Dynamics and Implications of Data-Based Disease Models in Public Health and Agriculture

by

Bruce Pell

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved August 2016 by the
Graduate Supervisory Committee:

Yang Kuang, Chair
Gerardo Chowell-Puente
John Nagy
Eric Kostelich
Carl Gardner

ARIZONA STATE UNIVERSITY

December 2016

ABSTRACT

The increased number of novel pathogens that potentially threaten the human population has motivated the development of mathematical and computational modeling approaches for forecasting epidemic impact and understanding key environmental characteristics that influence the spread of diseases. Yet, in the case that substantial uncertainty surrounds the transmission process during a rapidly developing infectious disease outbreak, complex mechanistic models may be too difficult to be calibrated quick enough for policy makers to make informed decisions. Simple phenomenological models that rely on a small number of parameters can provide an initial platform for assessing the epidemic trajectory, estimating the reproduction number and quantifying the disease burden from the early epidemic phase.

Chapter 1 provides background information and motivation for infectious disease forecasting and outlines the rest of the thesis.

In chapter 2, logistic patch models are used to assess and forecast the 2013-2015 West Africa Zaire ebolavirus epidemic. In particular, this chapter is concerned with comparing and contrasting the effects that spatial heterogeneity has on the forecasting performance of the cumulative infected case counts reported during the epidemic.

In chapter 3, two simple phenomenological models inspired from population biology are used to assess the Research and Policy for Infectious Disease Dynamics (RAPIDD) Ebola Challenge; a simulated epidemic that generated 4 infectious disease scenarios. Because of the nature of the synthetically generated data, model predictions are compared to exact epidemiological quantities used in the simulation.

In chapter 4, these models are applied to the 1904 Plague epidemic that occurred in Bombay. This chapter provides evidence that these simple models may be applicable to infectious diseases no matter the disease transmission mechanism.

Chapter 5, uses the patch models from chapter 2 to explore how migration in the

1904 Plague epidemic changes the final epidemic size.

The final chapter is an interdisciplinary project concerning within-host dynamics of cereal yellow dwarf virus-RPV, a plant pathogen from a virus group that infects over 150 grass species. Motivated by environmental nutrient enrichment due to anthropological activities, mathematical models are employed to investigate the relevance of resource competition to pathogen and host dynamics.

To William L. Hart

ACKNOWLEDGMENTS

First, I would like to sincerely thank my PhD advisor, Dr. Yang Kuang, for so many fruitful insights, kind suggestions, consistent encouragement and most of all, his patience. He has instilled in me many principles that have impacted my research and my character throughout my time working with him. I am deeply grateful.

I would also like to sincerely thank Dr. Gerardo Chowell, for all of his great suggestions, guidance and patience in the last year of my PhD study. Also, an extended thank you to the rest of my PhD committee, Eric Kostelich, Carl Gardner and John Nagy for their thoughtful insights and support.

It is also my great pleasure to acknowledge William L. Hart, for instilling in me a love for mathematics.

I am extremely grateful to my friends and colleagues, including, but not limited to, Amy Kendig for our discussions and letting me use her plant virus data, Daozhou Gao for his insights, and Tin Phan for his dedication to our research projects.

I am also indebted to Debbie Olson and Jennifer May for all the hard work they do for the graduate students at the School of Mathematics and Statistical Sciences.

Last but not least, I would like to thank my mother and sister, for their love and unwavering support throughout my life.

I am very grateful for the generous financial support from NSF grant DMS-1518529 and Arizona State University.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

1.1 Infectious Diseases

Epidemics of infectious diseases have been documented throughout history. Accounts describing epidemics of smallpox, typhus, tuberculosis and other deadly diseases can be traced back to ancient Greece and Egypt (Geddes, 2006; Retief and Cilliers, 1998). In fact, because of their often high mortality rates, infectious diseases have caused severe political and cultural changes. Indeed, evidence suggests that smallpox is responsible for the death of Pharaoh Ramses V of Egypt (Geddes, 2006) and an infectious disease outbreak in Athens is responsible for reducing 25% of its population during the time period of 430-427 BCE (Littman, 2009; Poole and Holladay, 1979).

One of the most devastating infectious diseases ever recorded is plague. Historians recognize three major plague pandemics, the Justinian Plague, the Black Death (or the Great Plague) and the Modern Plague. Taking place in the Eastern Roman Empire, the Justinian Plague (541-542 AD), was the first plague pandemic recorded and was followed by several smaller outbreaks over the next two hundred years. These smaller outbreaks would eventually lead to over 25 million people succumbing to the disease. Originated from China and spread along the Silk Road and other trade routes, the years of 1334-1353 brought the Black Death, in which an estimated 75 to 200 million people succumbed to plague. In Europe alone, the Black Death was estimated to be responsible for killing 60% of the population (Benedictow, 2005). In the 1860s, plague reappeared in China and spread to port cities around the world by

infected rats on ships. Over a 20 year period, the Modern Plague pandemic claimed approximately 10 million lives (Khan,2004).

It is important to clarify that the term "plague" has been used to describe other infectious disease epidemics that may not be the same disease responsible for the Justinian Plague, the Black Death and the Modern Plague. Indeed, "plague" has been used to describe the fall of the Han Dynasty (Lee, 2002) and was used to describe the death of Roman emperor, Lucius Verus, who died in 169, during the Antonine Plague pandemic (Duncan-Jones, 1996). Nine years later, the Roman historian, Dio Cassius, describes another epidemic of "plague" where up to 2,000 romans succumbed to the disease in one day (Gilliam, 1961). Even today, the causes of these plague epidemics have not been determined, but scholars believe that smallpox or measles is responsible (McLynn, 2011; Gilliam, 1961; Haeser, 1881).

Although it was declared eradicated in 1967, smallpox, is another example of an infectious disease that has been documented throughout history. In fact, evidence suggests that it emerged as early as 10,000 BCE. Smallpox was brought to North America when settlers arrived along the east coast in 1633 in Plymouth, Massachusetts. Their arrival brought devastating smallpox outbreaks among the Native American populations and played an important factor in colonization. Smallpox was also an important factor in the conquest of the Aztecs and the Incas by the Spaniards. In the 20th century alone, smallpox was responsible for killing over 300 million people.

Infectious diseases have also indirectly reduced human populations. During the time period of 1845-1852, Ireland experienced the Great Famine, in which mass starvation and disease caused approximately 1 million people to die, ultimately decreasing the population by 20% to 30%. The cause for the famine is partially accredited to *Pythophthora infestans*, or commonly known as potato blight, which destroyed potato crops throughout Europe during the 1840s.

1.2 Epidemiology

Epidemiology is the branch of medicine that deals with the incidence, distribution and possible control of diseases and other factors relating to the health of a population. The individuals who investigate these factors are called epidemiologists. They seek ways to reduce risk and occurrence of negative health outcomes through research, community education, and health policy. Without their efforts, it would be near impossible for assessing and controlling diseases spreading through populations.

One of the founding fathers of modern epidemiology is John Snow. Known for his work on cholera, Snow, traced the source of the disease to a public water pump during the 1854 cholera epidemic in Soho, London (Snow, 1855). His findings changed the way water and waste was managed in London and throughout the world, ultimately improving public health in general.

Another important historical figure is Ronald Ross. Ross, was a British physician who received a Nobel prize in 1902 for his work on the transmission of malaria. He proved that malaria was transmitted by mosquitoes with his discovery of the malarial parasite inside the digestive tract of a mosquito (Ross, 1897). He further laid the framework for developing methods for preventing the spread of the disease and in addition to his biological discovery, Ross, developed a mathematical model that related the number of mosquitoes and the rate of malaria infection in people (Ross, 1910).

Mathematics has played a role in epidemiology as well. In 1776, Daniel Bernoulli's statistical work on smallpox demonstrated the efficacy of vaccinating the public (Bacaër, 2011; Bernoulli and Blower, 2004). Other influential work is Kermack's and McKendrick's construction of the well-known SIR (susceptible, infected and removed) compartmental model that was used to model bubonic plague in Bombay (Kermack and

McKendrick, 1927). Their modeling ideas were later popularized by the work of May, (May *et al.*, 1979). Nowadays, more complex versions of the Kermack-McKendrick model that incorporate more realistic biology are used.

1.3 Disease Epidemics Discussed in this Dissertation

1.3.1 The 2013-2015 Ebola Virus Disease Epidemic

Ebola virus disease, also known as Ebola hemorrhagic fever (EHF) or commonly referred to as Ebola, is a viral hemorrhagic fever of humans and other primates that is caused by a negative-sense single stranded RNA virus of the family Filoviridae and genus Ebolavirus (Chowell and Nishiura, 2014). The five known virus species are named after the region where they were first identified: Zaire ebolavirus (EBOV), Sudan ebolavirus (SUDV), Tai Forest ebolavirus (TAFV), Bundibugyo ebolavirus (BDBV) and Reston ebolavirus (RESTV) (Chowell and Nishiura, 2014). Although, the reservoir host of Ebola virus remains unknown, researchers believe that the virus is animal-borne and resides in fruit bats (Leroy *et al.*, 2005). The deadliest of the five ebolaviruses, EBOV, was first discovered in 1976 near the Ebola River, in what is now the Democratic Republic of the Congo (formerly Zaire). Also, the 2013 Ebola outbreak was probably the first to occur in Western Africa as prior outbreaks in Africa were restricted to regions within Central and Eastern Africa. Ebola virus disease is usually transmitted through direct contact with blood or body fluids of a person who is sick with or has died from EVD or through direct contact with infected bats or primates (Funk and Kumar, 2015; CDC, 2014). Early nonspecific symptoms include sudden fever, vomiting, weakness, diarrhea, headache and sore throat, but rapidly progress to internal and external bleeding and possible organ failure.

The 2013 Ebola epidemic that occurred within villages, towns and cities of the

Western African nations of Guinea, Liberia and Sierra Leone became the most severe Ebola virus disease (EVD) outbreak in history, with a case fatality rate of 70.8% (95% confidence interval [CI], 68.6 to 72.8) and a hospitalized fatality rate of 64.3% (95% CI, 61.5 to 67.0) (Chowell and Nishiura, 2014; WHO Ebola Response Team, 2014). This epidemic was significantly different in both size and duration compared to previously reported EVD epidemics and as of April 13, 2016, 28,652 cases have been reported, of which 11,325 patients have succumbed to the disease, making it the deadliest Ebola epidemic in history (World Health Organization, 2015a). In addition to these outbreaks, smaller ones occurred in Mali, Nigeria and Senegal, but were quickly contained due to a rapid response and the availability and quality of diagnostics tools and overall health care (Cenciarelli *et al.*, 2015; Althaus *et al.*, 2015a; Frieden TR, 2015).

Standard practices to prevent the outbreak in these countries were not as effective partly due to their poor health infrastructure, including the lack of public health surveillance systems to rapidly detect emerging outbreaks (Frieden *et al.*, 2014). In addition, no licensed vaccine against EVD was available during the 2013-15 epidemic (World Health Organization, 2015c; Agnandji *et al.*, 2015). Instead, quarantine, isolation and education programs were used to mitigate the spread of the disease. This latest outbreak far surpassed the number of reported cases and deaths from ten major previous Ebola outbreaks combined with an estimated 1,531 cases and 1,002 deaths (Center for Disease Control, 2015).

1.3.2 Bubonic Plague Epidemic of 1905

In the late 19th century and early 20th century, bubonic plague was a yearly epidemic in Bombay. Due to the lethal nature of the bubonic plague – it was the leading cause of death in Bombay for two decades (Klein, 1986) – a commission was

formed in 1905 to investigate the mechanisms by which the plague was spreading. A large team of scientists and medical professionals gathered data from 12 different sections in Bombay including deaths due to plague on a bi-weekly basis (Commission *et al.*, 1907).

Bubonic plague is an infectious disease that is caused by the bacterium *yersinia pestis*. Bubonic plague is spread though inflected flea bites, as well as exposure to the tissue or fluids of an animal infected with the plague, but not directly from human to human, unless it is the uncommon pneumonic plague (Stenseth *et al.*, 2008). Untreated, mortality due to bubonic plague is estimated at 30%-60%. In 1906, the plague in Bombay exhibited a mortality rate of approximately 90% (Commission *et al.*, 1907).

When considering the bubonic plague in India, spread of the disease was markedly seasonal. Early attempts at forecasting plague were informed by weather: low hot weather and monsoon season temperatures were expected to result in high plague incidences (Rogers, 1933). Some have suggested the seasonality stems from the ability of the flea to transmit the bacteria to the new host during inclement temperatures, although new research suggests that temperature does not significantly affect efficiency of transmission (Schotthoefer *et al.*, 2011). Regardless, climate appears to be extremely important in vector borne diseases (Gage *et al.*, 2008).

Plague was first observed in the Mandvi region of Bombay in 1896 (Gatacre, 1897). Year after year, the plague epidemic appeared to persist in Bombay in a seasonal fashion. Roughly half the city attempted to flee in order to escape the plague, but the plague spread throughout India in the following years. By 1914, more than 8.5 million people had perished, and some areas of India reported plague deaths until the 1940s (Arnold, 1993).

1.3.3 Cereal Yellow Dwarf Virus

Cereal yellow dwarf virus-RPV, a member of the Polerovirus genus and Luteoviridae family, is a positive-sense single-stranded RNA virus (King, 2011). CYDV-RPV infects over 150 grass species and generally causes leaf discoloration, stunted growth and reduced seed production (D'Arcy *et al.*, 1995). The virus is transmitted by feeding aphids in a persistent manner, that is, the virus must be first taken up by the aphids stylet, then pass through the gut and the salivary glands before the aphid can infect another plant (Ali *et al.*, 2014). In addition, the virus is circulative, meaning it does not replicate while inside of the aphid (Ali *et al.*, 2014). However, viruses move relatively quickly through their plant host, infecting cells throughout the phloem in less than 24 hours (Carrigan *et al.*, 1983).

1.4 Background of Mathematical Infectious Disease Models

1.4.1 Ebola Virus Disease

Several studies have used mathematical models to quantify the effect that control interventions and behavior changes had on managing the 2014 Ebola epidemic. Althaus et al., employed an SEIR (susceptible-exposed-infectious-removed, (Anderson and May, 1991)) model and the estimated effective reproduction number to gain insights into the real-time intervention effects for the 2013-15 EVD epidemic(Althaus, 2014). They concluded that the effective reproduction numbers in Guinea and Sierra Leone decreased to around unity by the end of May and July 2014 due to sufficient control measures. However, that was not the case in Liberia where efforts needed to be improved. In a similar spirit, Chowell et al., employed the logistic model to capture early signs of intervention and behavior changes in the population (Chowell *et al.*, 2014a). Furthermore, they showed that phenomenological models are useful

for understanding early epidemic dynamics, specifically because of the small number of parameters that need to be estimated. With more complexity, Agusto et al., used a mathematical model to explore the effects of traditional belief systems and customs on the transmission process, concluding that the 2014 outbreaks may be controllable by using a moderately-effective basic public health intervention plan (Agusto *et al.*, 2015).

Other studies used mathematical models to investigate the affects of spatial structure on disease dynamics. For instance, Valdez et al., embedded a compartmental model into a 15-patch spatial framework (representing 15 counties of Liberia) and showed that reducing mobility only delayed the overall control of the epidemic (Valdez *et al.*, 2015). Their findings suggested that safe burials and hospitalizations were key to controlling EVD. In particular, if safe burials and hospitalizations were established in mid-July 2014, their model predicts that the epidemic would have been three months shorter and infected individuals would have been 80% less than if the controls were implemented in mid-August. Gomes et al., employed the Global Epidemic and Mobility Model that incorporates mobility and demographic data at a worldwide scale coupled to a stochastic epidemic model (Gomes *et al.*, 2014). They concluded that the probability of the disease spreading outside of Africa was highly unlikely. Merler et al., employed a spatial agent-based model to examine the effectiveness of safe burials, household protection kits and to estimate Ebola virus transmission parameters (Merler *et al.*, 2015). They suggested that the majority of infections occurred within hospitals and households. Their findings indicated that the decline in disease incidence is due in part by the increased number of Ebola treatment units, safe burials and household protection kits. Using a discrete, stochastic SEIR model that is embedded within a three-scale community network model, Kiskowski, showed that effects from community mixing along with stochasticity can explain the different growth rates of

reported cases observed in Sierra Leone, Liberia and Guinea (Kiskowski, 2014).

Multiple studies have used mathematical models for forecasting the potential number of future cases and estimating transmission parameters for the 2013-15 Ebola epidemic. Meltzer et al., constructed the EbolaResponse modeling tool that tracks patients through multiple stages of infection and categorizes patient infectiousness depending on whether they are in a hospital, a low-risk community setting or at home with no isolation (Meltzer *et al.*, 2014). The EbolaResponse model was used to estimate how control and prevention measures could stop the epidemic and to forecast future cases. Meltzer et al., suggested that policy makers rapidly increase the number of Ebola treatment units. In another study, Shaman et al. used a stochastic compartmental model is coupled with the Ensemble Adjusted Kalman Filter (EAKF) to forecast state variables and parameters six weeks into the future (Shaman *et al.*, 2014). The EAKF adjusts the parameters and ensemble state variables as more data becomes available. Parameter estimations provided some evidence that the epidemic growth was slowing down in Liberia.

For readers interested in more details of the studies see Table B.1.

1.4.2 Bombay Plague

The first influential mathematical model used to investigate the spread of bubonic plague in Bombay was constructed by Kermack and McKendrick (Kermack and McKendrick, 1927), a 3-compartment model featuring susceptible population, infected population, and immune (or recovered) population. While this model was able to fit the human plague deaths in 1906, it did not have the capability to shed light on future outbreaks. Additionally, as human-human infection is rare, occurring only in pneumonic plague, this modeling framework was not well suited for explaining disease dynamics.

Keeling and Gilligan (Keeling and Gilligan, 2000a) formulated a model that included susceptible rats, infected rats, recovered rats, average number of fleas per rat, and free fleas. From this epizootic model, they are able to estimate the number of human case that arise. Later, they included a human SIR model coupled to the previous 5-compartment model, but the humans do not transmit the disease within themselves. Their stochastic metapopulation model was used to predict the probability that the epidemic would persist in the rodent population for more than 1, 2, and 10 years (Keeling and Gilligan, 2000b).

Other researchers have furthered this model: Monecke et. al (Monecke *et al.*, 2009) made slight changes and were able to show agreement with the Bombay 1905-1906 epidemic data. Recently, Bacaer extended the model to consider the seasonality of the plague by defining the transmission of the plague from flea to human to be based on temperature, and were able to fit multiple years of data (Bacaër, 2012).

In addition to deterministic models, many have investigated spatiotemporal modeling in order to observe that patterns of propagation of the plague (Yu and Christakos, 2006; Yu *et al.*, 2007). One observation is that the plague often spread during the spring and autumn seasons, and virtually stopped during the summer, exhibiting the seasonal behavior described above.

1.4.3 Within-Host Models

Understanding how diseases affect growth and nutritional value of plants is a significant challenge for supporting a growing human population and satisfying its demand for sustainable food and fuel resources. From a disease ecology perspective, recent studies have shown that the dynamics of within-host pathogen populations are intimately linked to their resource supply and host elemental composition. For instance, researchers have found that nutrient supply alters prevalence and competitive

interactions among two coinfecting viruses in plants (Lacroix *et al.*, 2014). In another study, high host carbon to phosphorus ratios (C:P) were shown to inhibit viral production of *Paramecium bursaria Chlorella virus-1* (PBC-1) in the fresh water alga, Chlorella NC64A (Clasen and Elser, 2007). On the other hand, virus infection has been shown to mediate the effects of elevated C02 on plants and their vectors (Trębicki *et al.*, 2016).

The increase in recent studies linking disease ecology and human alterations of nitrogen, carbon and phosphorus cycles have motivated the development of mathematical modeling approaches for studying within-host pathogen dynamics that incorporate nutrient effects. While within-host pathogen models have been used for decades mainly for assessing various pathways and mechanisms, researchers have only recently started incorporating nutrient effects into mathematical models of within-host pathogen dynamics. For instance, mathematical models have suggested that phosphorus levels have implications for controlling cancer cell growth (Kuang *et al.*, 2004b; Everett *et al.*, 2015).

Several groups have used mathematical models to investigate plant virus dynamics. Spatiotemporal dynamics of plant virus infection was considered by work of Tromas et al. In particular, they used a type of patch model that considered the fraction of infected cells in each leaf of an infected plant host (Tromas *et al.*, 2014). They concluded that virus expansion between-cell is restricted, most likely due to the spatial structure of the host environment. Other researchers have employed delayed differential equations to model the time delay in post-transcriptional gene silencing (PTGS) and the maturation time of the growing plant tissue (Neofytou *et al.*, 2016b). While they did not parameterize the model, they did identify parameter regions associated with recovery and resistant phenotypes and possible chronic infections. In addition to modeling PTGS, Neofytou et al., introduced a new mathematical model to investigate

the role of RNA silencing in a plant infected with two competing viruses (Neofytou *et al.*, 2016a).

1.5 Motivation and the Influence of Data

The mathematical and computational modeling approaches discussed in this dissertation are motivated by the increasing number of novel pathogens that directly and indirectly (e.g. crop diseases) threaten the human population. Mathematical models of epidemic disease spread have been used for decades to gain insight into the transmission dynamics and potential effect of different control strategies, but only recently have researchers started to use the available computational power to simulate, calibrate and generate forecasts of epidemic spread using a variety of epidemic models. However, if available epidemic information is scarce during a developing infectious disease outbreak, mechanistic models that incorporate detailed transmission routes and infection processes for specific diseases often cannot be correctly calibrated quick enough to make informative predictions. Yet, because of their nature, simple mathematical models utilizing a small number of parameters have the potential to be provide a starting platform for quantifying uncertainty of infectious disease dynamics and informing policy makers as an outbreak begins to unfold. To this end, a large portion of this dissertation is concerned with developing and assessing the forecasting potential of simple phenomenological models.

An important theme throughout this dissertation is the use of data to guide the development and iteration of mathematical models. In chapter 2, the simple logistic equation is extended to incorporate spatial structure, because of the overwhelmingly rich amount of spatial data that was made available from the 2013-205 Ebola epidemic in West Africa. In chapter 3, synthetically generated epidemic data is used to compare two phenomenological model's abilities to estimate and forecast key quan-

tities of the epidemic. In this case, estimations can be compared to true values from the synthetically generated epidemic. In chapter 4, data from the Bombay plague epidemic of 1904, motivates the inclusion of a deceleration parameter into the logistic equation to capture the sub-exponential growth of the early epidemic phase seen in some of the parts of Bombay. Returning to spatial modeling, in chapter 5, a barren field separating an important section of Bombay from the rest of the city motivates the use of spatial models to investigate this natural barricade's effect on final epidemic size. Finally, in the last chapter, the best fit solutions from two mathematical models uncover patterns that otherwise would be hard to detect when viewing the data by itself.

1.6 Outline of this Dissertation

In chapter 2, logistic patch models are used to assess and forecast the 2013-2015 West Africa Zaire ebolavirus epidemic. We are motivated by the question: How can we use metapopulation models to understand the role of movement restrictions and migration patterns on the spread of infectious diseases? In particular, each models ability to forecast epidemic data was assessed by comparing model forecasting error, parameter distributions and parameter confidence intervals as functions of the number of data points used to calibrate the models. The patch models show an improvement over the logistic model in short-term forecasting, but naturally require the estimation of more parameters from limited data.

In chapter 3, three simple phenomenological models inspired from population biology are used to assess the Research and Policy for Infectious Disease Dynamics (RAPIDD) Ebola Challenge; a simulated epidemic that generated 4 infectious disease scenarios. Because of the nature of the synthetically generated data, model predictions are compared to exact epidemiological quantities used in the simulation.

In particular, we assess the accuracy and performance of estimating the effective reproduction, final epidemic size and future incidence forecasts.

In chapter 4, we apply the simple models from the previous chapter to the 1904 Plague epidemic that occurred in Bombay. Plague is considered in this chapter because of the different transmission mechanism (rats) that are the main driver of the disease burden. This chapter provides evidence that these simple models may be applicable to a diverse set of infectious diseases no matter the disease transmission mechanism. The effective reproduction number is estimated and the early growth phase is analyzed revealing prominent exponential growth in most parts of Bombay. However, evidence for sub-exponential growth is shown in some parts of Bombay as well.

