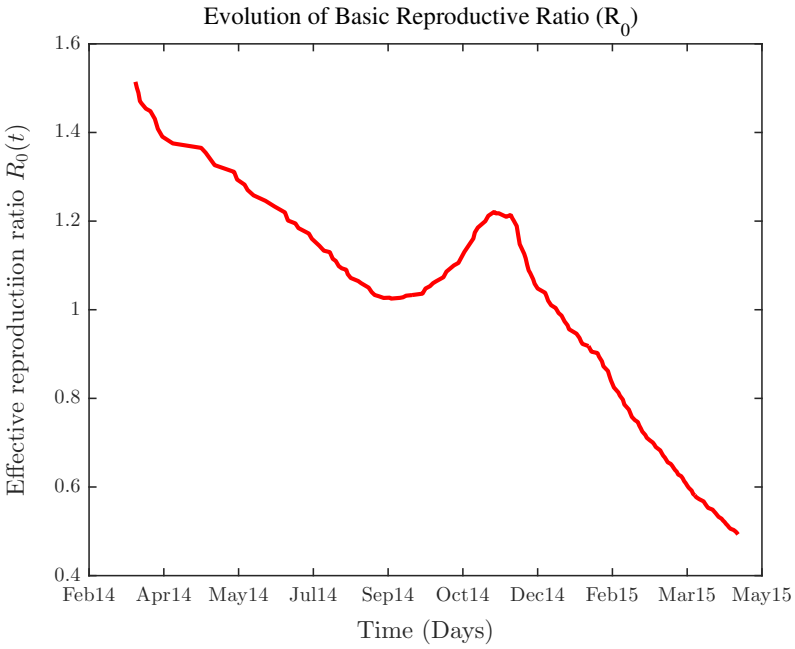


Figure 3.3: Changes in basic reproductive ratio over disease evolution

$\sigma_2 = 0.00085$. Power–law constant $\zeta$, for infected individuals, is 0.88 and for dead individuals, is 0.68.

Fig. 3.2 shows cumulative cases and deaths data and their estimation by the particle filter in Guinea. In our model, the basic reproductive ratio is equal to $R_0(t) = c\beta\gamma^{-1}$. The result indicates that transmission rate, latency rate and, recovery rate are not constant during the disease evolution. Therefore, as demonstrated in Fig. 3.3, $R_0(t)$ is not constant neither.

The maximum value of the basic reproductive number is 1.51 on March 2014 and it decreases until September 2014 which $R_0 \sim 1$. Afterward, a pick is occurred on October 2014 and after that it decreases. We can see in Fig. 3.4 that transmission rates change during the disease evolution. In[15], $R_0(t)$ is estimated as a single value with confidence interval, while in[37] the reproduction ratio, $R_0(t)$, is computed by fitting exponential growth curves to small successive time intervals of the Ebola outbreak. Instead, our method finds the basic reproductive ratio, $R_0(t)$, as a continuous function of time during the Ebola evolution. Using this method, we can also see that not only $R_0(t)$ changes over time, but also parameters such as $\beta$, $\lambda$ and $\gamma$ change.