Chapter 5 returns to the patch models defined in chapter 2 to explore the effect that mobility has on final epidemic size. In particular, bifurcation diagrams are produced by calibrating the 3-patch model with reported case data from sections in Bombay. This chapter highlights the importance of isolation and quarantine strategies used in controlling infectious disease epidemics.

The final chapter is an interdisciplinary project concerning within-host dynamics of cereal yellow dwarf virus-RPV, a pathogen from a virus group that infects over 150 grass species including crops. Motivated by environmental nutrient enrichment due to anthropological activities, mathematical models are employed to investigate the relevance of resource competition to pathogen and host dynamics.

Chapter 2

PATCH MODELS OF EBOLA VIRUS DISEASE (EVD) TRANSMISSION DYNAMICS

2.1 Introduction

The 2013-2015 West African Ebola epidemic provides a unique chance to assess the importance of spatial structure in forecasting an infectious disease epidemic due to the rich amount of spatial data. Our modeling goals here are to accurately describe the data with simple models.

We present a simple approach that phenomenologically connects the effects of behavior changes to mitigate transmission rates and population spatial structure. Our method derives the logistic equation from an assumption about the effect of population behavior and introduces spatial heterogeneity via logistic patch models. In particular, we contribute the following:

- The logistic model is derived from a susceptible-infected compartmental model in section 2.2.1, justifying its use in Chowell *et al.* (2014a).
- Formulas for the basic and effective reproduction numbers are presented in Section 2.2.2.
- We build upon the work done in Chowell *et al.* (2014a), by incorporating spatial heterogeneity via logistic patch models.
- Models are validated by comparing their fits to total reported case data in Section 2.4.1.

- As seen in Fig. 2.4, we show that these models improve upon the short term forecasting error in section 2.4.2. Furthermore, we perform Kruskal–Wallis tests to analyze the variation across the different models.

- Further model validation and comparison is presented in Section 2.4.3, via parameter estimations and confidence intervals. This section shows that patch models are not well constrained due to limited data.

- We provide estimates and 95% confidence intervals of R_0 for Liberia, Sierra Leone and Guinea respectively in section 2.4.4.

2.2 Modeling Methods

2.2.1 Logistic Equation as an Ebola Cumulative Infections Case Model

From a basic SI compartmental model and an assumption about population behavior we can derive the logistic equation. Assuming there are no births, natural deaths or immigration of susceptible individuals and that infected individuals do not return to the susceptible class, the classical Kermack and McKendrick infectious disease model can be adapted to obtain the following:

$$\begin{aligned} S(t)' &= -\frac{\beta S(t) I(t)}{S(t) + I(t)}, \\ I(t)' &= \frac{\beta S(t) I(t)}{S(t) + I(t)} - \mu I(t), \end{aligned} \tag{2.1}$$

where β is the infection rate and μ is the disease induced death rate. From system (2.1) the cumulative number of infections at time t, denoted by $x(t)$, has derivative $x'(t) = \beta \frac{SI}{S+I} \approx \beta I$, (assuming $\frac{S}{S+I} \approx 1$). Below we assume that $x'(t) = \beta I$.

As an increasing number of cases are reported during an outbreak, the behavior of the individuals in the affected region may change due to disease education programs, an increase in care or quarantine facilities and help from health care workers.

As an example, dead bodies infected with Ebola virus remain infectious, causing participants to unknowingly contract the infection during funeral burials. In the beginning stage of the outbreak, unsuspecting mourners would carry the infection back to other parts of the community and would infect more individuals. By having specific handling guidelines of human remains, communities were able to decrease exposure to the Ebola virus (World Health Organization, 2015d). In general, this is the notion of a positive behavioral change in the community. Based on these observations we make what we call the behavior assumption:

- **(Behavior assumption)**: During an epidemic, a change in behavior in the community that mitigates the transmission rates is expected as an epidemic unfolds. This response is modeled by a function of the total reported cases and has a decreasing affect on per-capita infection rate. That is,

$$\frac{I'(t)}{I(t)} = f(x(t)) \tag{2.2}$$

  is a decreasing function of the total number of reported cases $x(t)$.

In the following, we assume that $f(x(t)) = r(1 - ax(t))$ for some positive constants $r := \beta - \mu$ and a. Hence

$$I'(t) = rI(t)(1 - ax(t)) = \frac{r}{\beta}x'(t)(1 - ax(t)).$$

Therefore,

$$I(t) - I(0) = \frac{r}{\beta}\left(x(t) - \frac{a}{2}[x(t)]^2\right) - \frac{r}{\beta}\left(x(0) - \frac{a}{2}[x(0)]^2\right).$$

Since $I(0) = x(0) \approx 0$, we see that $I(t)$ can be approximated by

$$\frac{r}{\beta}\left(x(t) - \frac{a}{2}[x(t)]^2\right).$$

Therefore

$$x'(t) = \beta I(t) = r\left(x(t) - \frac{a}{2}[x(t)]^2\right) = rx(t)\left(1 - \frac{x(t)}{K}\right), \tag{2.3}$$

where $K = 2/a$. Here we interpret r as the *intrinsic infection rate*, a is a proportionality constant that corresponds to strength and effectiveness of disease interventions and preventive strategies and K is the *final epidemic size*.

In a study by Chowell et al., the saturation effect of the logistic equation was used to implicitly account for the behavior change in the population (Chowell *et al.*, 2014a). The above derivation provides a rigorous framework of this modeling effort and emphasizes the role behavior plays in the saturation effect. Fig. 2.1 shows the change in the 95% prediction band using the delta method as more data points are incorporated when fitting the logistic model to epidemic data (Bickel and Doksum, 2015).

2.2.2 Derivation of the Basic and Effective Reproduction Numbers

During an outbreak, there may not be enough data to calibrate mechanistic models of the exact transmission processes, thus the logistic model can provide useful insights into the early outbreak dynamics. To derive R_0 and R_e first observe that,

$$I(t+T) = R_e(t)\, I(t), \tag{2.4}$$

where T is the mean generation interval and is defined as the time between infection in an index case patient and infection in a patient infected by that index case patient WHO Ebola Response Team (2014). From equation (2.2), we have that $I'(t) = f(x(t))\, I(t)$, integrating both sides from t to $t+T$ yields

$$\ln(I(t+T)) - \ln(I(t)) = \int_t^{t+T} f(x(s))\, ds.$$

Solving for $I(t+T)$ and dividing by $I(t)$ yields $\frac{I(t+T)}{I(t)} = e^{\int_t^{t+T} f(x(s))ds}$, which from equation (2.4) yields

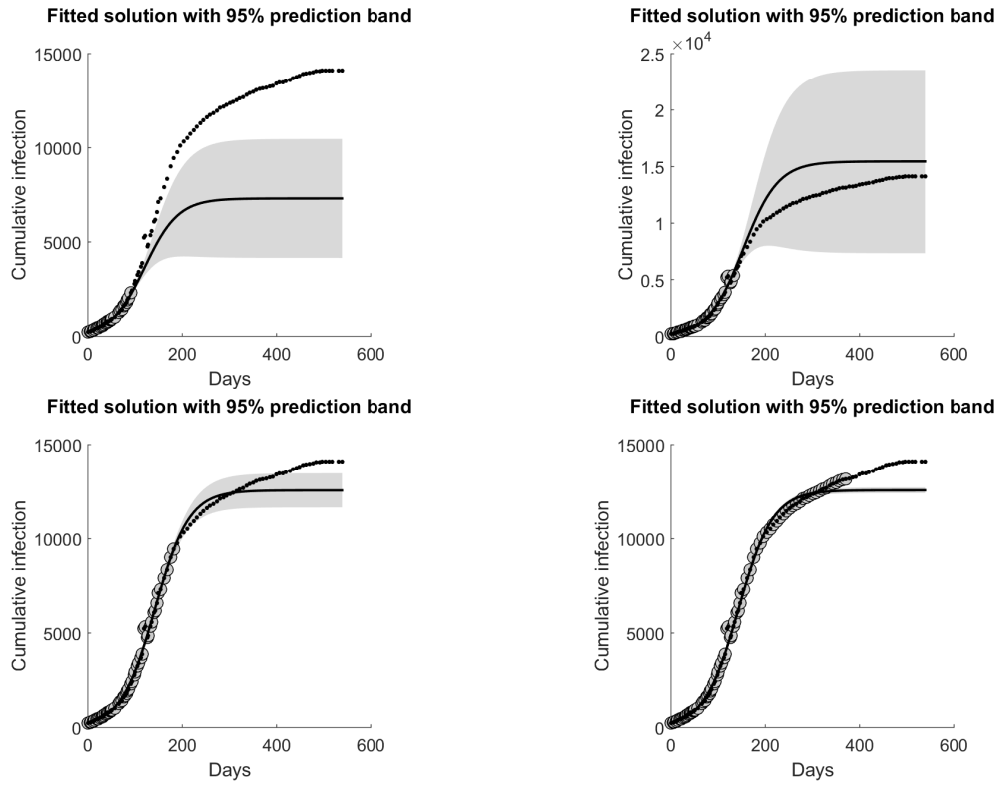

Figure 2.1: Predictions of the cumulative number of Ebola cases in Sierra Leone by the logistic growth equation (2.3). Data points start at June 2, 2014 and end December 23, 2015. 95% prediction bands are superimposed. Gray disks are data points for model calibration, while black dots are forecasting data points.

$$R_e(t) = e^{\int_t^{t+T} f(x(s))ds}. \tag{2.5}$$

Lastly, define $R_0 := e^{rT}$ which is approximately equal to the usual definition of the basic reproduction number, $\frac{\beta}{\mu}$, of model 2.1 when $\frac{\beta}{\mu}$ is close to 1.

2.2.3 Incorporating Population Heterogeneity: Multi-Patch Models

District geography, topology, health care centers and quarantined regions can influence population movement. This motivates the need for incorporating spatial structure in transmission models. We do this by partitioning a district into a network of two or more sub-districts (patches). In each sub-district, cumulative infections

obey logistic growth individually.

Let x_i be the **cumulative infections** in patch i and let m_{ij} be the rate of cumulative infections that travel from patch i to patch j, where $i, j = 1, 2, i \neq j$.

The equations for the two-patch model are:

$$\begin{aligned} x_1' &= r_1 x_1 \left(1 - \frac{x_1}{K_1}\right) - m_{12} x_1 + m_{21} x_2, \\ x_2' &= r_2 x_2 \left(1 - \frac{x_2}{K_2}\right) - m_{21} x_2 + m_{12} x_1. \end{aligned}$$

Similarly, the three-patch model is given by:

$$\begin{aligned} x_1' &= r_1 x_1 \left(1 - \frac{x_1}{K_1}\right) - (m_{12} + m_{13})\, x_1 + m_{21} x_2 + m_{31} x_3, \\ x_2' &= r_2 x_2 \left(1 - \frac{x_2}{K_2}\right) - (m_{21} + m_{23})\, x_2 + m_{12} x_1 + m_{32} x_3, \\ x_3' &= r_3 x_3 \left(1 - \frac{x_3}{K_3}\right) - (m_{31} + m_{32})\, x_3 + m_{13} x_1 + m_{23} x_2. \end{aligned}$$

In addition, consider two special cases of each model: symmetric migration (S) with $m_{ij} = m_{ji}$ and homogeneous migration (H) with, $m_{ij} = m$ for all i, j and $i \neq j$.

Assume that r_i and K_i are positive in the above models. It is easy to see that these patch models are cooperative in nature which generate a strictly monotone semiflow. It is shown that the positive solutions of the above models tend to a unique positive steady state (see Lemma 3.1 in Gao and Ruan (2012)). Furthermore, all solutions that start positive remain positive and are bounded. We prove this with the following two lemmas.

Lemma 2.2.1 (Positivity) *All solutions from a general n-patch model,*

$$x_i' = r_i x_i \left(1 - \frac{x_i}{K_i}\right) + \sum_{j \neq i}^{N} m_{ji} x_i - \sum_{j \neq i}^{n} m_{ij} x_i \tag{2.6}$$

for $i, j \in \{2, ..., n\}$ with positive initial conditions, remain positive for all $t > 0$.

Proof Assume $x_k(0) > 0$ for all $k \in \{2, .., n\}$. Further assume that there exists a $t_1 = \min\{t > 0 : x_i(t) = 0, \text{ for some } i \in \{1, 2, 3, ..., n\}$, that is, t_1 is the first time that one of the components goes to zero. Then for $t \in (0, t_1)$ we have,

$$\begin{aligned} x_i' &= r_i x_i \left(1 - \frac{x_i}{K_i}\right) + \sum_{j \neq i}^{n} m_{ji} x_i - \sum_{j \neq i}^{n} m_{ij} x_i \\ &> -\frac{r_i x_i}{K_i} - \sum_{j \neq i}^{n} m_{ij} x_i \\ &= -\left(\frac{r_i}{K_i} + \sum_{j \neq i}^{n} m_{ij}\right) x_i \\ &> -\alpha x_i \end{aligned}$$

which implies,

$$x_i(t) \geq x_i(0) e^{-\alpha t} > 0 \tag{2.7}$$

where $\alpha = \max_{i,j}\{\frac{r_i}{K_i} + \sum_{j \neq i}^{N} m_{ij}\}$. In particular, we find that $x_i(t_1) > 0$, a contradiction. Therefore, all solutions that start positive, remain positive. □

To see that solutions are bounded we have the following lemma,

Lemma 2.2.2 (Boundedness of solutions) *Let $r^\circ = \max r_i$, $r_\circ = \min r_i$ and $\bar{K} = \max K_i$, then $\sum_{i=1}^{n} x_i \leq n\bar{K}\frac{r^\circ}{r_\circ}$.*

Proof Assuming positivity of solutions and defining $z = \sum_{i=1}^{n} x_i$ we obtain,

$$\begin{aligned} z' &= \sum_{i=1}^{N} r_i x_i \left(1 - \frac{x_i}{K_i}\right) \\ &\leq \sum_{i=1}^{n} r_i x_i - \frac{r_i x_i^2}{\bar{K}} \\ &\leq \sum_{i=1}^{n} r^\circ x_i - \frac{r_\circ}{\bar{K}} \sum_{i=1}^{N} x_i^2. \end{aligned}$$

Noting that $z^2 = \left(\sum_{i=1}^n x_i\right)^2 = \sum_{i=1}^n x_i^2 + 2\sum_{i<j}^n x_i x_j$ and

$$z^2 = \left(\sum_{i=1}^n x_i\right)^2 \leq \sum_{i=1}^n x_i^2 + (n-1)\sum_{i=1}^n x_i^2$$

or

$$\sum_{i=1}^n x_i^2 \geq \frac{1}{n} z^2.$$

Therefore we obtain,

$$z' \leq r^\circ z - \frac{r_\circ}{n\bar{K}} z^2 = r^\circ z \left(1 - \frac{z}{n\bar{K}\frac{r^\circ}{r_\circ}}\right)$$

and solutions are bounded by applying a standard comparison argument. □

2.2.4 The Basic Reproduction Number

Let and $x = \sum_{i=1}^n x_i$. As with the derivation of R_e and R_0 for the logistic model above, define the basic reproduction number for an n-patch model as

$$R_e(t) \;=\; \exp\left(\hat{r}\int_t^{t+T} 1 - \frac{2}{\hat{K}} x(s)\, ds\right),$$

where $\hat{r} = \frac{\sum_{i=1}^n r_i K_i}{\hat{K}}$, $\hat{K} = \sum_{i=1}^n K_i$ are weighted averages and for simplicity we assume $T = 2$ weeks, instead of 2.18 days (Agnandji *et al.*, 2015). Similarly to above, we define $R_0 := e^{\hat{r}T}$.

2.3 Comparison Methods

We use district data from the World Health Organization (WHO) patient database, which contains weekly reported confirmed, suspected and probable infections from Liberia, Sierra Leone and Guinea (World Health Organization, 2015a). Data ranges from Mar. 1, 2014 to Aug. 5, 2015.

Table 2.1: Number of parameters for each model.

	Logistic	Two-patch (H)	Two-patch	Three-patch (H)	Three-patch (S)	Three-patch
Number of parameters	2	5	6	7	9	12

Table 2.1 lists the number of parameters of each model. By studying the special cases of the patch models we reduce the number of parameters that need to be estimated, which constrains model fits and reduces the likelihood of over-fitting the data.

We use Matlab's built-in function, fminsearch, to help locate optimized parameter values for data fitting. fminsearch is a derivative-free method that is based on the Nelder-Mead Simplex and searches for minimums, but does not guarantee global minimums (Lagarias *et al.*, 1998). We are searching for a biologically reasonable parameter set that minimizes the error between the simulations and the observed data. To this end, we define the weighted error function:

$$E_w = \frac{1}{N-P} \sum_{i=1}^{N} |y_i - \hat{y}_i| \, e^{-0.1\left(t_f - t_i\right)}, \tag{2.8}$$

where t_f is the final date that we have an observation for, P is the number of parameters and N is the number of observations. $\hat{y}_i$ denotes the observation at time t_i and y_i the value of our model at the i-th observation. We make the assumption that recent data has higher significance for forecasting future cases, as reflected by the exponential factor. The value of .1 in the exponential term is used because it gave a reasonable temporal-weight to the data points.

2.3.1 Ranking Models by Fitting and Forecasting Errors

To compare the models, we use absolute and relative errors that penalize models that have more parameters. The absolute error is calculated using the following equation,

$$E_{abs} = \frac{1}{\sqrt{N-P}} \sqrt{\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2} \tag{2.9}$$

and the relative error is given by,

$$E_{rel} = \frac{1}{\sqrt{N-P}} \sqrt{\sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{\hat{y}_i}\right)^2}. \tag{2.10}$$

Since we are interested in assessing and ranking the forecasting performance of all models, we define the *forecasting error* as follows:

$$E_{fcst} = \frac{1}{\sqrt{N-\hat{N}-P}} \sqrt{\sum_{i=i^*}^{\hat{N}} \left[y\left(t_i\right) - \hat{y}\left(t_i\right)\right]^2}, \tag{2.11}$$

where i^* corresponds to the temporal index at which we start forecasting our models, N is the total number of observations and $\hat{N}$ is the total number of observations used for model calibration and P is the number of parameters. If i^* was not an integer value, we took its floor value.

2.3.2 Parameters and Confidence Interval Assessment

To further compare and assess the models we compute 95% confidence intervals for the logistic, two-patch (H) and three-patch (H) models. Only these models were considered, because they have the least number of parameters which reduces the likelihood of overfitting the models to data.

Bootstrapping can be used as a way to estimate standard errors of parameter estimates in statistical models. The basic idea is to fit the model to data, find the residuals and add them to the data. Next, randomly sample with replacement B times, where B is large and fit the model to each of these newly created data sets to obtain B different parameter sets from the fitted model. This allows one to obtain a distribution of the parameters without assuming anything prior about them. For further details, see Davison and Hinkley (1997); Efron and Tibshirani (1994); Pardoe and Weisberg (2001).

Recall a statistical model, with $y = (y_1, ..., y_n)$ being explained by k explanatory variables $\boldsymbol{x} = (x_1, ..., x_k)$ using p parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$:

$$y_i = g(\boldsymbol{x_i}|\boldsymbol{\theta}) + \epsilon_i$$

for $i = 1, ..., n$. Where g is a mathematical model such as an ordinary differential equation model, partial differential equation model, algebraic model, etc. ϵ is the error and is a random variable and y is another random variable. Let G be the partial derivative matrix with respect to θ and the leverages, $h_1, ..., h_n$ be the diagonal elements of the $G(G^\dagger G)^{-1}G^\dagger$ matrix, where $\dagger$ denotes matrix transpose.

The bootstrapping method is described below.

1. Fit the model to the original data with an initial parameter set, $\hat{\boldsymbol{\theta}}$, and for each x_i, compute the corresponding residual $\hat{\epsilon}_i = y_i - \hat{y}_i$ for $i = 1, 2, ...n$, where n is the total number of data points and $\hat{y}_i = g(x_i, \hat{\boldsymbol{\theta}})$.

2. Correct for the potential heteroscedasticity in the residual variances by computing the **modified residuals**: $\hat{r}_i = \frac{\hat{\epsilon}_i}{\sqrt{1-h_i}}$ and compute the centered residuals $r_i^* = \hat{\epsilon}_i - \hat{r}_i$, where h_i are the leverages.

3. Sample with replacement from the n modified and centered residuals.

4. Generate bootstrap sample, $\bar{y}_i := \hat{y}_i + r_j^*$, for all i and where j is random.

5. Fit the model to these new $\bar{y}_i$ values and obtain a new set of parameter values, $\bar{\boldsymbol{\theta}}$.

6. Repeat steps 3, 4 and 5 a large number of times [1] of times (say 2,000). This generates 2,000 bootstrap samples and corresponding sets of parameter estimations.

7. Use the 2,000 parameter estimates to generate distributions to find confidence intervals.

2.3.3 Challenges

Whenever fitting a mathematical model to time series data that ranges from small values to very large values, deciding at what time to initiate the model can seriously influence its forecasting ability. For example, training a model on a large set of data that is relatively near zero except for the last couple of points will force the fitting to be heavily biased by the large amount of initial points near zero, thus not providing a good forecast. We remedied this by starting the models after there were no three consecutive weeks that had no infections and by using the weighted error (Eq. 2.8) for fitting. This was done due to the fact that smaller outbreak waves happened before the main wave of infections appeared.

Forecasting an ongoing disease outbreak in real-time brings many challenges. New data being available means that computer programs must be designed to process and incorporate new data sets with ease and in a timely fashion. In our case, fitting six

[1] Results from Efron and Tibshirani Efron and Tibshirani (1994) suggest that accurate results for confidence intervals can be obtained from 1000 bootstrap samples. For standard errors this number is reduced to 200.

models (including special cases) to forty-one data sets requires a significant amount of computing resources.

2.4 Results

2.4.1 Data Based Model Validation

To validate the patch models for epidemic modeling, we fit all models to all data sets and compare model fits and errors.

We report the means for the weighted, relative and absolute error (respectively equations 2.8, 2.9 and 2.10) for all 39 data sets. Observe that the patch models show an improvement over the logistic model when fitting the data. Additionally, we see that the homogeneous migration models perform better than their free migration versions.

Table 2.2: Mean error statistics

Model	Weighted Error	Relative Error	Absolute
Logistic	82.2822	1.3387	102.198
2-Patch (H)	53.6764	1.1271	63.1193
2-patch	58.6311	1.2124	72.7197
3-Patch (H)	48.709	1.1256	59.3391
3-Patch (S)	55.0951	1.1515	65.1694
3-Patch	54.215	1.1727	66.0885

In what follows, we summarize the different fitting and forecasting cases. Let FTG be the fitting error from Eq. 2.8 and FCST be the forecasting error from Eq. 2.11. We use the following convention to denote the different errors: FTG-Δ and FCST-Δ-Ω, where Δ is the fraction of data used for fitting and Ω is the number of weeks forecasted ahead.