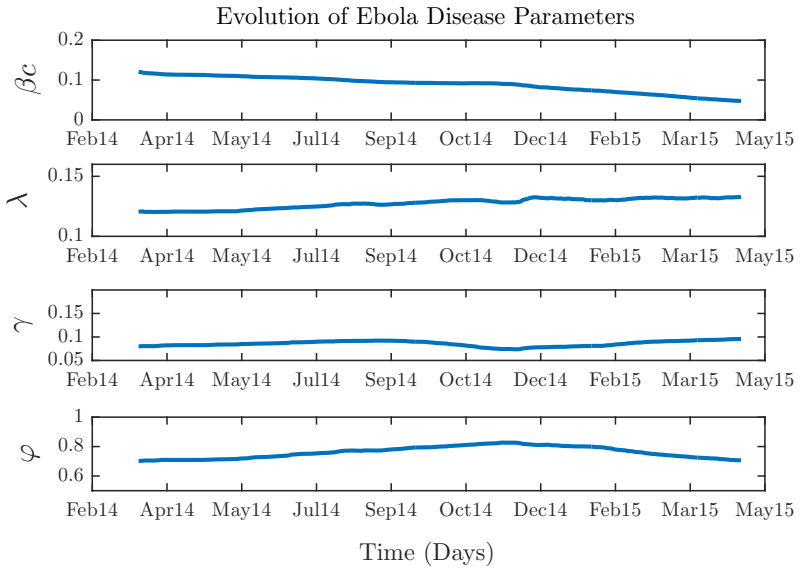


Figure 3.4: Parameters changes over Ebola disease evolution

## 3.5 Conclusion

Our analysis identifies a correlation between $R_0(t)$ temporal variations and important events in the 2014 Ebola outbreak. For instance, a reduction of $R_0(t)$ can be seen around the time WHO first announced the Ebola outbreak in Guinea. This reduction can be explained by taking into account the introduction of some initial medical support and public awareness. Conversely, a peak of $R_0(t)$ corresponds to the first Ebola cases in Europe and the USA.

# Chapter 4

# Conclusion

Immediate implementation of contact tracing strategy and hospitalization as two major mitigation strategies in early–stage of Ebola virus disease introduction have been studied in chapter 2. Our analysis identifies that immediate and accurate implementation of contact tracing protocols and hospitalization could result in reduction of total number of infections at the end of the disease evolution. Immediate hospitalization at early–stages of the disease spreading can decrease basic reproductive number significantly and impede the transmission chain of the Ebola disease spreading. Only implementation of contact tracing strategy may not result in a significant reduction in value of $R_0$. Moreover, It is observed that in early–stage of a disease introduction, using contact tracing strategy, detection of those contacts that are not infected is very difficult from exposed contacts, that may increase the financial burden on public health authorities. In particular, contact tracing strategy focuses on detection of contact of an infected individual more than detection of infectious contact of an infection.

Although immediate and accurate contact tracing strategy could identify infected individuals and their infectious contact properly, it cannot distinguish healthy contacts from exposed contacts. Therefore, to tackle the trade-off between financial burden and halting a virus transmission chain, combination of at least another mitigation protocol, such as hospitalization with contact tracing may result in a cost–effective strategy.

To study the effectiveness of implemented mitigation strategies in a real scenario that number of cases is high, in chapter 3, we investigate the reported data in Guinea. My analysis identifies a reduction in value of $R_0(t)$, when WHO announced Ebola outbreak in Guinea for the first time. This reduction can be explained by considering the introduction of initial medical supports and public awareness. Also, a peak in value of $R_0(t)$ when initial Ebola cases in Europe and the USA are recognized, is observed. Contact tracing as a major implemented mitigation strategy in Guinea does not result in an immediate reduction in value of $R_0(t)$. Value of $R_0(t)$ becomes less than one almost after nine month. This can be explained by the fact that contact tracing cannot distinguish healthy contacts from exposed contacts. So, in such cases that number of cases is high and financial sources are limited, only implementation of contact tracing strategy may not accelerate reduction of number of secondary infections.

Future works can have two main directions, 1) studying impact of temporal network topology on disease spreading and 2) further improvement of SMC method would involve incorporating spatial correlations in prediction and estimation schemes of the SMC method to consider several countries in West Africa at once. Furthermore, a more objective quantification of involved uncertainties and sensitivity analysis can be a great addition to the current work.

# Bibliography

[1] Centers for Disease Control and Prevention.

[2] World Health Organization, Ebola virus disease outbreak.

[3] Althaus, C. L. (2014). Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS currents*, 6.

[4] Browne, C., Gulbudak, H., and Webb, G. (2015). Modeling contact tracing in outbreaks with application to ebola. *Journal of theoretical biology*, 384:33–49.

[5] Chang, C. and Ansari, R. (2005). Kernel particle filter for visual tracking. *Signal processing letters, IEEE*, 12(3):242–245.

[6] Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., and Hyman, J. (2004). The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of Theoretical Biology*, 229(1):119–126.

[7] Chowell, G. and Nishiura, H. (2014). Transmission dynamics and control of ebola virus disease (evd): a review. *BMC medicine*, 12(1):196.

[8] Chowell, G. and Viboud, C. (2016). Is it growing exponentially fast?impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics. *Infectious Disease Modelling*.

[9] Chun-Lung, S. Betabuster Program, Department of Medicine and Epidemiology, University of California, Davis.

[10] Diekmann, O., Heesterbeek, J., and Metz, J. A. (1990). On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28(4):365–382.

[11] Dixon, M. G., Schafer, I. J., et al. (2014). Ebola viral disease outbreak?west africa, 2014. *MMWR Morb Mortal Wkly Rep*, 63(25):548–51.

[12] Doucet, A., De Freitas, N., and Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer.

[13] Eames, K. T. and Keeling, M. J. (2003). Contact tracing and disease control. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1533):2565–2571.

[14] Fisman, D., Khoo, E., and Tuite, A. (2014). Early epidemic dynamics of the west african 2014 ebola outbreak: estimates derived with a simple two-parameter model. *PLOS currents outbreaks*, 1.

[15] Gomes, M. F., y Piontti, A. P., Rossi, L., Chao, D., Longini, I., Halloran, M. E., and Vespignani, A. (2014). Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *PLoS currents*, 6.

[16] Greiner, M., Pfeiffer, D., and Smith, R. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive veterinary medicine*, 45(1):23–41.

[17] He, D., Ionides, E. L., and King, A. A. (2009). Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society Interface*.

[18] Heffernan, J., Smith, R., and Wahl, L. (2005). Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293.

[19] Jones, G. and Johnson, W. O. (2014). Prior elicitation: Interactive spreadsheet graphics with sliders can be fun, and informative. *The American Statistician*, 68(1):42–51.

[20] Kantas, N., Doucet, A., Singh, S. S., and Maciejowski, J. M. (2009). An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. In *15th IFAC Symposium on System Identification (SYSID), Saint-Malo, France.(invited paper)*, volume 102, page 117.

[21] Keeling, M. J. and Rohani, P. (2008). *Modeling infectious diseases in humans and animals*. Princeton University Press.

[22] Kiss, I. Z., Green, D. M., and Kao, R. R. (2005). Disease contact tracing in random and clustered networks. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1570):1407–1414.

[23] Klinkenberg, D., Fraser, C., and Heesterbeek, H. (2006). The effectiveness of contact tracing in emerging epidemics. *PloS one*, 1(1):e12.

[24] Lekone, P. E. and Finkenst"adt, B. F. (2006). Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics*, 62(4):1170–1177.

[25] Liu, S., Perra, N., Karsai, M., and Vespignani, A. (2014). Controlling contagion processes in activity driven networks. *Physical review letters*, 112(11):118702.

[26] Newman, M. (2010). *Networks: an introduction*. Oxford University Press.

[27] Nishiura, H. and Chowell, G. (2014). Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014. *Euro Surveill*, 19(36):20894.

[28] Pellis, L., Ball, F., Bansal, S., Eames, K., House, T., Isham, V., and Trapman, P. (2015). Eight challenges for network epidemic models. *Epidemics*, 10:58–62.

[29] Perra, N., Gonçalves, B., Pastor-Satorras, R., and Vespignani, A. (2012). Activity driven modeling of time varying networks. *Scientific reports*, 2.

[30] Rivers, C. M., Lofgren, E. T., Marathe, M., Eubank, S., and Lewis, B. L. (2014). Modeling the impact of interventions on an epidemic of ebola in sierra leone and liberia. *PLoS currents*, 6.

[31] Rizzo, A., Pedalino, B., and Porfiri, M. (2016). A network model for ebola spreading. *Journal of Theoretical Biology*.

[32] Shahtori, N. M., Scoglio, C., Pourhabib, A., and Sahneh, F. D. (2016). Sequential monte carlo filtering estimation of ebola progression in west africa. *American Control Conference*.

[33] Sheinson, D. M., Niemi, J., and Meiring, W. (2014). Comparison of the performance of particle filter algorithms applied to tracking of a disease epidemic. *Mathematical biosciences*, 255:21–32.

[34] Shirley, M. D. and Rushton, S. P. (2005). The impacts of network topology on disease spread. *Ecological Complexity*, 2(3):287–299.

[35] Skvortsov, A. and Ristic, B. (2012). Monitoring and prediction of an epidemic outbreak using syndromic observations. *Mathematical biosciences*, 240(1):12–19.

[36] Team, W. E. R. et al. (2014). Ebola virus disease in west africa?he first 9 months of the epidemic and forward projections. *N Engl J Med*, 371(16):1481–95.

[37] Towers, S., Patterson-Lomba, O., and Castillo-Chavez, C. (2014). Temporal variations in the effective reproduction number of the 2014 West Africa Ebola outbreak. *PLoS currents*, 6.

[38] Xue, H. (2014). *Data Assimilation Based on Sequential Monte Carlo Methods for Dynamic Data Driven Simulation*. PhD thesis, Georgia State University.