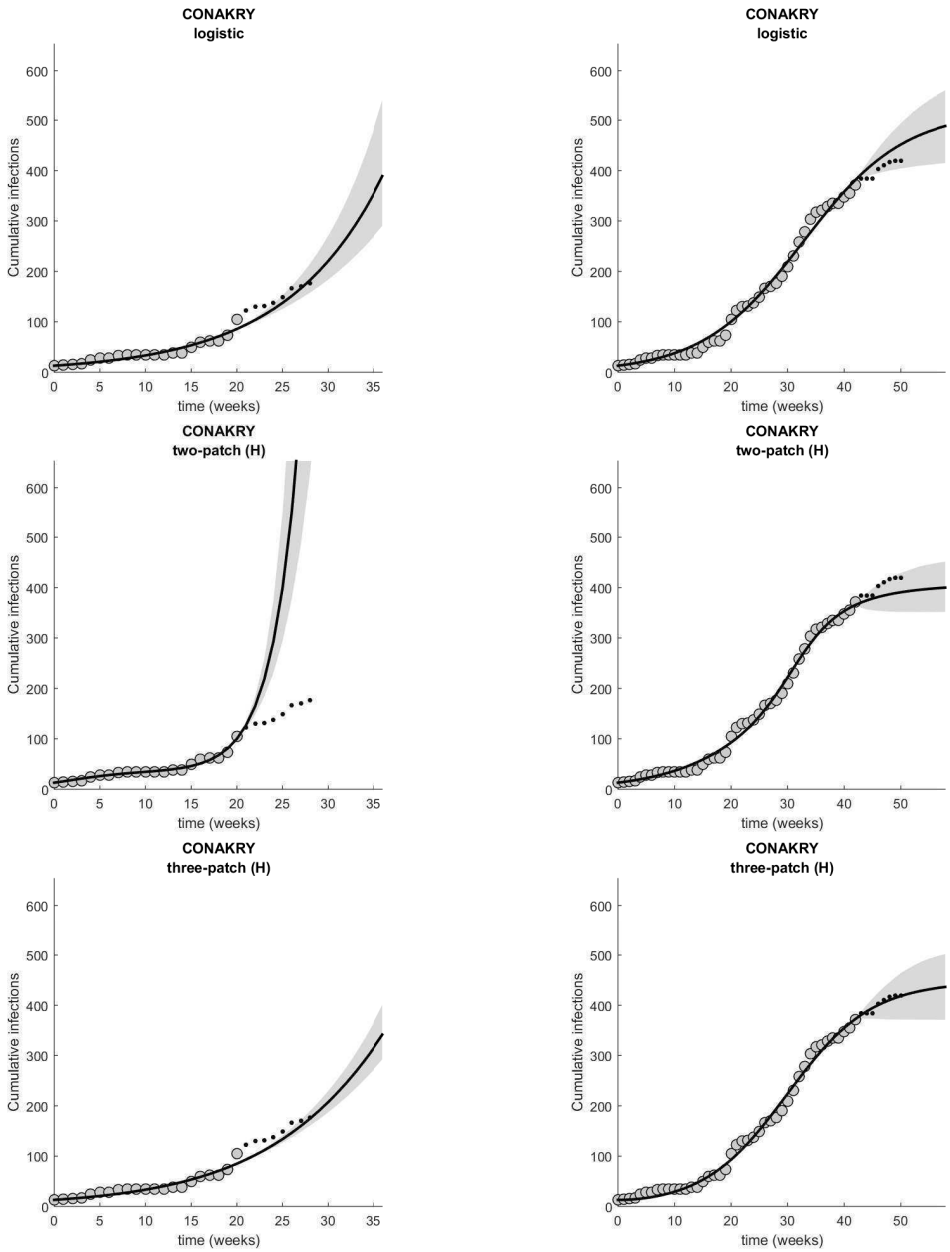

Figure 2.2: Illustration of the model fitting and forecasting for Conakry Left column: models trained on the first one-third data. Right column: models trained on the first two-thirds of data. Gray shaded region represents 95% prediction bands.

Fitting errors were calculated using Eq. 2.8 and the first one-third and the first two-thirds of each data set. All fitting errors are provided in Table A.1 given in the appendix. From Fig. 2.3, most of the patch models had smaller mean fitting error than the logistic model.

Four and eight week forecasts were made after training all models to the first one-third and first two-thirds of the data set. Figure 2.3 shows that in all cases, the patch models had smaller mean forecasting errors. This supports the hypothesis that modeling spatial structure within the district improved forecasting error. Additionally, all models perform better when forecasting the short-term rather than long-term epidemic trajectory. Forecasting error variance was lowest with FCST-2/3-4. In contrast, the variance was the largest with FCST-1/3-8.

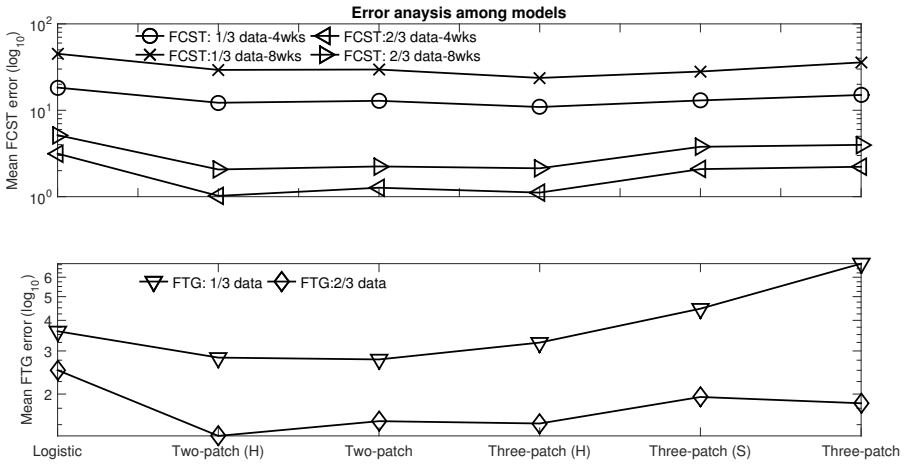

Figure 2.3: Mean forecasting and fitting errors. Models are along the x-axis and variance is along the y-axis. We connect points for aesthetic purposes.

Results of Kruskal-Wallis tests were not significant for FCST-1/3-4, FCST-1/3-8, FCST-2/3-4 and FCST-2/3-8; the mean ranks for all forecasting cases did not significantly differ. We include the p-values (95%) in table 2.3 for this test.

2.4.2 Forecasting Error as a Function of Forecasting Points

Forecasting error for Port Loko, Guinea, Liberia and Sierra Leone was calculated for varying amounts of forecasting points.

The forecasting error for Port Loko in Fig. 2.4, suggests that the patch models have smaller forecasting errors than the logistic equation for short-term forecasts (4 to 70

Case	p-value
FCST-3-4	0.9806
FCST-3-8	0.9872
FCST-23-4	0.9933
FCST-23-8	0.9894

Table 2.3: P-values of the Kruskal-Wallis test show forecasting errors do not significantly differ across models.

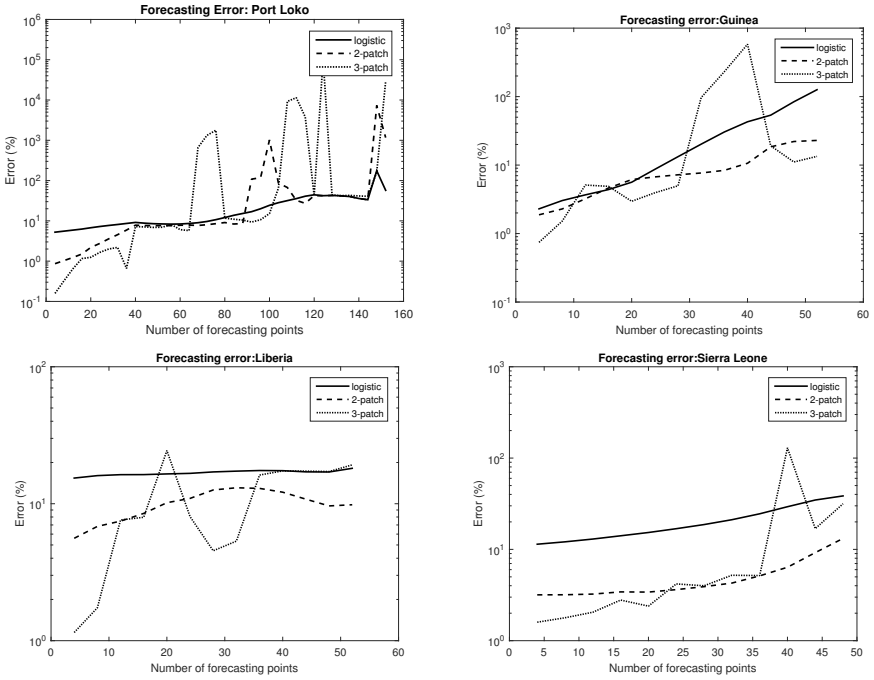

Figure 2.4: Relative error as a function of forecasting points for the logistic, two and three patch models with homogeneous migration rate: Port Loko, Guinea, Liberia and Sierra Leone.

days). Additionally, it shows erratic long-term forecasting of the patch models for Port Loko, because they are not well constrained due to the limited data. Fig. 2.4 further shows lower short-term error for Sierra Leone and Liberia by the two-patch model. We note that the three-patch model yielded the smallest error when forecasting ten prediction points or less (4-10 days).

2.4.3 *Confidence Interval Assessment*

Parameter confidence intervals for the logistic equation decrease in length as we decrease the number of prediction points (Fig. 2.5) for Port Loko. Similar assessments were done using data from Sierra Leone, Liberia and Guinea at the country level. Results were similar as the Port Loko case except for Liberia, where confidence interval lengths begin to increase when we forecast less data points. In summary, the logistic model shows well behaved parameter values when we fit to an increasing number of data points for three out of four data sets used.

The patch models tell a different story. The confidence intervals are larger and show erratic behavior when forecasting a large number of points. Indeed, for the two-patch model (Fig. 2.6), the confidence intervals for r_1 actually increase when we are predicting a small number of data points from Port Loko. This variability is seen to be worse in the confidence intervals for the final epidemic sizes (K_i's) for both two and three patch models, but they are so erratic that they cannot be shown in a reasonable way and therefore are not included. The fact that the patch models have more parameters allows for different parameter sets that produce a well fit curve, but allow for large variability in the parameter sets. The same is seen in the confidence interval assessment using data from Sierra Leone, Guinea and Liberia, (not shown here).

2.4.4 *Implications for Liberia, Sierra Leone and Guinea: R_0*

From the bootstrapping method, we calculated 95% confidence intervals for R_0 in Guinea, Liberia and Sierra Leone, see Table 2.4 for our estimations compared to values found in the literature.

We use the models and calculate R_0 and R_e for each Guinea, Liberia and Sierra

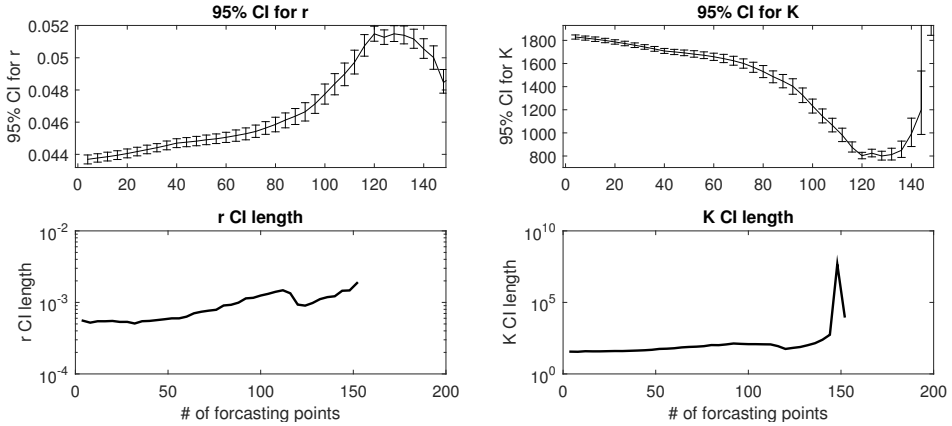

Figure 2.5: 95% CI for r and K from equation 2.3. (Bottom) Plot of the length of the CI for r and K as a function of the number of forecasting points. District: Port Loko.

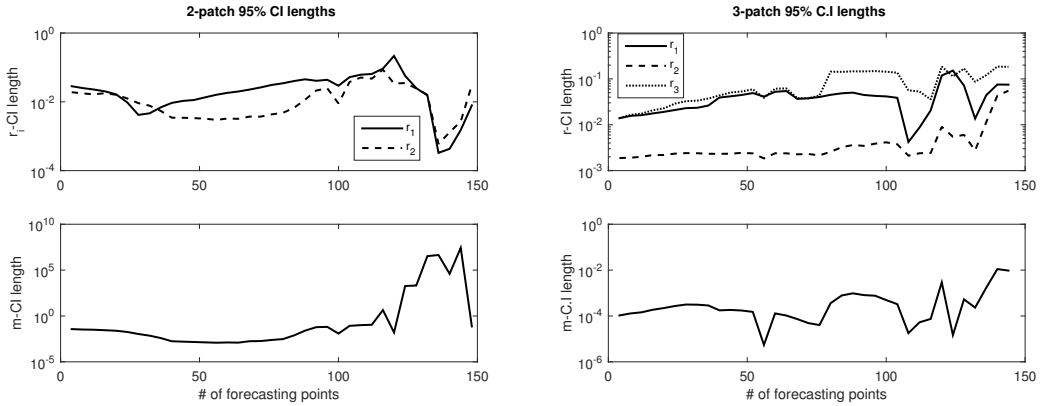

Figure 2.6: 95% CI for r_i and m for $i = 1, 2, 3$ from 2.6 for Port Loko. (Left) two-patch confidence interval lengths for intrinsic infection rate and migration parameter. (Right) Three-patch 95% confidence interval lengths for intrinsic infection rate and migration rate. Note the variability for high numbers of prediction points for both models and the high variability in r_1 for the two-patch model for low numbers of prediction points.

	(Althaus, 2014)	(WHO Ebola Response Team, 2014)	Logistic	2-Patch (H)	3-Patch (H)
Guinea	1.51 (1.50-1.52)	1.71 (1.44 - 2.01)	1.252 (1.249,1.255)	1.52 (1.42, 1.92)	1.45 (1.39, 1.51)
Liberia	1.59 (1.57-1.60)	1.83 (1.72 - 1.94)	2.11 (2.07,2.15)	1.45 (1.12, 1.94)	1.43 (1.06, 2.199)
Sierra Leone	2.53 (2.41-2.67)	2.02 (1.79 to 2.26)	2.28 (2.25, 2.32)	2.27 (2, 2.62)	2.12 (1.87, 2.26)

Table 2.4: 95% confidence intervals for R_0

Leone. In table 2.4 we provide estimations and 95% confidence intervals for R_0 for the three models. We further provide plots of R_e in Fig. 2.7.

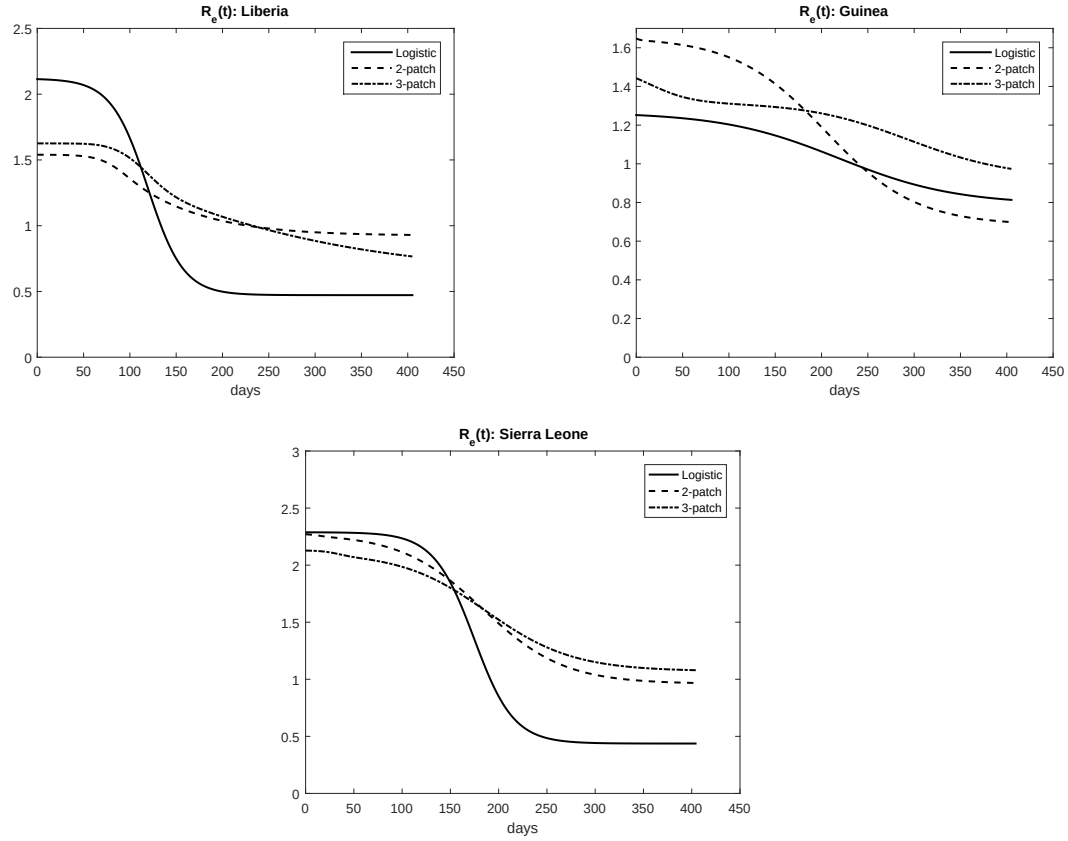

Figure 2.7: Effective reproduction numbers for (a) Liberia: simulations start at June 10, 2014, (b) Guinea: simulations start at March 25, 2014 and (c) Sierra Leone: Simulations start at May 27, 2014.

2.5 Discussion

In this chapter, a family of logistic patch models were preliminarily evaluated for use in disease modeling and forecasting. An explicit formula for the cumulative number of infectious individuals was derived from a SI compartmental model which takes the form of the well known logistic model. This derivation follows from the behavior change assumption, Eq. (2.2). We then extended the logistic model to include spatial population heterogeneity by using multi-patch models that incorporate

migration between patches and logistic growth within each patch. Each model's ability to forecast epidemic data was assessed by comparing model forecasting error, parameter distributions and parameter confidence intervals as functions of the number of data points used to calibrate the models. The patch models show an improvement over the logistic model in short-term forecasting, but naturally require the estimation of more parameters from limited data.

The models were tested by fitting them to the total reported case data from 39 districts in West Africa. In particular, the means of the weighted, relative and absolute errors of the patch models are less than the logistic model's, suggesting that spatial structure improved the data fitting. Next, models were compared by their forecasting capabilities in two ways: comparing forecasting error and comparing parameter confidence intervals. These latter efforts were restricted to the logistic, two-patch and three-patch models with homogeneous migration. The forecasting errors from Fig. 2.3 show that the patch models forecast better than the logistic model. However, Fig. 2.4 shows long-term forecasting variability from the patch models, because of the limited data. In contrast to these results, the Kruskal-Wallis test showed no significant difference in the forecasting errors across the models.

The value of R_0 during the outbreak in Liberia, Guinea and Sierra Leone were estimated to be in the same range as previous studies that were based on compartmental models Althaus (2014); Gomes *et al.* (2014); Khan *et al.* (2015); Yamin *et al.* (2015). In particular, from Table 2.4 the estimates from the two and three patch models for R_0 are similar with Althaus et al., but our confidence intervals are not as small Althaus (2014). This agreement further supports the reliability of the logistic and patch models with homogeneous migration.

In reality, early in the Ebola 2013-15 epidemic, the public's behavior in Liberia, Sierra Leone and Guinea did not swiftly change in a manner that mediated disease

transmission nor has there been any evidence supporting that the per-capita infection rate decreased linearly. Actually, the public's misunderstanding of the disease, lack of resources and fear fostered high-risk behaviors and resulted in an increased disease transmission in West Africa during the epidemic World Health Organization (2015b); Nielsen *et al.* (2015). However, health-care workers supplied valuable public awareness programs and medical resources that helped manage the spread. Our modeling assumptions approximate these notions and provide immediate behavior change in the spirit of Eq. (2.2), but this is modeled simultaneously everywhere in space and is one reason why the logistic model does not fit the data well. The patch-models overcome this issue by modeling behavior changes at different times, rates and locations, but require more data to be constrained.

Although the patch models performed better in many ways than the logistic equation, the patch models are not well constrained and parameter identifiability is an issue. Hence a direction for future work would be to assess parameter identifiability of the patch models and obtain mobility data for constraining the migration parameters. Identifiability analysis can be implemented by many different approaches, such as the Differential Algebra Identifiability of Systems (DAISY) (Saccomani *et al.*, 2003), the Exact Arithmetic Rank (EAR) (Karlsson *et al.*, 2012) or the Profile Likelihood (PL) approached as proposed in (Raue *et al.*, 2009). Work by (Raue *et al.*, 2014) provides a comparison and review of these methods. Similar work can be done with parameter sensitivity by using methods such as the Morris and Sobol' methods and Latin hypercube sampling-partial rank correlation coefficient (LHS-PRCC), to find which parameters are most sensitive and should be restricted.

Further work can be done with between-country and between-district scales. The latter would allow for more parameter constraint, but would have to be restricted to a small number of patcheshe number of parameters increase quickly as more patches

are introduced. that represent a small number of neighboring districts. The problem with incorporating all districts is that it ultimately requires a high-dimensional patch model with many parameters on a complicated network. This may be remedied with a partial differential equation model or by using mobility data to constrain the migration parameters. In addition, exploring different behavior functions would be another direction to expand this work.

Although the logistic model is phenomenological, it is capable of fitting the sigmoid curves that usually result from plotting the cumulative reported cases of disease outbreaks. The logistic and the patch models provide a general framework for disease modeling, because they do not model specific disease transmission processes. Specifically, they are based on two fundamental mechanisms that influence disease outbreaks: behavior change in the community and movement of individuals within that community. We find that incorporating the latter mechanism decreased forecasting errors with respect to the logistic model, but also require more data for model calibration.

Chapter 3

USING PHENOMENOLOGICAL MODELS FOR FORECASTING THE 2015 EBOLA CHALLENGE

3.1 Introduction

In this chapter, we use two simple phenomenological models to forecasts synthetically derived data from the 2015 Research and Policy for Infectious Disease Dynamics (RAPIDD) Ebola challenge. Our modeling approach is guided by data from a synthetically generated epidemic.

The RAPIDD Ebola challenge was designed to test the forecasting ability of mathematical models during an epidemic in real-time. Goals of the contest included (www.ebola-challenge.org, 2016)),

1. Improving predictive capabilities for future emergencies
2. Guiding the implementation of control measures
3. Illustrating how data quality and availability affect prediction accuracy

In this spirit, synthetic epidemic data was generated by a modified version of the model published by Merler et al., a detailed individual-based model of Ebola transmission (Merler *et al.*, 2015). Synthetic epidemic data was released at five different time points with a test release on Sept. 18, 2015. Five batches of data were released during the contest at time points: Oct. 6, Oct. 25, Nov. 11, Nov. 24 and Dec. 20, 2015, respectively. Model predictions were due two weeks later after each time point.

A diverse set of mathematical models comprising a total of seven independent teams competed in the contest. Models ranged from simple phenomenological to

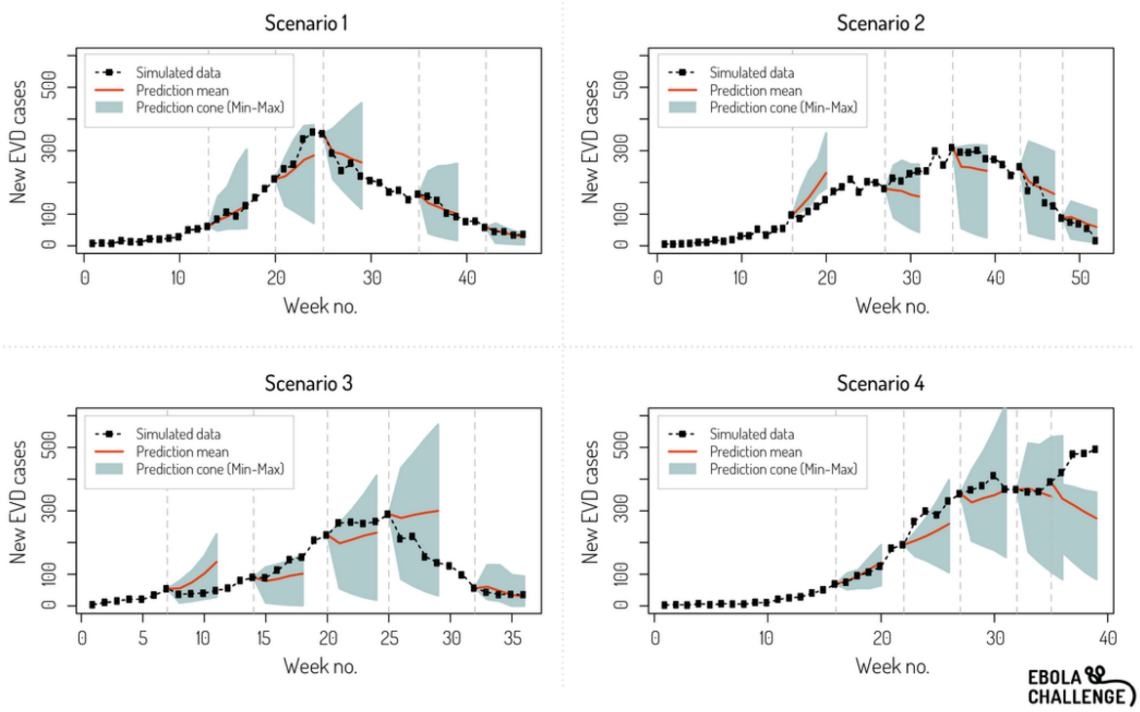

Figure 3.1: Ensemble results from the participating teams. Simulated data (black squares), ensemble prediction mean (red line) across all model predictions, and prediction cone (shaded gray) are shown. Results are shown for all time points and all scenarios. Taken from (www.ebola-challenge.org, 2016).

complex mechanistic models, with the goal of assessing and estimating key epidemiological quantities. During the challenge, we used the logistic model to forecast future cases of the epidemic given by the Ebola-challenge. Fig. 3.1 shows mean ensemble results. For more detailed information about the data see sub-section 3.2.2.

In this chapter we carry out a systematic comparison of simple phenomenological models namely the logistic growth model and the generalized-growth model that incorporates a flexible range of epidemic growth profiles including early sub-exponential and exponential growth epidemics in the context of the Ebola challenge based on synthetic data derived from a detailed individual-based model of Ebola transmission. Specifically, we assess the reproduction number and forecasts of the epidemic trajectory and the final epidemic size.

3.2 Materials and Methods

3.2.1 Model Description

The well-known logistic growth model was previously employed for epidemic forecasting the 2015 Ebola epidemic Chowell *et al.* (2014a), and was the model originally employed by the Arizona State Team (BP & YK) during the 2015 Ebola Challenge. This simple model is given by the following differential equation:

$$C' = rC\left(1 - \left(\frac{C}{K}\right)\right), \tag{3.1}$$

where $C'(t)$ models the rate of change in the number of new cases at week t. The logistic growth model relies on two parameters, the intrinsic infection rate, r, and the final epidemic size K.

For comparative purposes, we also analyzed the performance of the generalized Richards model Chowell *et al.* (2016a), which has been recently devised in order to capture the possibility of early sub-exponential growth epidemics and is given by:

$$C' = rC^p\left(1 - \left(\frac{C}{K}\right)^a\right). \tag{3.2}$$

The GRM is an enhanced version of the Richards model (Richards, 1959) by incorporating the generalized-growth model (GGM; $C' = rC^p$) (Viboud *et al.*, 2016). Specifically, the GRM incorporates a deceleration of growth parameter p to model a range of early epidemic growth profiles ranging from constant incidence ($p = 0$), polynomial ($0 < p < 1$) and exponential growth dynamics ($p = 1$). The GRM model was recently employed to generate forecasts of the Zika epidemic in Antioquia, Colombia (Chowell *et al.*, 2016a). All parameter values are positive: r is the growth rate, K is the final epidemic size, and a is a parameter that modulates the peak timing.

3.2.2 Data

As discussed above, synthetic epidemic data was generated by a modified version of the model published by Merler et al. that was calibrated for an EVD outbreak in Liberia (Merler *et al.*, 2015). Synthetic epidemic data was released at five different time points with a test release on Sept. 18, 2015. Five batches of data were released during the contest at time points: Oct. 6, Oct. 25, Nov. 11, Nov. 24 and Dec. 20, 2015, respectively. Model predictions were due two weeks later after each time point. During the contest, we only used the country level incidence time series data for our predictions (see 3.1). Basic demographic information was made available during the challenge in the form of age dependent infection probabilities, proportion of household sizes within Liberia and the population of infected counties. Contained in each of the five batches of released data, four scenarios representing different epidemiological conditions, behavioral changes, intervention measures and data availability were prepared for use in forecasting the epidemic (www.ebola-challenge.org, 2016):

- Scenario 1 consisted of the ideal data-rich scenario, in which district locations of newly reported EVD cases and Ebola treatment units (ETUs) were disclosed.

- Scenario 2 partially withheld location information of new EVD cases and ETUs and had a slightly later deployment of safe burial protocols and ETUs than scenario 1, at weeks 20 and 18 respectively.

- Scenario 3 took action the quickest by deploying an ETU and initiating contact tracing 20 weeks into the epidemic followed by safe burials by week 25.

- Scenario 4 was the worst-case scenario in which the disease showed no sign of slowing even with deploying a small ETU at week 18 of the epidemic, enforcing

safe burials 20 weeks into the epidemic, initiating contact tracing at week 16 and deployment of future ETUs.

Each scenario dataset contained outbreak situation reports, transmission tree data and weekly reported new EVD cases at the county and country level. All scenarios included a 20-30% noise level applied to incidences and missing info in patient-level data. New EVD cases were forecasted at one, two, three and four weeks past each time point, see Figure 3.1.

3.2.3 The Generation Time

The generation time is defined to be the time between infection in an index case patient and infection in a patient infected by that index case patient (WHO Ebola Response Team, 2014). We used transmission tree data (www.ebola-challenge.org, 2016). that was made available for scenarios 1, 3 and 4 to derive their generation time distributions respectively. In particular, given patient x, we found their infector, patient y, and calculated how long it took for patient y to infect patient x. This was done for all recorded patients, except the initially infected ones. For scenario 2 we used estimations from scenario 1.

3.2.4 The Effective Reproduction Number

The effective reproduction number, $R_e(t)$, is defined as the average number of new infections generated by one infectious individual in the population at time t (Nishiura and Chowell, 2009). $R_e(t)$ was numerically evaluated by training each model on an increasing amount of data (Chowell *et al.*, 2016a,b) based on the discretized renewal equation [10,12] (Nishiura and Chowell, 2009; Fraser, 2007):

$$R_e(t_i) = \frac{I_i}{\sum_{j=0}^{i} I_{i-j}\rho_j} \tag{3.3}$$

where I_i denotes incidence at time t_i, ρ_j denotes the discretized probability distribution of the generation interval, which we assumed to be gamma distributed with a mean of 16 days (WHO Ebola Response Team, 2014) and the denominator represents the total number of cases that contribute (as primary cases) to generating new cases (as secondary cases) (Nishiura and Chowell, 2009).

3.2.5 Performance Statistics and Epidemiological Targets

All teams that participated in the challenge had their models assessed according to a predefined set of performance metrics, which were used to systematically compare forecasting performance across the participating models. All metrics were calculated using model predicted incidences and observed incidences (synthetic incidence data). Performance metrics included: R^2, Pearson's correlation coefficient, mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and the mean absolute percentage error (MAPE). Incidence targets consisted of incidence predictions (new EVD cases) at 1, 2, 3 and 4 weeks after the last observed time point for a given scenario (see Figure 3.1). The challenge assessed each team's model performance by comparing incidence targets and nonincidence targets using the metrics above. Nonincidence targets consisted of effective reproduction number, peak time, incidence at peak time and final epidemic size.

3.2.6 Uncertainty Method 1

During the Ebola challenge, incidence targets, effective reproduction number, final epidemic size and peak timing predictions were generated by employing MATLAB's (The Mathworks, Inc.) built-in function, LSQCURVEFIT, with the Levenberg-

Marquardt option to find optimized parameter values for the best fit solution of the logistic model to the cumulative reported EVD cases (Marquardt, 1963; Moré, 1978). During the challenge we consistently employed a residual bootstrapping method to obtain the 25th and 75th percentiles for parameter estimates that is described in (Pell *et al.*, 2016). In short, we fit the model once and randomly added the residuals back into the original incidence data to create a new data set. A new optimized parameter set was then obtained by fitting the logistic model to this new data set and then the process was repeated 2000 times.

3.2.7 Uncertainty Method 2

For model comparison, Equations (3.1) and (3.2) were fit to the reported incidence data using the built-in MATLAB function LSQCURVEFIT (The Mathworks, Inc.). With this method, confidence intervals for model parameters and epidemiological forecasting targets were constructed as in prior studies (Chowell *et al.*, 2009, 2007; Viboud *et al.*, 2016) by simulating 200 realizations of the best-fit curve using parametric bootstrap with a Poisson error structure. 95% confidence intervals were calculated by taking the 2.5 and 97.5 percentiles from the generated parameter distributions.

Incidence forecast estimations were generated by extending the 200 realizations of the best-fit trajectory of a model 4 weeks into the future after the forecasting time point. The 95% confidence bands for the incidence targets were constructed with the distributions of incidence predictions at each time point.

3.3 Results

Post-challenge incidence forecasting performance metrics are summarized in Table 3.1 and post-challenge incidence forecast trajectories are illustrated for all scenar-

ios in Figure 3.2. Using Uncertainty Method 2, the GRM model provided improved incidence target forecasts compared to the logistic model when the models were calibrated on an increasing set of incidence data. In particular, the GRM had lower mean RMS values in every scenario than the logistic model (see Table 3.1). For example, mean RMS decreased from 66.80 (logistic) to 48.39 (GRM) in scenario 2. Furthermore, the GRM performed better across all scenarios and time points than the logistic model. In particular, RMS averaged across all scenarios decreased from 78.00 (logistic) to 60.80 (GRM) (Table 3.1). Similar improvements were seen when taking the mean across all scenarios and time points for Pearsons R score and the mean absolute percentage error; Pearsons R score increased from .15 (logistic) to .36 (GRM) (an R score closer to 1 means better agreement with the incidence data) and the MAPE decreased from .38 (logistic) to .32 (GRM).

The GRM slightly outperformed the logistic model in scenario 1 with incidence RMS decreasing by 1.01% when averaging across all time points (Table 3.1). Additionally, the GRM had better agreement with the trend of incidence targets with the higher Pearson R score of .55 than the logistic models 0.33 (Table 3.1).

In scenario 2, the GRM displayed better performance than the logistic model with incidence RMS decreasing by 27.56% when averaging across all time points. As in scenario 1, the GRM showed better agreement with the trend of incidence targets with a higher Pearson R score (GRM: .51, logistic: .47) (Table 3.1).

Once again, the GRM displayed better performance in scenario 3 than the logistic model with incidence RMS decreasing by 11.68% when averaging across all time points. The GRM showed better agreement with the incidence targets with a higher Pearson R score than the logistic (GRM: .31, logistic: -0.10) (Table 3.1).

Scenario 4 displayed the biggest difference in incidence forecasting with the GRM outperforming the logistic model with a 32.36% decrease in incidence RMS when

averaging across all time points. Again, the GRM showed better agreement with the incidence targets with a Pearson R score of .36 compare to the logistic models -0.08 (Table 3.1).

We did not include time point 1 in our analysis for final epidemic size predictions, because of an insufficient amount of data for model calibration that did not constrain estimations of K in scenario 4. Considering time points 2-5 and scenarios 1-4, the overall uncertainty in the predicted epidemic size was reduced as more data was made available for model calibration, but the GRM achieved better coverage of the observed final epidemic size than the logistic. In particular, Figure 3.4 shows that 95% confidence bars of final epidemic size predictions provided by the GRM contained the true epidemic size 8 out of 16 times (50% success rate) and had an average MAPE of 0.30 across all scenarios. In contrast, the logistic model consistently underestimated the final epidemic size in all scenarios during time points 2-5 with an average MAPE of 0.31 across all scenarios and 95% confidence bars that never contained the epidemic size, see Figure 3.4.

Estimations of the generation interval assuming a gamma distribution yielded reasonably good fits, with mean generation times in the range of 11.9-17.1 days and variance in the range of 8.3-42.3 days across scenarios 1, 3 and 4.

Using the estimated mean generation time and variances from transmission tree data from scenario 1, 3 and 4 to calculate the effective reproduction number yielded overestimates. Most notable are the estimations by both models in scenario 4, where the variance was the largest at 23.7 days. Across all scenarios, the GRM performed better that the logistic with an MAPE of 2.37, while the logistic model had an MAPE value of 2.64. In contrast, estimates of the effective reproduction number provided reasonable predictions under the assumption of a gamma distributed generation interval with a mean of 16 days and variance of 8 days, see Figure 3.3 and Table 3.2.

In particular, the GRM again outperformed the logistic model with an MAPE of 0.08 compared to 0.10, averaged across all scenarios and time points.

Mean estimates of the deceleration of growth parameter (p) during the early growth phase derived by fitting the GGM to the first 10 weeks of the epidemic ranged from .35-.71, .51-.92, .26-.5 and .17-.85 for scenarios 1, 2, 3 and 4 respectively (Figure 3.5). The range of p for scenarios 1 and 3 support sub-exponential growth profiles with substantial uncertainty. In contrast, scenario 2 and 4 yield ranges of p that may potentially indicate near exponential growth, but also include values that support sub-exponential growth.

During the challenge the logistic equation coupled with Uncertainty Method 1 consistently underestimated the final epidemic size, peak timing and the number of cases at peak timing with an average MAPE of 0.49, 0.36 and 0.40 respectively (Figure 3.8). Estimations of the effective reproduction number showed similar behavior with an average MAPE across all scenarios of 0.22 (Figure 3.7). In contrast, quantitative improvements were seen with the logistic model by using Uncertainty Method 2 (Figure 3.7). For instance, the mean MAPE across all scenarios and time points of the effective reproduction number decreased to 0.10 with Uncertainty Method 2. Similarly, across all scenarios, incidence target RMS decreased from 177.83 to 78.00 using Uncertainty Method 1 and Uncertainty Method 2, respectively (see Table 3.3 and Table 3.1). Performance statistics for the logistic model using Uncertainty Method 1 are reported in Table 3.3 and should be compared with results from Uncertainty Method 2 in Table 3.1.

3.4 Discussion

To improve the understanding of real time forecasting of epidemics, we have assessed the forecasting performance of two simple phenomenological models using the

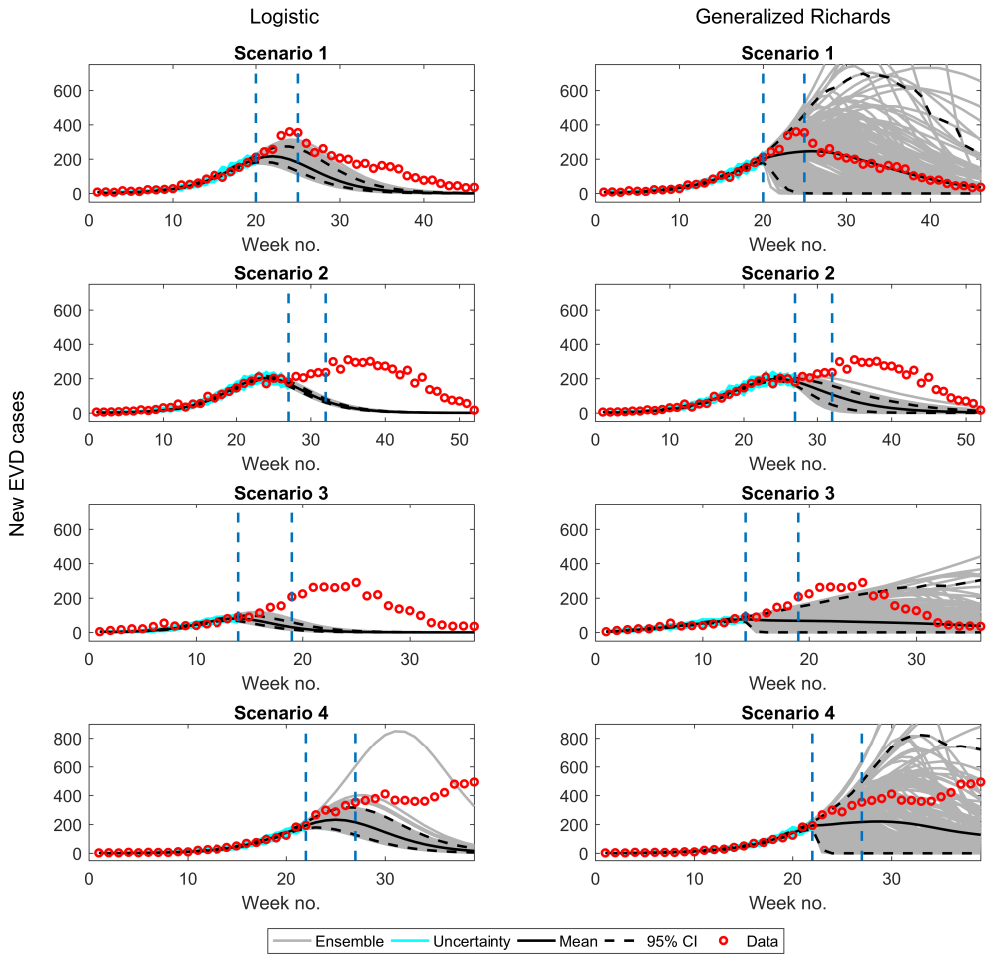

Figure 3.2: Epidemic forecasts based on the logistic (Equation 3.1; left column) and the generalized Richards model (Equation 3.2; right column) calibrated on epidemic data up to the second time point (left vertical line): 20, 27, 14 and 23 epidemic weeks for scenarios 1, 2, 3 and 4, respectively. The four data points (red circles) contained in the interior of the region bounded by the two vertical lines are the forecasting incidence targets used to calculate the statistics in Table 3.1. The mean (solid black line) and 95% uncertainty bounds for the calibrated model (light blue curves) and the 95% confidence bands (dashed black lines) of 200 forecasting ensembles (gray curves).

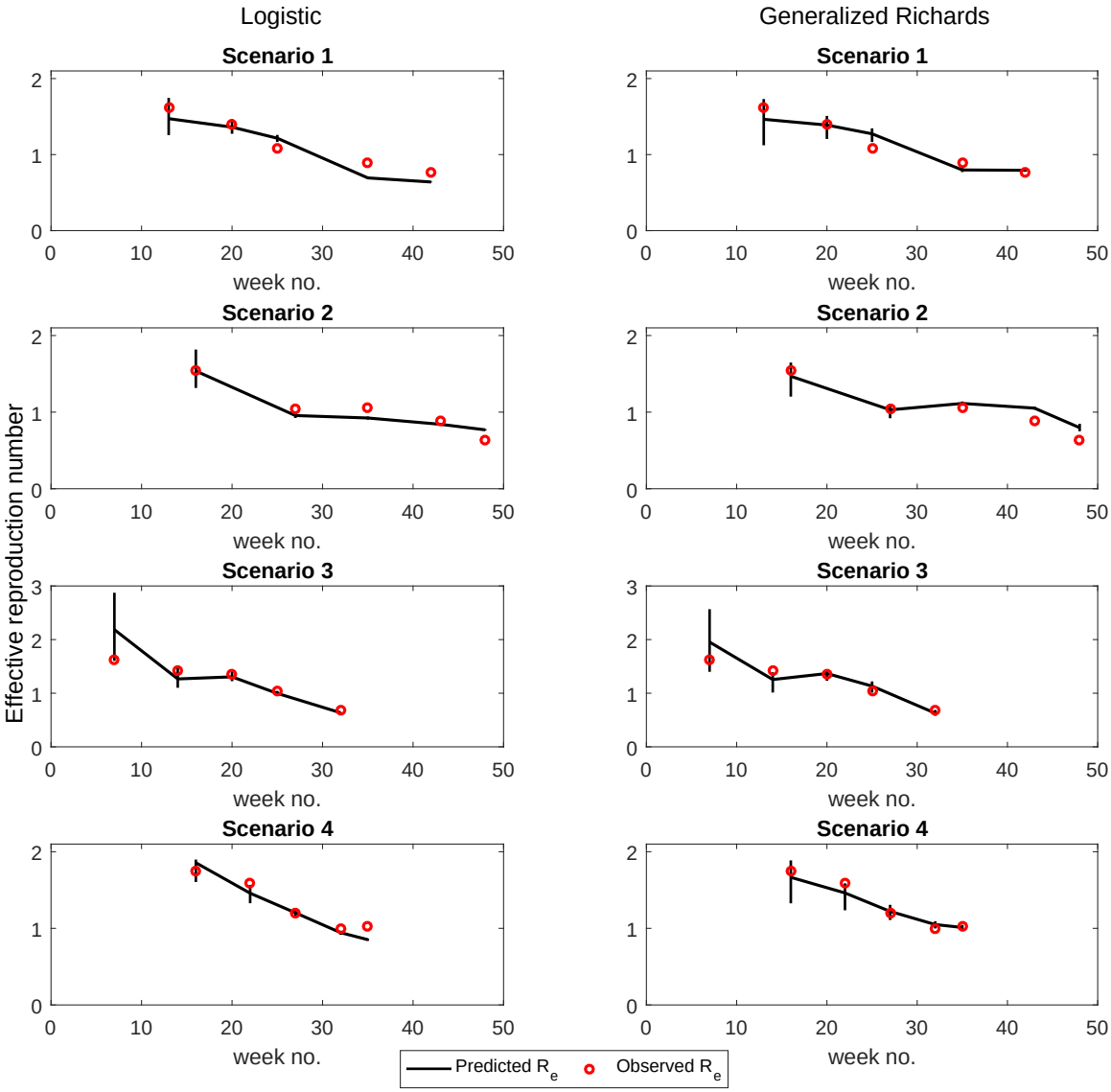

Figure 3.3: Mean estimates of the effective reproduction number from the logistic growth model (Equation 3.1; left column) and the generalized Richards model (Equation 3.2; right column) derived from Equation 3.3. Models provided reasonable forecasts among all scenarios (rows). Predictions for each scenario are obtained by fitting the corresponding model to an increasing amount of epidemic data: time points 1, 2, 3 and 4 respectively and using Equation 3 with a gamma distributed generation time with mean 16 days and variance of 8 days.

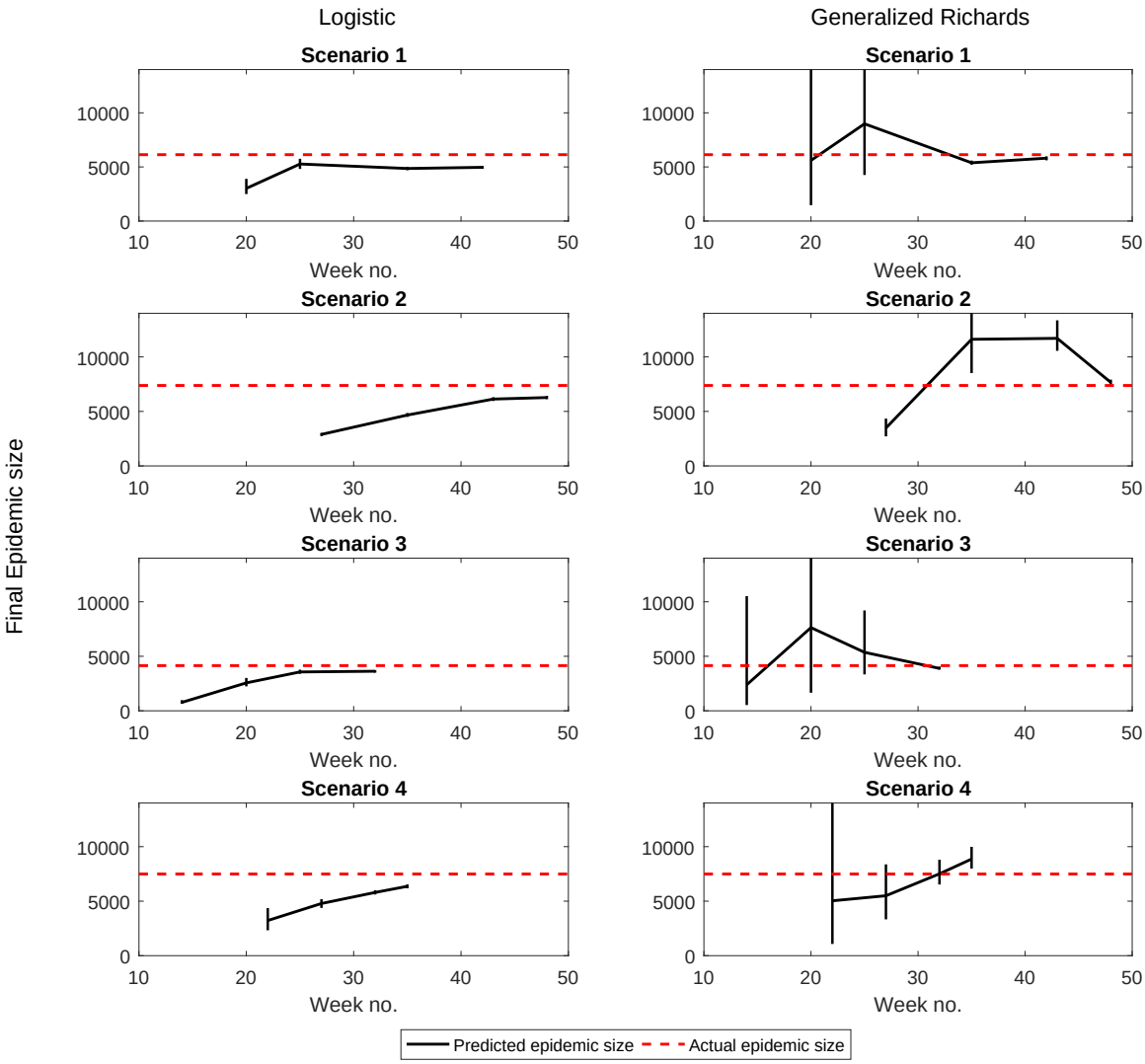

Figure 3.4: The Generalized Richards model (right column) provided improved forecasts over the logistic model (left column) of the expected epidemic size using data of the evolving epidemic at different time points namely 20, 27, 14 and 23 epidemic weeks.

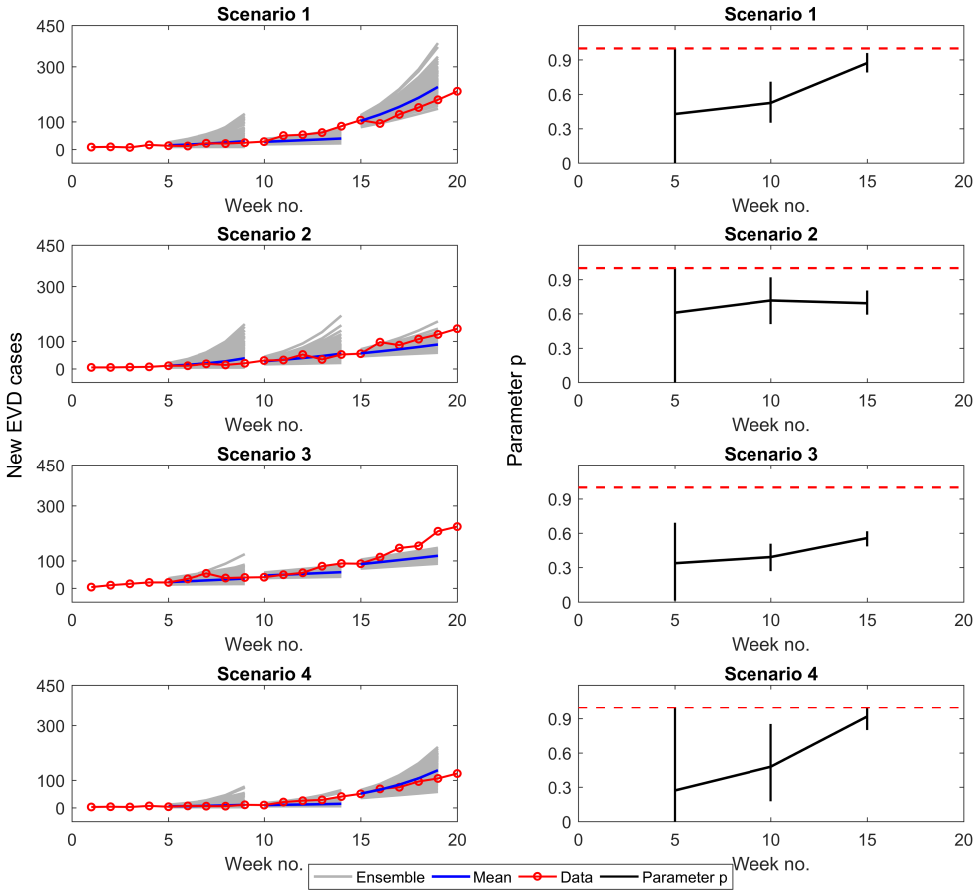

Figure 3.5: (left column) Short-term epidemic forecasts based on the generalized-growth model calibrated using an increasing amount of epidemic data (red line with circles): 5, 10, and 15 epidemic weeks for each scenario. The mean (blue solid line) of the model fit and forecasting ensembles (gray curves) are shown. (right column) Mean estimates and corresponding 95% confidence intervals of the deceleration of growth parameter, p, derived using the generalized-growth model fitted to an increasing amount of case incidence data: 5, 10, and 15 epidemic weeks.

Scenario	Logistic			GRM		
	Pearson's R	RMS	MAPE	Pearson's R	RMS	MAPE
Scenario 1						
Time point 1	-0.13	38.43	0.32	0.65	45.17	0.4
Time point 2	-0.79	105.02	0.28	0.95	80.95	0.22
Time point 3	0.76	73.27	0.29	-0.65	140.61	0.56
Time point 4	0.97	80.44	0.64	0.97	51.08	0.4
Time point 5	0.84	31.21	0.78	0.85	7.24	0.16
Scenario mean	0.33	65.67	0.46	0.55	65.01	0.35
Scenario 2						
Time point 1	0.98	17.16	0.11	0.98	34.62	0.24
Time point 2	-0.83	107.98	0.46	-0.85	78.89	0.33
Time point 3	0.58	143.75	0.49	0.81	28.14	0.09
Time point 4	0.72	51.79	0.28	0.75	80.9	0.52
Time point 5	0.91	13.37	0.44	0.9	19.41	0.64
Scenario mean	0.47	66.81	0.36	0.52	48.39	0.36
Scenario 3						
Time point 1	-0.89	21.16	0.51	0.89	19.48	0.45
Time point 2	-0.97	77.11	0.48	-0.97	62.15	0.41
Time point 3	-0.46	76.43	0.28	0.38	17.86	0.05
Time point 4	0.91	18.14	0.08	0.93	69.41	0.41
Time point 5	0.87	10.38	0.23	0.88	10.56	0.24
Scenario mean	-0.11	40.64	0.31	0.42	35.89	0.31
Scenario 4						
Time point 1	0.98	61.12	0.48	0.98	31.34	0.28
Time point 2	0.57	73.15	0.23	0.83	93.54	0.31
Time point 3	-0.13	89.53	0.21	-0.13	73.05	0.17
Time point 4	-0.94	171.94	0.4	-0.96	90.24	0.18
Time point 5	-0.9	298.75	0.61	-0.87	181.53	0.37
Scenario Mean	-0.08	138.9	0.39	-0.03	93.94	0.26
Mean across all scenarios	0.15	78.01	0.38	0.37	60.81	0.32

Table 3.1: Incidence performance statistics for the logistic growth model and generalized Richards equation.

synthetic incidence data generated from the 2015 Ebola Challenge. During the contest we employed the logistic equation to provide estimates of epidemic size, peak timing and the effective reproduction number using Uncertainty Method 1. The simplicity of this approach allowed us to provide fast estimates, but produced poor forecasting estimates when coupled with Uncertainty Method 1 because it failed to capture the uncertainty associated with the best fit to data. Our retrospective analysis indicates that improved uncertainty measures can be obtained using parametric bootstrap with Poisson error structure (Uncertainty Method 2). We compared the performance of the logistic model and the generalized Richards model calibrated with varying amount

Scenario	GRM	Logistic	Observed R
Scenario 1			
Time point 1	1.46	1.47	1.66
Time point 2	1.38	1.36	1.39
Time point 3	1.27	1.21	1.08
Time point 4	0.79	0.69	0.89
Time point 5	0.79	0.63	0.76
Scenario 2			
Time point 1	1.46	1.53	1.54
Time point 2	1.02	0.95	1.04
Time point 3	1.11	0.92	1.05
Time point 4	1.05	0.84	0.88
Time point 5	0.79	0.76	0.64
Scenario 3			
Time point 1	1.95	2.18	1.62
Time point 2	1.25	1.26	1.41
Time point 3	1.36	1.30	1.35
Time point 4	1.13	0.99	1.03
Time point 5	0.62	0.63	0.68
Scenario 4			
Time point 1	1.66	1.85	1.74
Time point 2	1.46	1.45	1.59
Time point 3	1.22	1.20	1.19
Time point 4	1.04	0.94	0.99
Time point 5	1.01	0.85	1.02

Table 3.2: Predicted and observed values of the effective reproduction number, using Uncertainty Method 2.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	All Scenarios
R2	-0.41	-2.36	-1.91	-13.73	-0.95
Pearson's R	0.72	0.58	0.55	-0.24	0.7
MSE	15172.34	26988.96	20969.51	63362.58	31623.35
RMSE	123.18	164.28	144.81	251.72	177.83
MAE	99.4	143.79	124.03	235.55	150.69
MAPE	0.64	0.75	0.79	0.6	0.7

Table 3.3: Summary of mean performance statistics of the incidence targets for the logistic growth model using method 1, during the challenge. The last columns averages values across all scenarios.

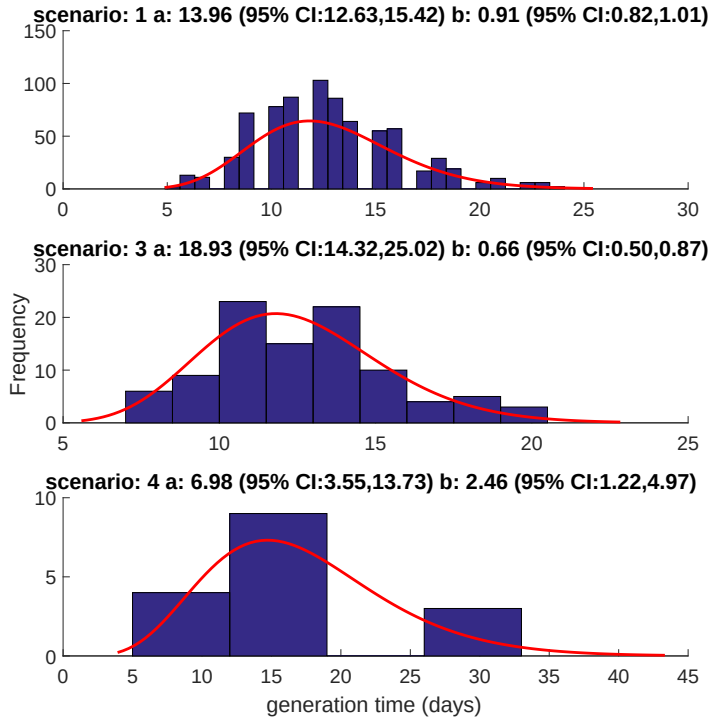

Figure 3.6: Distribution of generation times (blue bars) and the fitted gamma distribution with parameters a and b (red line) for scenarios 1, 3 and 4. Note that scenario 2 did not provide any transmission tree data.

of epidemic data. By changing the method used to model error in the best fit to data, we improved the performance of the logistic models ability to estimate the effective reproduction number. This highlights the sensitivity the calibration process can have on a models ability to estimate key quantities. Although, the logistic model coupled with Uncertainty Method 2 was an improvement, we saw an even further improvement when using the GRM. In particular, GRM obtained closer final epidemic size estimations with less data than the logistic. Finally, the logistic equation and the GRM provided similar estimates of the reproduction number and provided reasonably accurate results given their phenomenological nature.

Inclusion of the parameter p in the GRM is motivated by studies that have recently shown support for the presence of early sub-exponential growth dynamics (Viboud

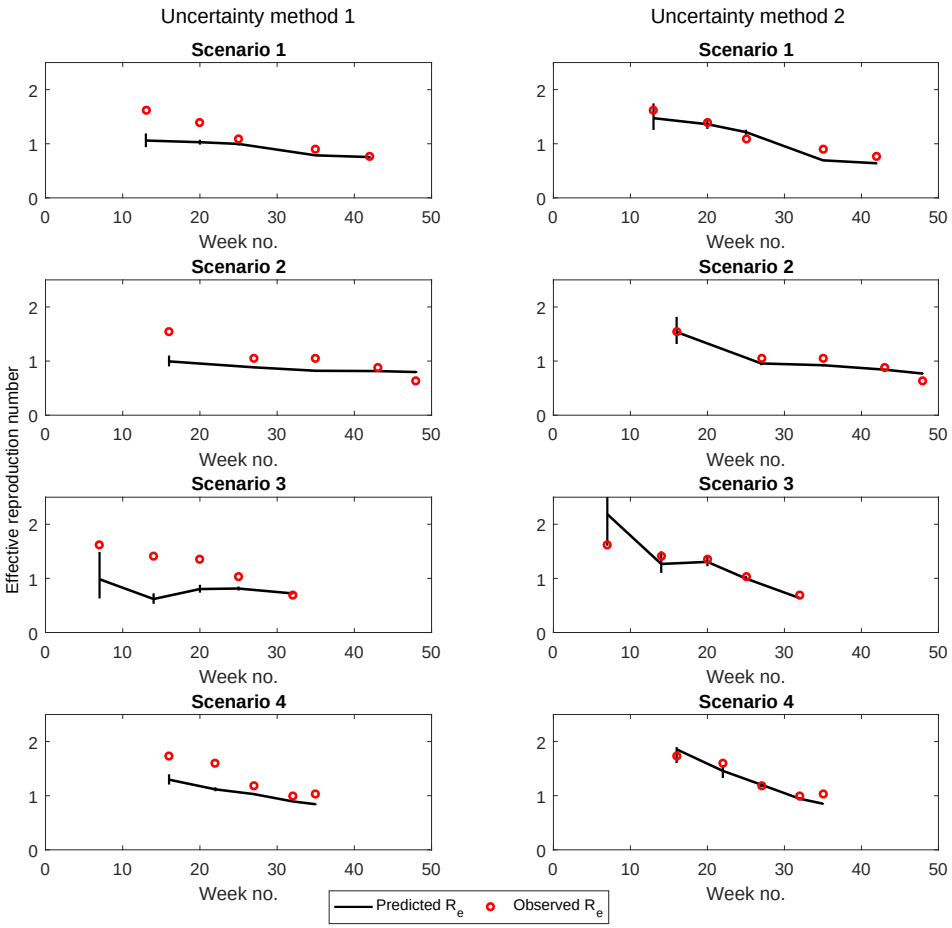

Figure 3.7: Means of the effective reproduction number $R_e(t)$ throughout the challenge predicted by the logistic growth model using Uncertainty Method 1 (left column) and Uncertainty Method 2 (right column). For each scenario, estimations are based on an increasing number of data points that were made available during time points 1-5. Observed effective reproduction numbers (red circles) are displayed for comparison.

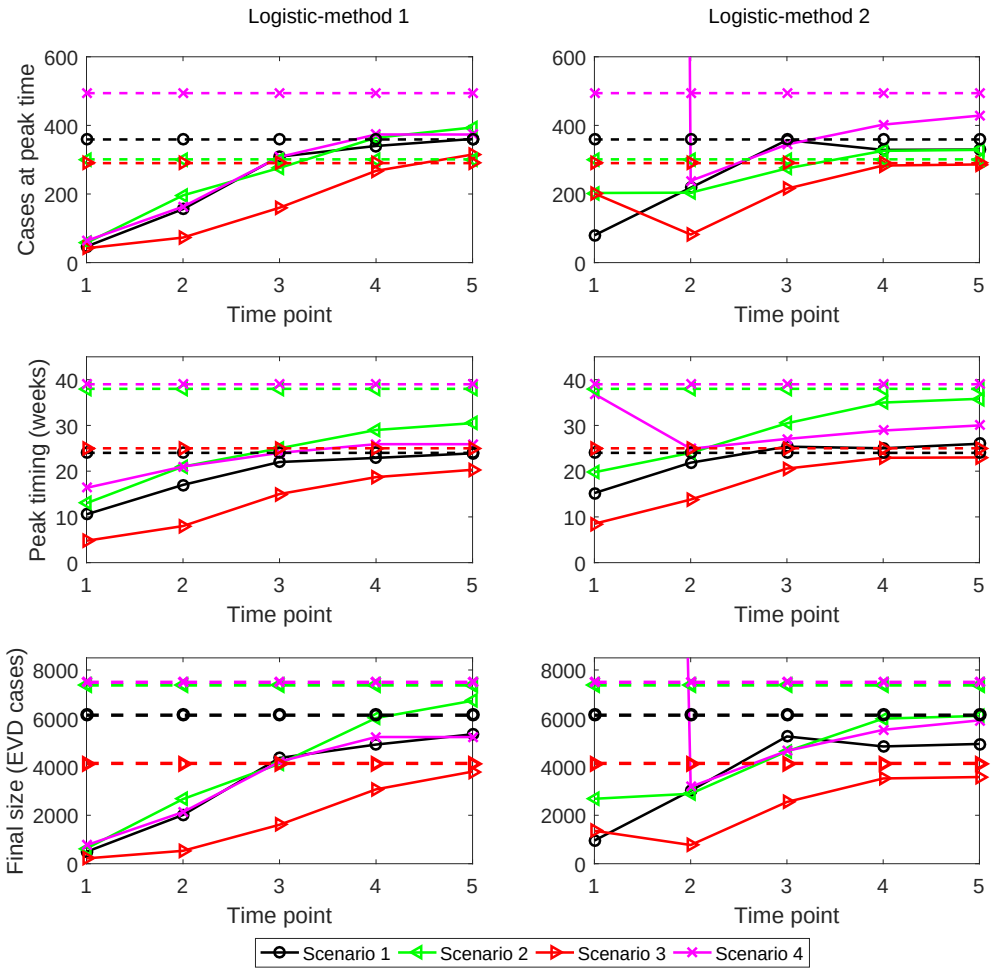

Figure 3.8: Means of the predicted (solid lines) epidemiological quantities and their true values (dashed lines) for the logistic growth model with method 1 (left column) and method 2 (right column).

et al., 2016; Chowell and Nishiura, 2015). In addition, the logistic model and GRM both assume that as more cases accumulate, the susceptible population is depleted. However, this phenomenological saturation effect in these models only becomes important during the later stages of the epidemic and could captures behavior changes, public health interventions and other disease prevention strategies that may take place during an evolving epidemic.

Because of a misunderstanding during the challenge, estimations were submitted for the basic reproduction number instead of the effective reproduction number.

Consequently, this led to incorrect predictions of the reproduction number during the challenge. Here we provide corrected results using Equation 3.3, which are displayed in(3.7 (left column) and summarize the rest of the predictions made by the logistic model during the challenge in Table 3.3 and Figure 3.8 (left column).

Our mean synthetic estimates of the reproduction number during the early epidemic growth phase are in broad agreement with published estimates of the reproduction number derived from real Ebola epidemics including for past outbreaks in Central Africa (Chowell *et al.*, 2004; Legrand *et al.*, 2007) and estimates derived for the 2014-15 Ebola in West Africa (Chowell and Nishiura, 2014; Althaus, 2014; Towers *et al.*, 2014; Fisman *et al.*, 2014; Nishiura and Chowell, 2014) or estimates based on a transmission tree of Ebola cases in Guinea during March-August 2014 Faye *et al.* (2015). Moreover, it is worth noting that our estimates of the effective reproduction number follow a declining trend during the early growth phase, a pattern that is in line with polynomial rather than exponential early epidemic growth dynamics (Chowell *et al.*, 2014b; Viboud *et al.*, 2016; Chowell *et al.*, 2016b). Polynomial epidemic growth could result from a number of factors including contact network characteristics (Funk *et al.*, 2010) and reactive behavior changes that gradually mitigate the transmission rate (Chowell *et al.*, 2014b). Simple phenomenological models composed of a small number of equations and parameters have shown promise in generating forecasts of epidemic impact based on early outbreak data (e.g., (Chowell *et al.*, 2014a; Nishiura and Chowell, 2014; Fisman *et al.*, 2014; Hsieh and Cheng, 2006)). For instance, the well-known logistic model provides a simple description of a single epidemic outbreak using only two parameters: the growth rate r and the final epidemic size K. However, a limitation of this and other models is the rigid assumption of early exponential growth dynamics. Using the logistic model, the exponential growth assumption was shown to work relatively well to describe and generate forecasts of the 2014 Ebola

epidemic in Liberia (Chowell *et al.*, 2014a), but it failed to provide a good fit to the early epidemic phase of the Ebola epidemics in Guinea and Sierra Leone (Chowell *et al.*, 2014a) where polynomial growth better characterized the early epidemic growth phase of the epidemic in those countries (Chowell *et al.*, 2014b). Our work here based on synthetic Ebola epidemic data derived from a detailed agent-based model (Merler *et al.*, 2015) and a recent analysis of a Zika epidemic in Antioquia, Colombia (Chowell *et al.*, 2016a) further emphasize the importance of designing models that reliably capture the epidemic growth phase of epidemic outbreaks in order to generate improved disease forecasts.

Reliably assessing a developing infectious disease outbreak as quickly as possible allows for policy makers to make swift and well informed decisions on the type and intensity of interventions that would be needed to ensure epidemic control. When substantial uncertainty surrounds the transmission, clinical, or epidemiological characteristics of the infectious agent encumbers the development of mechanistic transmission models that incorporate details about transmission modes, epidemiological stages, and effects of interventions, phenomenological models (e.g. (Chowell *et al.*, 2014a; Fisman *et al.*, 2013; Hsieh and Chen, 2009)) based on a few number of equation and parameters have the potential for providing a starting point to forecast epidemic impact (e.g. epidemic size), assess the early growth phase during the first few disease generations, and characterize the reproduction number, and represent a starting point towards a first response suite of mathematical models for addressing emerging infectious disease outbreaks.

Chapter 4

ANALYZING THE 1905-1906 PLAGUE EPIDEMIC IN BOMBAY

4.1 Introduction

Up until now, we have used mathematical models to assess Ebola virus disease. In this chapter, we change gears and assess the ability of simple phenomenological models to fit epidemic data from a completely different pathogen with a different transmission process. In particular, 1906 plague epidemic data from Bombay is used, because of its in-depth account that was recorded by the research commission established in Bombay during that time.

As discussed in chapter 1, Ebola virus disease is usually transmitted through direct contact with blood or body fluids of a person who is sick with or has died from EVD or through direct contact with infected bats or primates (Funk and Kumar, 2015; CDC, 2014). In contrast, plague was spread by infested fleas as well as exposure to the tissue and fluids of an animal infected with the plague (*Mus rattus* (house rat), *Mus decumanus* (field rat) and *Nesokia bandicota* (bandicoot)), but not entirely human to human.

We compare and contrast the abilities of three simple models to fit plague data for each of the twelve sections in Bombay during the year 1906. We use the basic logistic equation, Richards model (Richards, 1959) and the generalized Richards model.

In section 4.2, we introduce the mathematical models that we will be comparing and contrasting. Section 4.3 contains the results of best-fit parameter estimations. In section 4.4, we discuss our results and propose further directions.

4.2 Methods

4.2.1 Models

Two mathematical models are *nested* if one (the restricted model) is obtained from the other (the full model) by setting some parameters to zero (i.e. removing terms from the model), or some other constraint on the parameters.

We employ a family of nested models which for simplicity we'll call Model 1, Model 2 and Model 3:

$$C' = rC\left(1 - \left(\frac{C}{K}\right)\right), \tag{4.1}$$

$$C' = rC\left(1 - \left(\frac{C}{K}\right)^a\right), \tag{4.2}$$

and

$$C' = rC^p\left(1 - \left(\frac{C}{K}\right)^a\right). \tag{4.3}$$

Observe that Model 1 is a nested model of Model 2 and Model 2 is a nested model of Model 3. Note that Model 1 is the logistic equation, Model 2 is Richards equation and Model 3 is a generalized version of Richards equation. As in previous chapters of this dissertation we interpret r as the intrinsic rate of infection, a measures the temporal behavior change intensity, K as the final epidemic size and p $(0 < p \leq 1)$ as a deceleration parameter of the early epidemic growth phase (Viboud *et al.*, 2016).

4.2.2 Derivation of Model 1 and Model 2

The model derivation given in chapter 2 can be extended to derive Richards' model (Richards, 1959). As before, we make the assumption that the per-capita

infection rate is a decreasing function of the cumulative number of infected cases,

$$\frac{I'(t)}{I(t)} = f(C(t)) \tag{4.4}$$

In the following, we assume that $f(C(t)) = r(1 - (bC(t))^a)$ for some positive constants r, b and a. Hence

$$I'(t) = rI(t)(1 - (bC(t))^a) = \frac{r}{\beta}C'(t)(1 - (bC(t))^a).$$

Therefore,

$$I(t) - I(0) = \frac{r}{\beta}\left(C(t) - \frac{b^a}{a+1}[C(t)]^{a+1}\right) - \frac{r}{\beta}\left(C(0) - \frac{b^a}{a+1}[C(0)]^{a+1}\right).$$

Since $I(0) = C(0) \approx 0$, we see that $I(t)$ can be approximated by

$$\frac{r}{\beta}\left(C(t) - \frac{b^a}{a+1}[C(t)]^{a+1}\right).$$

Therefore

$$C'(t) = \beta I(t) = r\left(C(t) - \frac{b^a}{a+1}[C(t)]^{a+1}\right) = rC(t)\left(1 - \left(\frac{C(t)}{K}\right)^a\right), \tag{4.5}$$

where $K = \frac{(a+1)^{1/a}}{b}$. Here we interpret r as the *intrinsic infection rate*, b is a proportionality constant that corresponds to strength and effectiveness of disease interventions and preventive strategies, K is the *final epidemic size* and a corresponds to a nonlinear change in the behavior response which can be interpreted as the intensity of the behavior response as more cases are reported. Observe that Model 1 is a special case of Model 2 with $a = 1$.

4.2.3 Data

Data comes as reported death counts from (Commission *et al.*, 1907). The built-in MATLAB function LSQCURVEFIT was used to obtain parameters estimations for the best model fit.

4.2.4 Parameter Estimation

All parameters can be estimated by nonlinear least-square curve fitting to the. As in previous chapters, we use the built-in MATLAB function lsqcurvefit, to estimate r, a, p, and K. The initial number of cases $C(0)$ is fixed according to the first observation.

4.2.5 Statistical Comparison

To compare the best fits using Models 1-3, we perform two F-tests (Model 1 vs Model 2 and Model 2 vs Model 3). An F-test can be used to compare two nested models used to fit the same data set to determine whether the model with more parameters statistically improves the fit. The improvement is said to be statistically significant if the resulting p-value is less than .05. The F-ratio is defined as

$$F = \frac{(RSS_R - RSS_F)/(df_R - df_F)}{RSS_F/df_F}, \tag{4.6}$$

where RSS is the sum of squared residuals between model predictions and observed data. The subscripts F and R, denote the full and restricted model respectively. We note, that Model 1 is a restricted version of Model 2, Model 2 is a restricted version of Model 3. The degrees of freedom associated with RSS is $df = N - P$, where N is the number of data points and P is the number of fitted parameters. Lastly, to compute the associated p-value, we calculated the F-distribution evaluated at the F-ratio with $(df_R - df_F, df_F)$ degrees of freedom. Comparisons between models was performed individually for all districts.

4.3 Results

4.3.1 Best fit: Model 1 vs Model 2

We compared the best fits of using Model 1 and Model 2 by performing an F-test, which determines which one of the two nested models provides a better data fit from a statistical standpoint (Section 4.2.5). Results given in Table 4.1 show that Model 2 provides significantly better fits than Model 1 for 9 out of the 12 data sets (with the p-value$<$ 0.05). For the other data sets (districts 2,4 and 9), the F-test shows that there is a statistical trend that supports Model 1 (with the p-value ranging from .07 to .23). We have given the root mean square error (RMSE) Table 4.1 for further model comparison.

District	p-value	$RMSE_R$	$RMSE_F$
1	$<.05$	7.40	6.59
2	0.19	11.27	10.72
3	$<.05$	8.85	5.86
4	0.07	6.68	6.25
5	$<.05$	16.39	10.31
6	$<.05$	8.15	7.21
7	$<.05$	10.41	8.30
8	$<.05$	11.00	7.17
9	0.23	7.05	6.72
10	$<.05$	6.19	4.96
11	$<.05$	6.61	3.91
12	$<.05$	4.52	3.63

Table 4.1: Model comparison results for Model 1 vs Model 2

District	p-value	RMS_R	RMS_F
1	$< .05$	6.59	4.17
2	$< .05$	10.72	8.64
3	1.00	5.86	5.86
4	0.32	6.25	5.97
5	1.00	10.31	10.25
6	1.00	7.21	7.21
7	1.00	8.30	8.30
8	1.00	7.17	7.17
9	$< .05$	6.72	4.41
10	0.18	4.96	4.69
11	$< .05$	3.91	2.78
12	1.00	3.63	3.63

Table 4.2: Model comparison statistics for Model 2 vs Model 3

4.3.2 Best Fit: Model 2 vs Model 3

Results from comparing Model 2 and Model 3 support that Model 2 (the restricted model) is the better model. Indeed, results given in Table 4.2 show that Model 3 provides significantly better fits than Model 2 for only 4 out of the 12 data sets (with the p-value< 0.05), while Model 2 has 8 better fits out of the 12.

4.4 Discussion

A family of nested models based on the logistic equation was assessed for use is modeling the spread of plague. We extended work in Pell *et al.* (2016) by deriving the Richards model from a behavior change assumption and showed that the logistic is a special case when $a = 1$. Furthermore, we gave an interpretation of the parameter a as a measure of the nonlinear change in behavior. We introduced the generalized

Richards model and related the parameter p to a previous study where it was established as a measure of sub-exponential, exponential or super-exponential growth in the early phase of an epidemic depending on if $p < 1$, $p = 1$ or $p > 1$, respectively. In general, adding more complexity by incorporating the nonlinear behavior change into the Logistic model gave better fits to plague reported case data from the 1904 epidemic..

Statistical comparison using the F-test, showed that Model 2 is statistically better than Model 1 and Model 3. Indeed, Model 2 provides significantly better fits for 9 out of the 12 data sets than Model 1 and provides significantly better fits for 8 out of the 12 data sets than Model 3, see Figure 4.1. These results highlight the importance of the parameter a in the Richards model. Indeed, as explained above, a can be interpreted as the nonlinear behavior change of the population as a function of the number of reported cases. In contrast to parameter a, p is not important for modeling plague in the generalized Richards model, except when $p = 1$ which reduces the generalized Richards model to the Richards model.

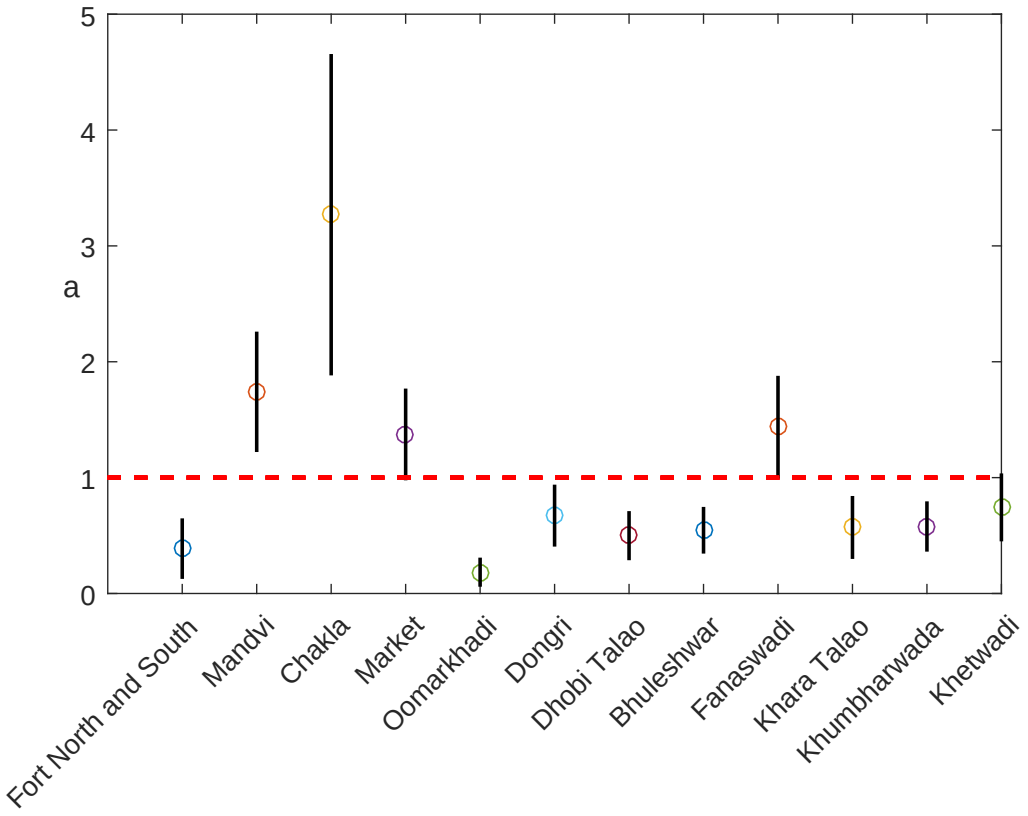

Figure 4.1: Confidence intervals for parameter a of the Richards model for all district data used. Values near 1 show evidence of a linear behavior change. The red dashed line represents 1.

Our results support near exponential growth during the 1904 plague epidemic in Bombay, as found in (Viboud *et al.*, 2016). However, Fort North and South show signs of sub-exponential growth ($p \approx .4$). One possible reason for this is the overall better quality of the buildings in this part of Bombay. Indeed, the report by the commission (Commission *et al.*, 1907), describes the houses here being built better than in the native quarter. Additionally, Fort North and South was surrounded by the esplanade, where plague was not commonly found and ultimately isolated Fort North and South from the rest of the city.

Our results also support the use of the very common assumption of exponential growth in more mechanistic models of plague. Interestingly, not all diseases have

shown this same characteristic (Viboud *et al.*, 2016).

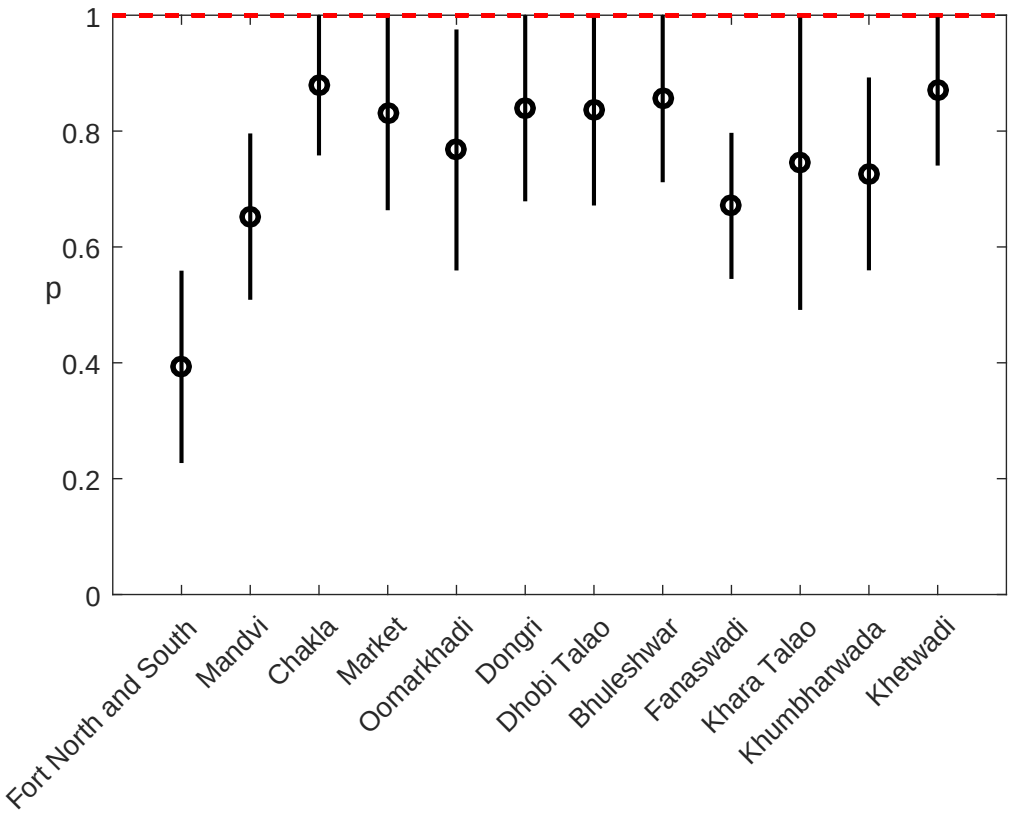

Figure 4.2: Confidence intervals for parameter p for the generalized Richards model for all district data used. Values near 1 show evidence of exponential growth. The red dashed line represents 1.

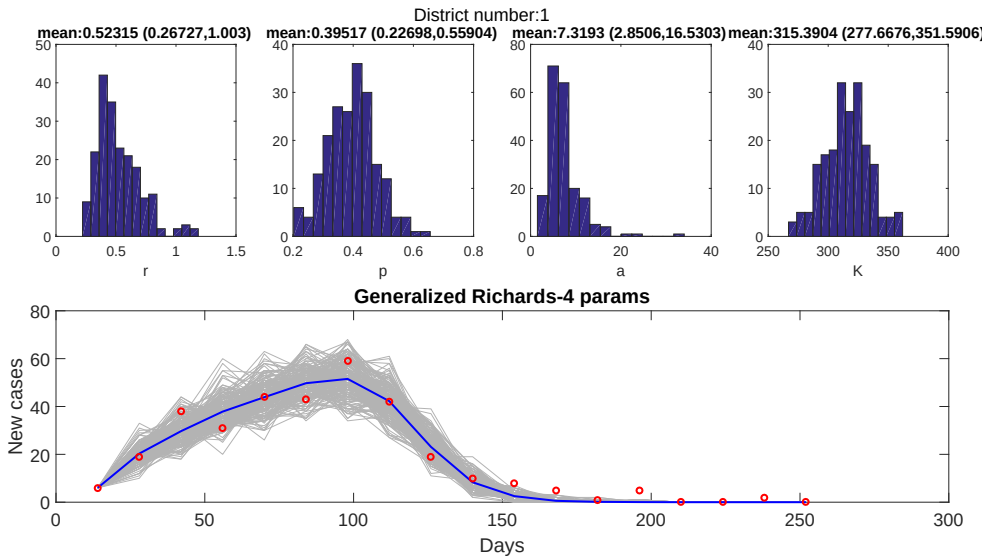

Figure 4.3: Best fit of the generalized Richards model to Fort North and South case data. Parameter distributions and 200 realizations of the best fit model solution generated from parametric bootstrap with Poisson error structure. Notice the sub-exponential growth in the early course of infection.

Chapter 5

THE INFLUENCE OF MIGRATION ON FINAL EPIDEMIC SIZE

5.1 Introduction

In this chapter, we numerically explore the impact that migration has on the final epidemic size and the effective reproduction number by using the 1905 Bombay epidemic as a case study.

The city of Bombay is a narrow island that is roughly 10 miles long and 3 miles across at its greatest stretch and roughly comprises 22.4 square miles. During the 1905-06 plague epidemic, Bombay, was joined to the larger island of Salsette by two causeways and two railway lines. Within the city, many districts existed that were all interconnected. Interestingly, Fort North and South, a district within Bombay, was mostly isolated from the other districts by a large barren field, called the Esplanade (which also was notably exempt from plague).

5.2 Methods

We wish to apply the three-patch model from chapter 2 to model the impact that the Esplanade has on the final epidemic size and the basic reproduction number during the 1905 epidemic. In general, we wish to study how the final epidemic size and the basic reproduction number changes as we change migration between the patches. With this in mind, we let Forth North and South be patch 1. From the map of Bombay city, we designate the second patch to be comprised of Mandvim, Chakla, Market, Oomarkhadi, and Dongri. Lastly, Dhobi Talao, Bhuleshwar, Fanaswadi, Khara Talao, Khumbharwada and Khetwadi. This partitions part of Bombay into

Param.	r_1	r_2	r_3	m_{12}	m_{21}	m_{13}	m_{31}	m_{23}	m_{32}	K_1	K_2	K_3
Value	0.088	0.062	0.189	0.054	0.004	0.010	0.001	0.607	0.601	333.357	2366.903	2520.110

Table 5.1: Resulting best fit parameter set.

three main patches that we have data for, see figure 5.1.

Next, we simultaneously fit each x_i to the corresponding incidence data that was described above to obtain a best fit parameter set. Throughout the exploration, we shall keep all parameters fixed to these values, unless we specify.

Recall the three-patch model from chapter 1, where x_i, r_i, and K_i represent the cumulative number of infections, intrinsic infection rate and final epidemic size in patch i. Lastly, m_{ij} is the rate at which cumulative infections travel from patch i to patch j.

$$\begin{aligned} x_1' &= r_1 x_1 \left(1 - \frac{x_1}{K_1}\right) - (m_{12} + m_{13})\, x_1 + m_{21} x_2 + m_{31} x_3, \\ x_2' &= r_2 x_2 \left(1 - \frac{x_2}{K_2}\right) - (m_{21} + m_{23})\, x_2 + m_{12} x_1 + m_{32} x_3, \\ x_3' &= r_3 x_3 \left(1 - \frac{x_3}{K_3}\right) - (m_{31} + m_{32})\, x_3 + m_{13} x_1 + m_{23} x_2. \end{aligned}$$

Let and $x = \sum_{i=1}^{N} x_i$. As with the derivation of R_e and R_0 for the logistic model above, define the basic reproduction number for an N-patch model as

$$R_e(t) = \exp\left(\hat{r} \int_t^{t+T} 1 - \frac{2}{\hat{K}} x(s)\, ds\right),$$

where $\hat{r} = \frac{\sum_{i=1}^{N} r_i K_i}{\hat{K}}$, $\hat{K} = \sum_{i=1}^{N} K_i$ are weighted averages and for simplicity we assume $T = 7$ days (Ali *et al.*, 2014). Similarly to above, we define $R_0 := R_e(0) \approx e^{rT}$.

Although we are not concerned with assessing the model fit to data, we present it in Figure 5.3 and present the best fit parameter values in Table 5.1.

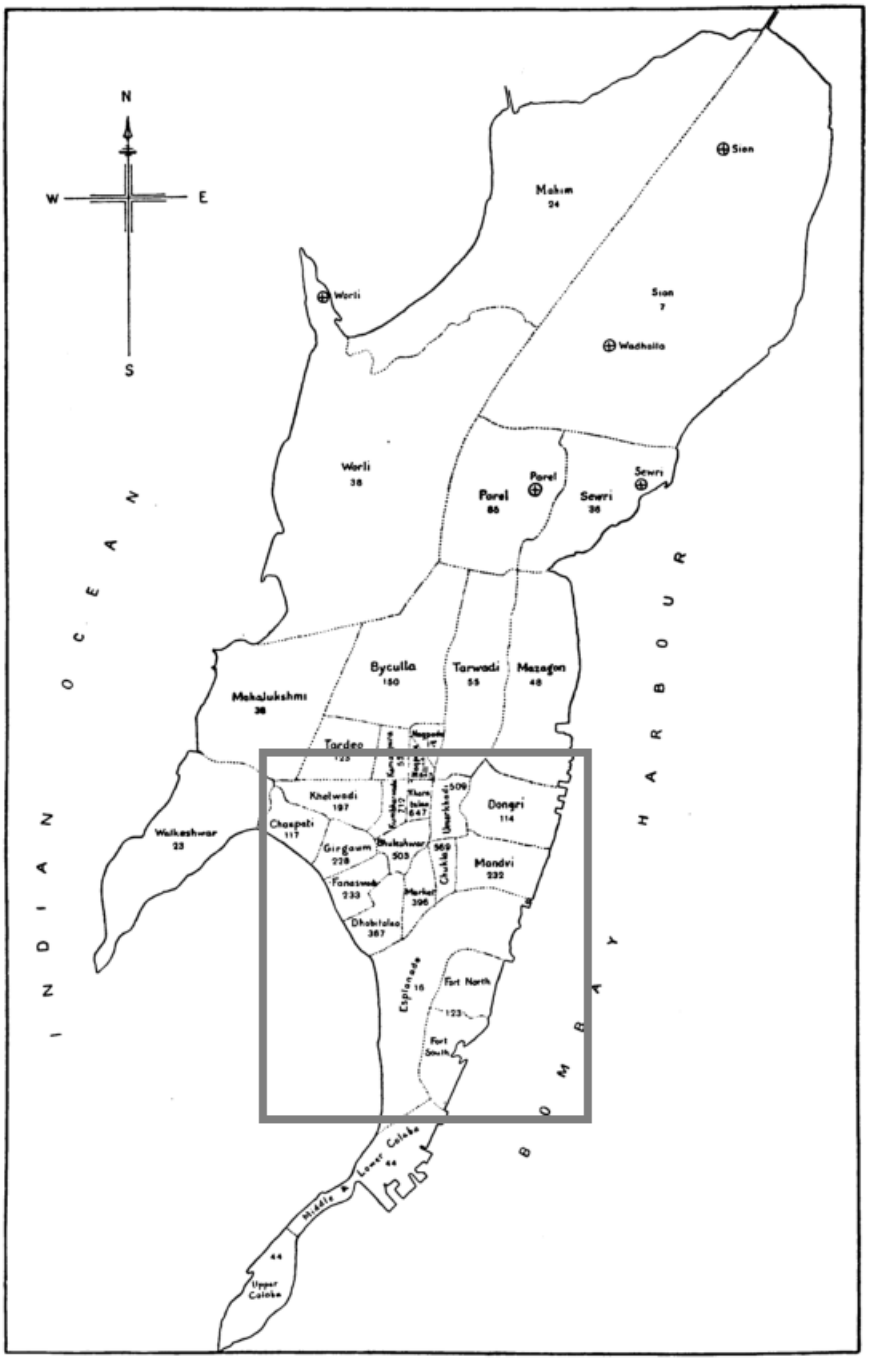

Figure 5.1: Map of Bombay. The rectangular region is the study area. Taken from (Commission *et al.*, 1907)

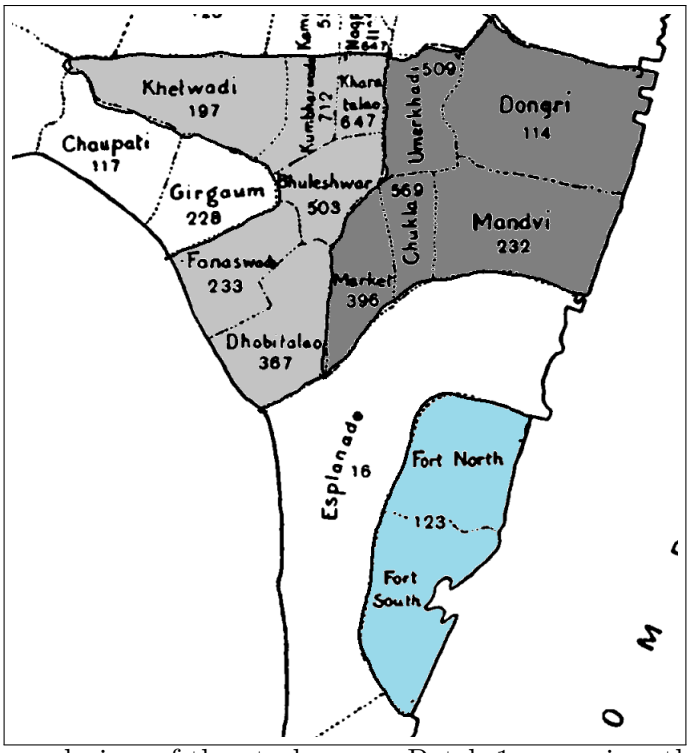

Figure 5.2: Enlarged view of the study area. Patch 1 comprises the light blue area, patch 2 comprises the dark gray area and finally, light gray represents the 3rd patch. Note how the Esplanade surrounds Fort North and South, creating a barrier between patch 1 and patch 2 and 3. The numbers show the population per acre.

5.3 Controlling migration in Bombay

Case 1 (migration into patch 1): As migration into patch 1 by the other two patches increases, the final epidemic size of the system decreases. That is, by sending people into the patch with smaller intrinsic infection rate (r_1) and smaller isolated final epidemic size (K_1), the overall final epidemic size will decrease (Figure 5.4; left).

Case 2 (migration out of patch 1): In contrast to case 1, by increasing migration out of patch 1 into the rest of the city (patch 2 and 3), the overall final epidemic size increases for migration values (m_{12} and m_{13}) near 0, but ultimately decreases the final epidemic size as migration increases. Simply put, the model predicts that by sending people into a more volatile patch does not have as large effect as to sending them to a less volatile patch (Figure 5.4; right).

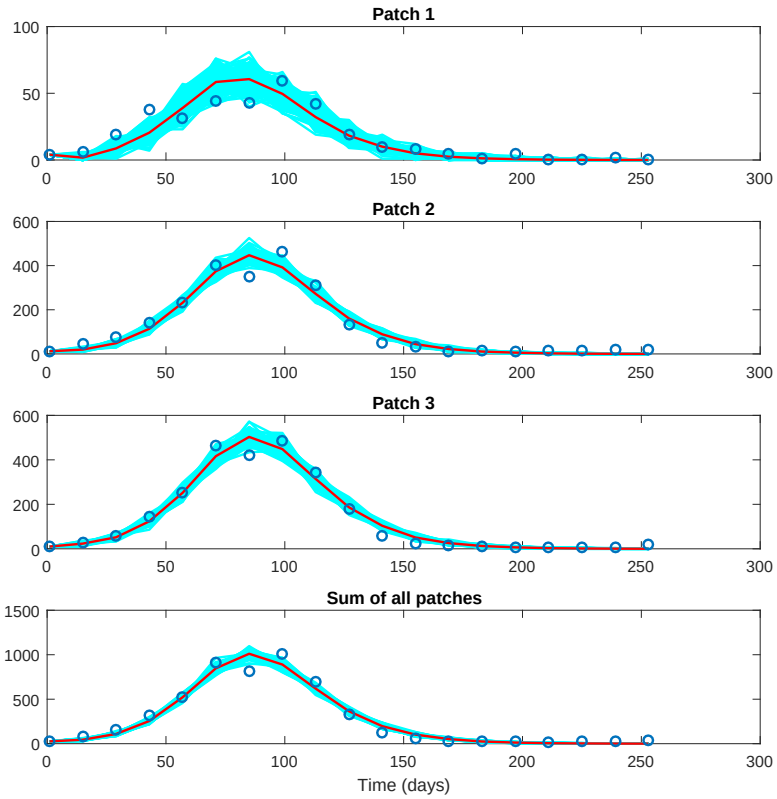

Figure 5.3: Model Fit to Incidence Data from Bombay. Patch 1 consists of Fort North and South. Patch 2 consists of Mandvim, Chakla, Market, Oomarkhadi, and Dongri. Lastly, patch 3 comprises, Dhobi Talao, Bhuleshwar, Fanaswadi, Khara Talao, Khumbharwada and Khetwadi.

Case 3 (migration between patch 2 and patch 3): This case models what happens to the final epidemic size, as migration is changed within the city (patch 2 and 3). We see that, when both patches have approximately equal migration rates (m_{32} and m_{23}), the overall final epidemic size increases above the value $K_1 + K_2 + K_3$ (the final epidemic size when all patches are isolated).

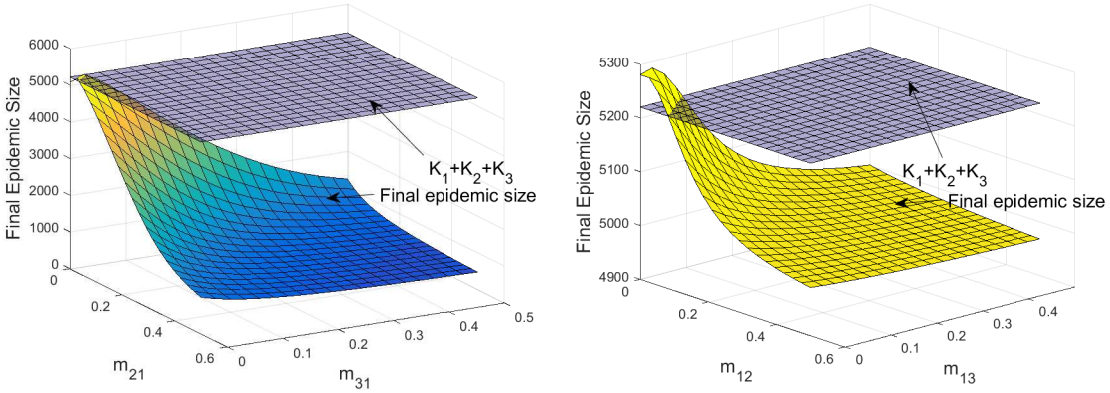

Figure 5.4: The effect of controlling migration rates in and out of Fort North and Fort South and how they influence the final epidemic size. (Left; migration into Fort North and South) In this case the model shows that the final epidemic size is a decreasing functions of m_{21} and m_{31}. (Right) The effect of letting people from Fort North and Fort South into the rest of the city can possibly increase the final epidemic size above what it would be if all patches were isolated from each other. In this case, migration does not have as big an impact as in the previous case.

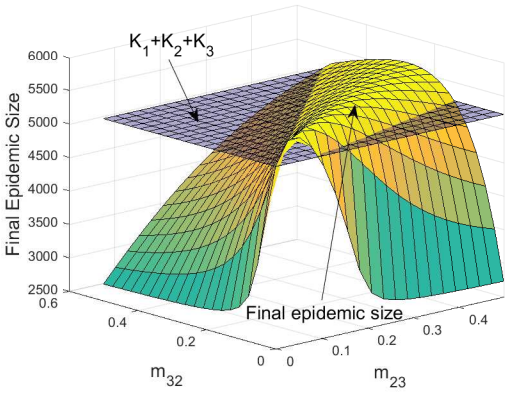

Figure 5.5: Final epidemic size as a function of m_{23} and m_{32}. The model predicts that it's possible for the overall final epidemic size to be larger than $K_1 + K_2 + K_3$.

5.4 Discussion

The numerical exercise reveals the role of migration and spatial structure in projecting disease burden. It further shows the insight that can be gained by modeling the cumulative number of infected individuals with patch models.

The effect of controlling migration rates in and out of Fort North and Fort South

and how they influence the final epidemic size. In the left plot shown in Figure 5.4, we see that the final epidemic size is a decreasing functions of m_{21} and m_{31}. In contrast, the right panel shows the effect of letting people from Fort North and Fort South into the rest of the city. Ultimately, the model predicts that this can possibly increase the final epidemic size above what it would be if all patches were isolated from each other. Lastly, Figure 5.5 shows that migration can have a positive effect on the final epidemic size.

In all of these bifurcations we can see the need for quarantine and travel control. Since the horizontal line represents when all three parts of the district are isolated from one another the bifurcation diagrams show the impact isolation protocols have on final epidemic size.

Chapter 6

WITHIN-HOST DYNAMICS OF CEREAL YELLOW DWARF VIRUS IN AVENA SATIVA

6.1 Introduction

In this chapter, we change gears from modeling at the population level and model within-host interactions of Cereal Yellow Dwarf virus-RPV (CYDV-RPV) from an experiment by Kendig et al. As noted in chapter 1, CYDV-RPV, is a plant virus that infects over 150 species of grasses and is vectored by *Rhopalosiphum padi* (the "RP" in RPV). Like previous chapters, the development and iteration of the two mathematical models presented here are motivated and guided by data. As we shall see, the best fit solutions from our models uncover patterns that otherwise would be hard to identify when viewing the data by itself.

After being injected by the stylus from an infected aphid, CYDV-RPV enters the phloem cells of its plant host, where it replicates. The phloem cells in plants are used to transport sucrose throughout the plant (Raven *et al.*, 2005) and help in spreading the virus. Phloem cells that make up grasses come in two general types: sieve tube elements (the cells that "transport" sucrose) and cells that support the sieve tube elements. This latter class of cells include companion cells and other specialized parenchyma cells.

Since CYDV-RPV infect sieve tube elements and ultimately damages or destroys them, they inhibit the movement of crucial resources and carbohydrates to the roots and leaves. The reduction of nutrients to these plant structures is the proposed mechanism by which CYDV-RPV reduces the growth of its plant host (Erion and

Riedell, 2012).

In an experiment by Kendig et al., *Avena sativa* (common oats) were inoculated with CYDV-RPV. Plant soil then underwent four different nutrient solution treatments to test the effects of nutrient concentration on virus dynamics within the plants. Plants were treated with a control solution (CTRL), a nitrogen addition solution (+N), a phosphorus addition solution (+P) and finally a treatment with both nitrogen and phosphorus addition (+NP), see Fig.6.1. To investigate the dynamics of the virions (virus particles), we turn to mathematical models.

6.2 Data and Questions

Fig. 6.2 shows interesting dynamics that change under different nutrient treatments. In the control experiment, virion concentrations remain low, until 18 days post inoculation, when the virion population grows quickly. Furthermore, the virion population in the control experiment is larger than in all other experiments. In the nitrogen addition experiment, virion population is dramatically reduced when compared to the control experiment. The phosphorus addition experiment also reduces virion population, but not as much as the previous case. Additionally, two subtle "bumps" before the high growth of the virion population can be seen in the CTRL data.

Motivated by the reduced virion population under different nutrient regimens compared to the control experiment and the "bumps", we focus on developing mathematical models motivated by the two following questions:

1. How to derive a within-host pathogen model where growth of the virus population is dependent on the resource nutrient concentration taken up by the plant?

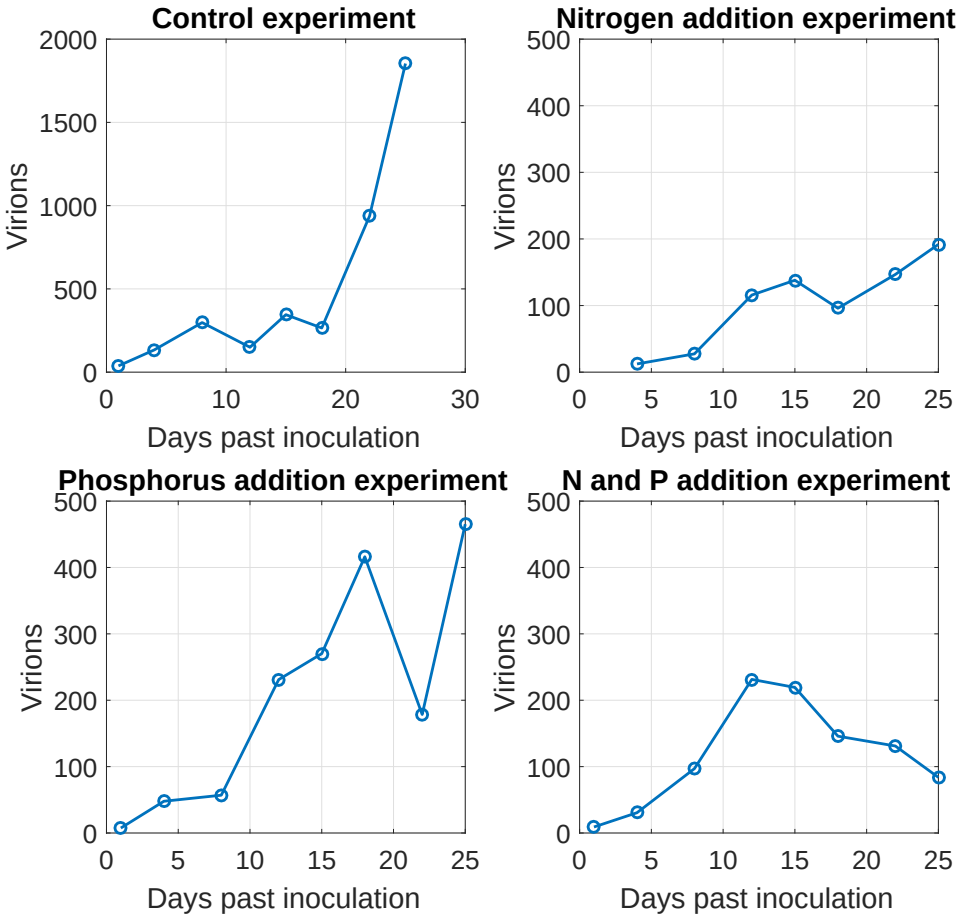

Figure 6.1: Virion data from plants grown under different nutrient solutions.

2. Can the shape of the data in from the different nutrient regimens be explained by a delay in virus production?

6.2.1 The Droop Equation

To tackle our first goal, we construct a mathematical model that describes the dynamics of susceptible cell, infected cell, virus populations and nutrient supply. In addition, as a first iteration model, we choose to only model one nutrient.

Since virions can only replicate within a host's cell, it is important to accurately describe the amount of nutrient in the host's cells. Nutrient-controlled growth rates

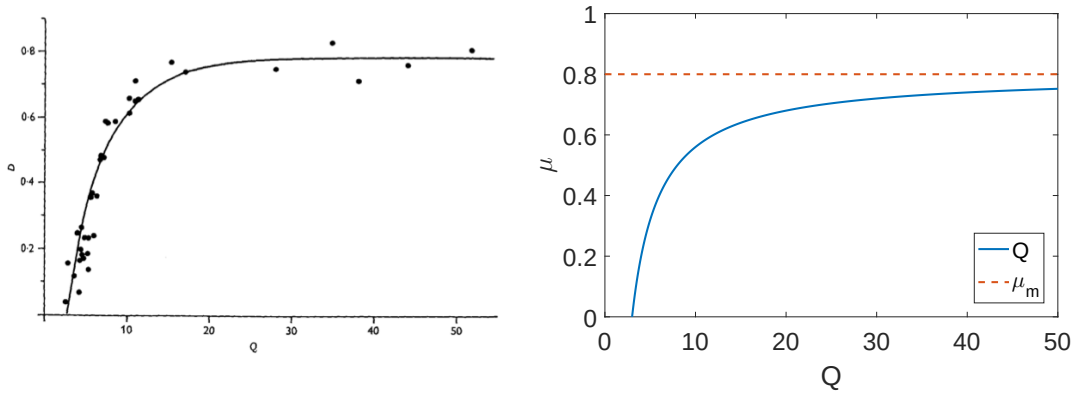

Figure 6.2: Relation between dilution rate (D) and cell quota (Q) where $\frac{D}{\mu}$ is assumed to be constant. M. Droop 1968 (left). Plot of Droop equation (Equation 6.1), $q = 3$, $\mu_{max} = .8$ (right). Note: it is useful to remember that dilution rate is equal to growth rate when the chemostat is at steady state.

can be employed via the Droop cell model (Droop, 1973a,b, 1974). Droop showed that there existed an empirical relationship between algal specific growth rate and the intracellular concentration of vitamin B_{12} inside the chemostat in which algal was being grown. Droop defined the cell quota, Q, as the total cell nutrient per unit biomass. He discovered a simple relationship between specific growth rate (μ) and the cell quota:

$$\mu = \mu_m \left(1 - \frac{q}{Q}\right), \tag{6.1}$$

where *the subsistence quota*, q, is interpreted as the minimum Q required for life and μ_{max} is the maximum specific growth rate. We present a comparison of Droop's original plot of the cell quota equation and Equation 6.1 in Figure 6.2.

Recently, the Droop equation has been applied to model dynamics of cancer growth (Portz *et al.*, 2012; Everett *et al.*, 2015, 2014) and has been used to derive the logistic equation (Kuang *et al.*, 2004a).

To generalize to n nutrients one could use Liebigs law of the minimum: an organism's growth will be limited by the resource that is in lowest supply with respect to

the organism's needs, to obtain a similar growth function,

$$\mu = \mu_m \min\left\{\left(1 - \frac{q_1}{Q_1}\right), ..., \left(1 - \frac{q_n}{Q_n}\right)\right\}.$$

6.3 Nutrient Growth Model Derivation

Let S and I be the number of susceptible and infected cells respectively and V be the number of virions. Since virus particles that cause cereal yellow dwarf disease are restricted to the phloem of host plants we let S represent a typical phloem cell. Consider the following model derivation with nutrient limited virus and cell growth.

6.3.1 Healthy Phloem Cells and Nutrient Uptake

We assume that in the absence of infection, the susceptible cells obey nutrient limited growth via the Droop equation Droop (1973b,a, 1974). That is, let N_S and N_f be the nutrient in the plant's cells and the free nutrient respectively, then the total amount of nutrient in the (closed) system is $N_t = N_S + N_f$. Let $Q = Q(t)$ be the plant's cell nutrient quota. With this, we may write down a governing equation for S,

$$\frac{dS}{dt} = \mu_m \left(1 - \frac{q}{Q}\right) S - mS,$$

where the maximum specific growth rate is μ_m and we have included a natural death rate, m, of the cells.

To formulate a governing equation for Q, we let the rate of change of free nutrient, N_f, change according to

$$\frac{dN_f}{dt} = -\alpha S N_f + DSQ, \tag{6.2}$$

where the first term approximates the loss of free nutrient by the uptake by cells, and the second term represents the nutrient that is released back into the environment

when cells die. Since $N_S = QS$ we obtain $N_t = N_f + QS$. Solving for Q and differentiating with respect to t and using the fact that $\frac{dN_f}{dt} = -\alpha N_f + DSQ$, we obtain

$$\begin{aligned}
\frac{dQ}{dt} &= \frac{-SN_f' - (N_t - N_f)\, S'}{S^2} \\
&= \frac{S\left(\alpha S N_f - DQS\right) - (N_t - N_f)\left(\mu_m\left(1 - \frac{q}{Q}\right) S - DS\right)}{S^2} \\
&= \alpha N_f - \mu_m\left(Q - q\right) \\
&= \alpha\left(N_t - SQ\right) - \mu_m\left(Q - q\right).
\end{aligned}$$

Therefore, we arrive at a simple model for nutrient cell growth in the absence of infection,

$$\begin{aligned}
\frac{dS}{dt} &= \underbrace{\mu_m\left(1 - \frac{q}{Q}\right) S}_{\text{growth}} - \underbrace{mS}_{\text{death}} \\
\frac{dQ}{dt} &= \underbrace{\alpha\left(N_t - SQ\right)}_{\text{Uptake}} - \underbrace{\mu_m\left(Q - q\right).}_{\text{loss from cell growth}}
\end{aligned} \tag{6.3}$$

Under the assumption that all parameters are positive, Everett showed this model exhibits a unique positive steady that is globally asymptotically stable (Everett, 2015).

6.3.2 Infected Phloem Cells

We assume that infection of healthy phloem cells is governed by mass action and release virus at rate δ. With these assumptions the governing equation for the rate of change of infected phloem cells is

$$\frac{dI}{dt} = \underbrace{\beta SV}_{\text{infection}} - \underbrace{\delta I.}_{\text{viral shedding}} \tag{6.4}$$

Due to the introduction of infected cells and virus particles, we include the cell

quota contained within infected cells and virus particle, $I\,(Q+\theta)$ and θ respectively in the equation for total nutrient,

$$N_t = N_f + SQ + I(Q+\theta) + \theta V. \tag{6.5}$$

Where we have assumed that the infected cell population have their growth mechanism hijacked by the virus and the amount of nutrient contained within a virus particle is equal to θ.

6.3.3 Free Virions

Our modeling approach as of now has been similar to work by Fuhrman et al., but we now diverge from their work by incorporating nutrient dependent virus growth (Fuhrman *et al.*, 2011).

Virions are relatively homeostatic and must process nutrients within the host to replicate. Furthermore, virions are simple in structure (a genome and protein capsid) and elemental composition. Thus we make the assumption that the cell nutrient inside of a virus particle is constant for all time and is equal to θ.

One way to model the growth function of the virion population is to simply assume that it also behaves like the Droop equation. That is

$$b(Q) = b_m \max\{\left(1 - \frac{q_v}{Q}\right), 0\}, \tag{6.6}$$

where q_v is the minimum amount of nutrient needed for virion production. In the case when $Q < q_v$, we assume that there is not enough nutrient for virus production and is therefore 0, which ultimately amounts to introducing the maximum function. In short, this function relates the growth rate of the virus ($b(Q)$) to the nutrient concentration inside the cells of the host plant (Q).

Free virions are produced at a maximum rate, b_m, per infected phloem cell per unit of time and are destroyed or cleared by the immune system at rate d. Free virions that are not destroyed or cleared are assumed to be absorbed by neighboring phloem cells.

$$\frac{dV}{dt} = \underbrace{b\,(Q)\,\delta I}_{\text{New virions}} - \underbrace{dV}_{\text{Cleared Virions}} - \underbrace{\beta SV}_{\text{absorption}} \tag{6.7}$$

6.3.4 Free Nutrient Equation

With the added infected cell and virus populations we obtain an updated equation that governs the free nutrient within the system, equation (6.5). As we did before, to find an updated governing equation for Q, we let the change of free nutrient change according to

$$\frac{dN_f}{dt} = -\alpha N_f(S+I) + mSQ + d\theta V + \delta I\hat{Q}, \tag{6.8}$$

where the first term is the loss of free nutrient due to cell uptake, the second and third are from the release of nutrient due to healthy cell death and virus death and the third term represents nutrient that is released when infected cells die, where,

$$\hat{Q} = \underbrace{Q}_{\substack{\text{Quota initially}\\ \text{in cell}}} + \underbrace{\theta}_{\substack{\text{quota from virions}\\ \text{in cell}}} - \underbrace{\theta b(Q)}_{\substack{\text{loss from produced}\\ \text{virions}}} + \underbrace{\frac{\mu(Q-q)}{\delta}}_{\substack{\text{production of}\\ \text{quota over length of}\\ \text{the production cycle}}}. \tag{6.9}$$

$\hat{Q}$ is interpreted as the remaining nutrient from an infected cell after virus release that has not been taken up by the virus particles (Fuhrman *et al.*, 2011).

6.3.5 Full Model

The full model with nutrient dependent virus growth is summarized below:

$$\begin{aligned}
\frac{dS}{dt} &= \underbrace{\mu_m\left(1-\frac{q}{Q}\right)S}_{\text{Growth}} - \underbrace{\beta SV}_{\text{Loss from Infection}} - \underbrace{mS}_{\text{cell degeneration}} \\
\frac{dI}{dt} &= \underbrace{\beta SV}_{\text{growth}} - \underbrace{\delta I}_{\text{viral shedding}} \\
\frac{dV}{dt} &= \underbrace{b\left(Q\right)\delta I}_{\text{New virions}} - \underbrace{dV}_{\text{Cleared Virions}} - \underbrace{\beta SV}_{\text{absorption}} \\
\frac{dQ}{dt} &= \underbrace{\alpha N_f}_{\text{Uptake}} - \underbrace{\mu_m\left(Q-q\right)}_{\text{loss from cell growth}} \\
\frac{dN_f}{dt} &= -\alpha N_f(S+I) + mSQ + d\theta V + \delta I\hat{Q}
\end{aligned} \tag{6.10}$$

where $b\left(Q\right) = b_0 \max\{\left(1-\frac{q_v}{Q}\right), 0\}$. If we assume $b(Q) = b_0$ we obtain the model by Fuhrman et al., who have studied this system fully (Fuhrman *et al.*, 2011).

The assumption that the nutrient is indeed constant allows this model to be reduced to 4 equations. Indeed, to see this conservation law note that total nutrient is $N_t = N_f(t) + S(t)Q(t) + I(t)(Q(t)+\theta) + \theta V(t)$ and therefore,

$$\begin{aligned}
N_t' &= N_f' + S'Q + SQ' + I'Q + IQ' + \theta I' + \theta V' \\
&= -\alpha N_f\left(S+I\right) + mSQ + d\theta V + \delta I\hat{Q} \\
&+ \mu\left(1-\frac{q}{Q}\right)SQ - \beta SVQ - mSQ \\
&+ \alpha N_f S - \mu\left(Q-q\right)S \\
&+ \beta SVQ - \delta IQ \\
&+ \alpha N_f I - \mu\left(Q-q\right)I \\
&+ \theta\beta SV - \delta\theta I \\
&+ b(Q)\delta\theta I - d\theta V - \beta SVQ.
\end{aligned}$$

Canceling like terms, this reduces to

$$\begin{aligned} N_t' &= \delta I\hat{Q} - \delta IQ - \mu\left(Q-q\right)I - \delta\theta I + b(Q)\delta\theta I \\ &= \delta I\hat{Q} - \delta I\left(Q + \frac{\mu(Q-q)}{\delta} + \theta - b(Q)\theta\right) \\ &= \delta I\hat{Q} - \delta I\hat{Q} \\ &= 0 \end{aligned}$$

and allows us to reduce the model to a system of 4 equations.

$$\begin{aligned} \frac{dS}{dt} &= \underbrace{\mu_m\left(1-\frac{q}{Q}\right)S}_{\text{Growth}} - \underbrace{\beta SV}_{\text{Loss from Infection}} - \underbrace{mS}_{\text{cell degeneration}} \\ \frac{dI}{dt} &= \underbrace{\beta SV}_{\text{growth}} - \underbrace{\delta I}_{\text{viral shedding}} \\ \frac{dV}{dt} &= \underbrace{b\left(Q\right)\delta I}_{\text{New virions}} - \underbrace{dV}_{\text{Cleared Virions}} - \underbrace{\beta SV}_{\text{absorption}} \\ \frac{dQ}{dt} &= \underbrace{\alpha(N_t - SQ - I(Q+\theta) - \theta V)}_{\text{Uptake}} - \underbrace{\mu_m\left(Q-q\right)}_{\text{loss from cell growth}} \end{aligned} \tag{6.11}$$

We would like to note that to keep our modeling approach simple, we have decided not to model explicitly the virus population within the infected cell population. Furthermore, due to the lack of data and literature on this specific virus, we assume that in each infected cell there are b virus particles and therefore the total virus population is $bI + V$.

6.4 Numerical Work

We fit our mathematical model to the virion data from the four different experiments carried out by Kendig et al.

6.4.1 Susceptible Cells Population Estimation

We use root mass and shoot mass from healthy plants that were grown under CTRL, +N, +P and +NP and convert this mass into numbers of healthy cells under the assumption that there are roughly 10^9 plant cells for 1 gram of plant tissue (Wayne, 2009).

6.4.2 Parameter Estimations

We decided to derive parameter estimations by conducting two model fitting rounds. The first round of parameter estimations used healthy plant data and the disease free model of susceptible cell model 6.3. We used nonlinear least squares (implemented in MATLAB with built-in function FMINSEARCH) with susceptible cells data to estimate parameters μ, m, α, q and initial conditions $S(0)$ and $Q(0)$. These parameter estimations were then fixed when fitting the complete model (model 6.11) in the second round, except for $Q(0)$ which was refitted in the second round. This was done for all four nutrient experiments.

The second round was conducted by fitting $b_m I + V$ to virion data to generate parameter estimations for β, q_v, m, d b_m and $Q(0)$. δ was held constant at 1/13 day^{-1} as estimated in (Eweida *et al.*, 1988) for a similar virus as CYDV-RPV. As discussed above, during this second round of parameter estimations, we fixed the previous estimations of μ, m, α q, $S(0)$. In both rounds, best fit parameters were obtained by minimizing the following error function:

$$\text{err} = \sum_{i}^{N} \left((b_m I_i + V_i) - \bar{V}_i \right)^2. \tag{6.12}$$

Virion data is given by $\bar{V}_i$ and the analogous value given by our model is $bI_i + V_i$. We use $b_m I_i + V_i$ because $\bar{V}_i$ represents all virions (inside infected cells and free). N

Parameter	Fitted (CTRL)	Fitted (+N)	Fitted (+P)	Fitted (+NP)	Units
$S(0)$	1.12e+8	1.34e+8	1.24e+08	1.93e+8	cells
$Q(0)$	0.06	.08	0.015	0.15	fmol
μ_m	0.66	0.38	0.107	0.29	day^{-1}
m	0.55	0.28	6.609e-08	0.26	day^{-1}
α	11.52	19.25	36.55	21.37	fmol $\text{cell}^{-1}\text{day}^{-1}$
q	0.0025	0.0034	0.017	0.003	fmol
β	1.937e-10	3.901e-10	3.52e-10	1.09e-09	cells virion^{-1} day^{-1}
d	1.1336	0.97	0.55	1.94	day^{-1}
δ	1/13	1/13	1/13	1/13	day^{-1}
b_m	66.794	19.7	36.7	11.4	virions cell^{-1} day^{-1}
q_v	0.0006	0.0009	0.018	0.034	fmol
N_t	10^7	10^7	10^7	10^7	fmol
θ	4.106×10^{-4}	4.106×10^{-4}	4.106×10^{-4}	4.106×10^{-4}	fmol

Table 6.1: Parameter values used in Fig. 6.4 and Fig. 6.3

is the number of data points.

Model fittings are presented in Figure 6.4 and Figure 6.5 and fitted parameters can be found in Table 6.4.2. It is important to note that the model predicts nutrient limited virion growth in Figure 6.5, which is a stark difference than in the control and nitrogen addition experiments. As for now we shall delay our discussion of these results until the end of this chapter and now focus on our second question: Can the delay of virion production be a mechanism that explains the "bumps" seen in the data?

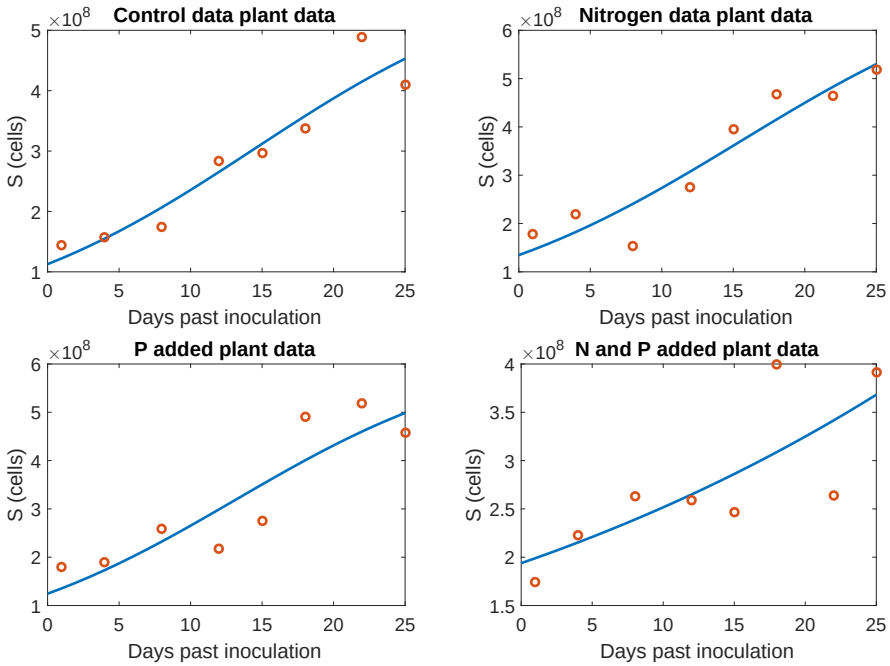

Figure 6.3: Model fits to healthy plant data. Top row (left to right): control data, nitrogen addition data, nitrogen and phosphorus addition and phosphorus addition.

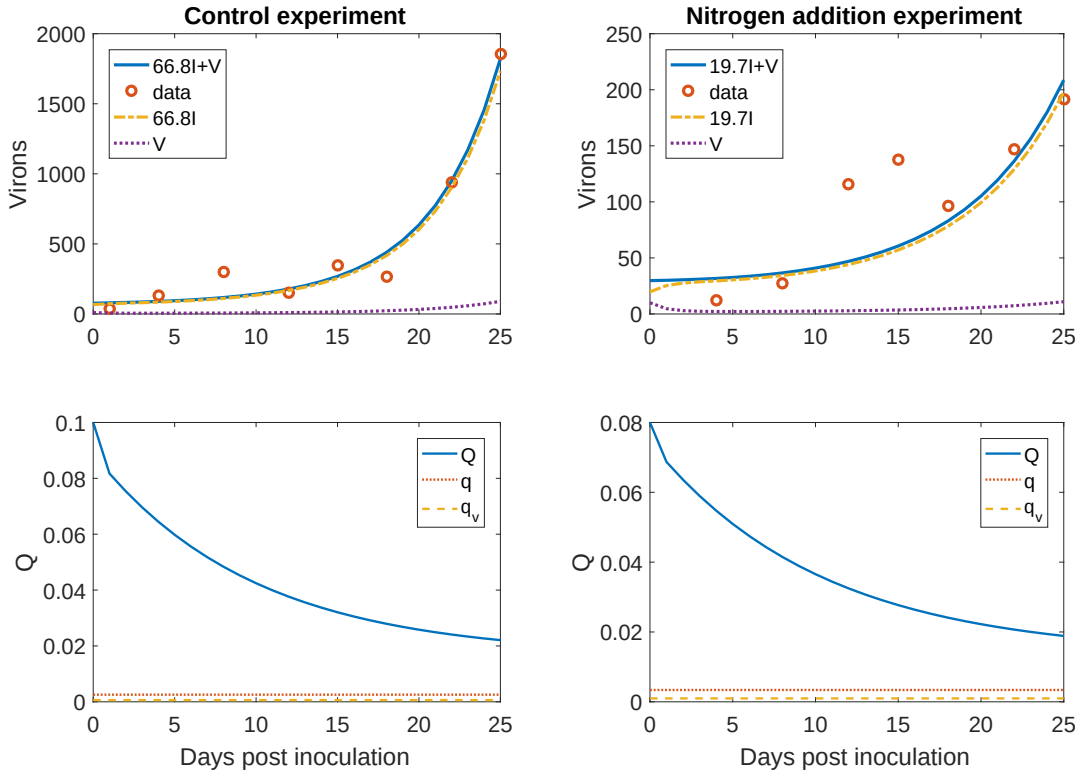

Figure 6.4: Model fits to virus data. (Left column: top) Fitted solutions of $b_m I + V$ against virion data (red dots) from the control experiment. (Left column: bottom) trajectory of the cell quota. Q does not go below the minimum cell quota for virus growth, q_v, and therefore V does not undergo a decline in virus production. (Right column: top) Fitted solutions of $b_m I + V$ against virion data (red dots) from the nitrogen addition experiment. (Right column: bottom) trajectory of the cell quota of the infected plant. In the nitrogen addition experiment, Q does not go below the minimum cell quota for virus growth, q_v, and therefore V does not undergo a decline in virus production

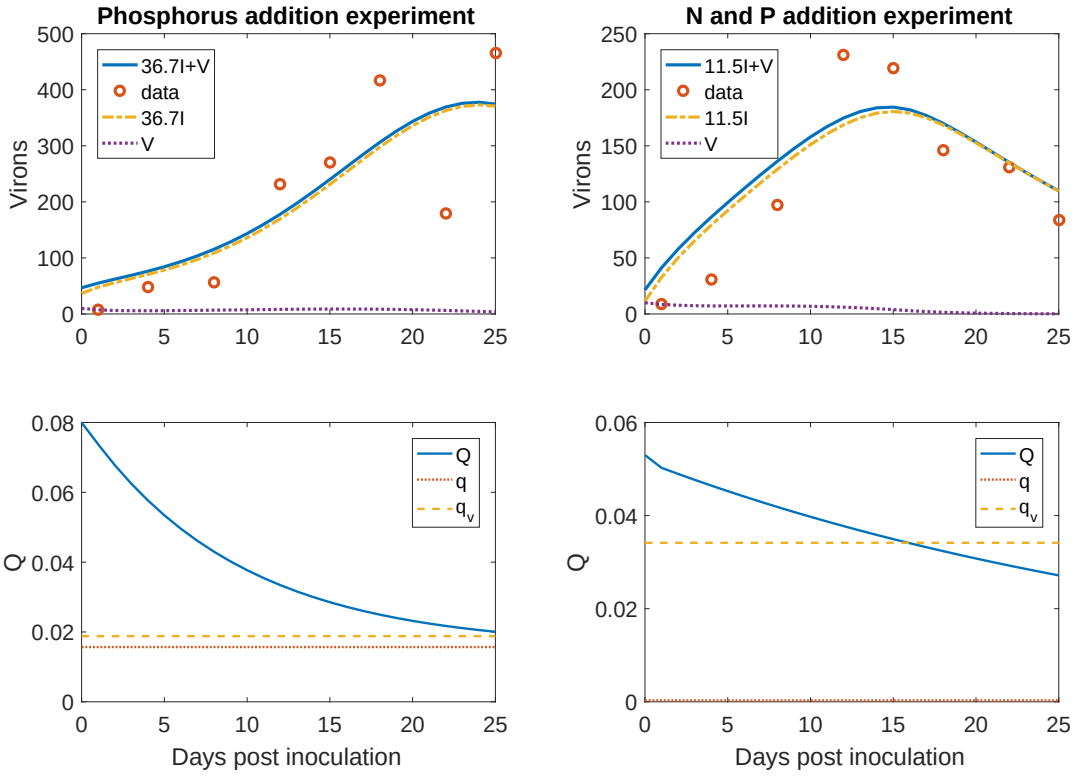

Figure 6.5: Model fits to virus data. (Left column: top) Fitted solutions of $b_m I + V$ against virion data (red dots) from the phosphorus experiment. (Left column: bottom) trajectory of the cell quota. Q approaches the minimum cell quota for virus growth, q_v, and therefore V does undergoes a decline in virus production. (Right column: top) Fitted solutions of $b_m I + V$ against virion data (red dots) from the nitrogen and phosphorus addition experiment. (Right column: bottom) trajectory of the cell quota of the infected plant. In the nitrogen and phosphorus addition experiment, Q does go below the minimum cell quota for virus growth, q_v, and therefore V stops virus production.

6.5 Modeling Virion Production Delay

To answer the second question, we develop a mathematical model that incorporates a delay in virus production. As we'll see, this leads naturally to a system of delay differential equations.

As in the previous model, let S and I be the number of susceptible and infected cells respectively and V be the number of virions (virus particles). Consider the following model derivation with a delayed virus production.

6.5.1 Healthy Phloem Cells

We assume that in the absence of infection, the susceptible cells obey logistic growth with carrying capacity K. Furthermore, we assume a maximum per-capita proliferation rate μ and a constant per-capita death rate m:

$$\frac{dS}{dt} = \mu\left(1 - \frac{S}{K}\right)S - mS. \tag{6.13}$$

6.5.2 Rate of Infection

We employ a standard incidence rate of infection: $\beta S(t)V(t)/N(t)$, where $N(t) = S(t) + I(t)$, where β can be interpreted as the maximum rate at which virions infect healthy phloem cells or the probability that a single virion infects any phloem cell in a healthy plant.

6.5.3 Infected Phloem Cells

We assume the virus is produced with a time delay τ. Where we interpret that each infected phloem cell on average produces b_m viruses in its lifetime with average life expectancy $\frac{1}{\delta}$ after infection of τ days.

At any time t, the density of the infected plant cells, $I(t)$, is obtained by integrating

$$\frac{\beta S(t-\theta)V(t-\theta)e^{-\delta\theta}}{S(t-\theta) + I(t-\theta)}$$

for $\theta \geq 0$. Here $\frac{\beta S(t-\theta)V(t-\theta)}{S(t-\theta)+I(t-\theta)}$, represents the rate of infection at previous times, and $e^{-\delta\theta}$ represents the probability of a cell surviving the infection from $t-\theta$ to t with natural mortality rate δ. Finally, we note that for any time, t, we do not integrate past $-\tau$, because these virions have already been released and left the I class. With these observations we obtain a governing equation for $I(t)$:

$$I(t) = \int_0^{\tau} \frac{\beta S(t-\theta)V(t-\theta)e^{-\delta\theta}}{S(\theta)+I(\theta)} d\theta.$$

With a change of variables $s = t - \theta$ we obtain:

$$I(t) = \int_{t-\tau}^{t} \frac{\beta S(s)V(s)e^{\delta(s-t)}}{S(s)+I(s)} ds$$

and differentiating with respect to t yields

$$\frac{dI(t)}{dt} = \frac{\beta S(t)V(t)}{S(t)+I(t)} - e^{-\delta\tau}\frac{\beta S(t-\tau)V(t-\tau)}{S(t-\tau)+I(t-\tau)} - \delta I.$$

We note that equation 6.5.3 can also be derived from the Mckendrick-von Foerster age-structured model (Gourley *et al.*, 2008).

6.5.4 Free Virions

Free virions are produced at rate b_m per infected phloem cell per unit of time and are destroyed or cleared by the (innate) immune system ate rate d. Free virions that are not destroyed or cleared are absorbed by neighboring phloem cells. With these observations we obtain the governing equation for V,

$$\frac{dV}{dt} = be^{-\delta\tau}\frac{\beta S(t-\tau)V(t-\tau)}{S(t-\tau)+I(t-\tau)} - dV - \frac{\beta SV}{S+I}.$$

6.5.5 Full Model

With the above considerations we obtain the following delayed virus production compartmental model,

$$\begin{aligned}
\frac{dS}{dt} &= \hat{r}\left(1 - \frac{S+I}{\hat{K}}\right)S - \frac{\beta SV}{S+I} \\
\frac{dI}{dt} &= \frac{\beta SV}{S+I} - \delta I - e^{-\delta\tau}\frac{\beta S(t-\tau)V(t-\tau)}{S(t-\tau)+I(t-\tau)} \\
\frac{dV}{dt} &= be^{-\delta\tau}\frac{\beta S(t-\tau)V(t-\tau)}{S(t-\tau)+I(t-\tau)} - dV - \frac{\beta SV}{S+I}.
\end{aligned} \tag{6.14}$$

For simplicity and parameter estimation purposes we rewrite the equation for S as

$$\frac{dS}{dt} = \hat{r}\left(1 - \frac{S+I}{\hat{K}}\right)S - \frac{\beta SV}{S+I} \tag{6.15}$$

where $\hat{r} = r - m$ and $\hat{K} = K\frac{\hat{r}}{r}$.

6.5.6 Initial Data

Initial data for the system takes the form of

$$\begin{gathered} S(s) = S_0(s) \geq 0, \quad I(s) = I_0(s) \geq 0, \quad V(s) = V_0(s) \geq 0 \\ \text{and } K \geq S(s) + I(s) > 0, I(0) > 0 \text{ for } s \in [-\tau, 0], \end{gathered} \tag{6.16}$$

where

$$I_0(0) = \int_{-\tau}^{0} \frac{\beta S_0(s) V_0(s) e^{\delta s}}{S_0(s) + I_0(s)} ds. \tag{6.17}$$

6.5.7 The Basic Reproduction Number

To derive $\tilde{R}_0$ we first observe that at the average number of virions produced per infected cell per timestep is $\beta b e^{-\delta\tau}$ and the average lifetime of a virion is $\frac{1}{d+\beta}$. Multiplying these together yield the basic reproduction number for the model:

$$\tilde{R}_0 := \frac{b\beta e^{-\delta\tau}}{d+\beta}.$$

6.6 Preliminary Analysis

6.6.1 Positivity of Solutions

When modeling populations, it is important that the model makes biological sense. In particular, populations should never become negative and also be bounded above by some finite number.

We show below that solutions of system (6.14) with nonnegative initial conditions remain nonnegative and bounded (and hence exists for all time) for all $t > 0$.

Proposition 6.6.1 *Each component of system (6.14), subject to initial conditions (6.16), remains nonnegative and bounded for all $t > 0$.*

Proof Observe that system (6.14) is locally Lipschitz at $t = 0$. Hence the solution of system (6.14), subject to initial conditions (6.16), exists and is unique on $[0, c)$ for some $c > 0$. Furthermore, if $S(0) = 0$, then $S(t) \equiv 0$ for all $t > 0$. Therefore we may assume that $S(0) > 0$. In a similar fashion, if $V(0) = 0$, then $V(t) \equiv 0$ for all $t > 0$. Therefore we may also assume that $V(0) > 0$.

To see that $S + I$ is bounded, observe that,

$$\begin{aligned}\frac{d(S+I)}{dt} &= \hat{r}\left(1 - \frac{S+I}{\hat{K}}\right)S - \frac{\beta S(t-\tau)V(t-\tau)}{S(t-\tau)+I(t-\tau)} \\ &\leq \frac{-\hat{r}}{\hat{K}}\left(S + I - \hat{K}\right)S.\end{aligned}$$

This implies that

$$\frac{d(S+I-\hat{K})}{dt} \leq \frac{-\hat{r}}{\hat{K}}\left(S + I - \hat{K}\right)S$$

and integrating yields

$$S(t) + I(t) \leq \hat{K} + \left[S(0) + I(0) - \hat{K}\right]\exp\{\frac{-\hat{r}}{\hat{K}}\int_0^t S(s)ds\}.$$

Therefore, if $S(0) + I(0) < \hat{K}$, then $S(t) + I(t) < \hat{K}$ for all $t > 0$.

Case 1: By way of contradiction we assume that there exists $t_1 \in (0, c)$ such that $S(t_1) = 0$ and $S(t) > 0$, $I(t) > 0$ and $V(t) > 0$ for $t \in (0, t_1)$. Observe that for $t \in [0, t_1]$,

$$\begin{aligned}\frac{dS}{dt} &= \hat{r}\left(1-\frac{S+I}{\hat{K}}\right)S - \frac{\beta SV}{S+I} \\ &\geq -\left(\frac{\beta V}{S+I}\right)S \\ &\geq -\max_{t\in[0,t_1]}\left\{\frac{\beta V}{S+I}\right\}S \\ &= -\alpha S\end{aligned}$$

Where $\alpha = \max_{t\in[0,t_1]}\left\{\frac{\beta V}{S+I}\right\}$ and integrating yields,

$$S(t) \geq S(0)e^{-\alpha t} > 0,$$

which is a contradiction when $t = t_1$. Therefore $S(t) > 0$ for $t \in [0, c)$.

Case 2: Assume that there exists a $t_1 \in (0, c)$ such that $V(t_1) = 0$, $I(t) > 0$ and $S(t) > 0$, $I(t) > 0$ for $t \in (0, t_1)$. Therefore, for $t \in [0, t_1]$,

$$\begin{aligned}\frac{dV}{dt} &= be^{-\delta\tau}\frac{\beta S(t-\tau)V(t-\tau)}{S(t-\tau)+I(t-\tau)} - dV - \frac{\beta SV}{S+I} \\ &\geq -\left(d+\frac{\beta S}{S+I}\right)V \\ &\geq -\max_{t\in[0,t_1]}\left\{d+\frac{\beta S}{S+I}\right\}V \\ &= -\alpha_1 V\end{aligned}$$

where $\alpha_1 = \max_{t\in[0,t_1]}\left\{d+\frac{\beta S}{S+I}\right\}$. This implies,

$$V(t) \geq V(0)e^{-\alpha_1 t} > 0, \tag{6.18}$$

which yields a contradiction for $t = t_1$. Therefore $V(t) > 0$ for $t \in [0, c)$.

Case 3: Assume that there exists a $t_1 \in (0, c)$ such that $I(t_1) = 0$ and that $S(t) > 0$, $I(t) > 0$ $V(t) > 0$ for $t \in (0, t_1)$. Since $I(t) = \int_{t-\tau}^{t} \frac{\beta S(s)V(s)e^{\delta(s-t)}}{S(s)+I(s)}ds$, we have

$$I(t_1) = \int_{t_1-\tau}^{t_1} \frac{\beta S(s)V(s)e^{\delta(s-t_1)}}{S(s)+I(s)}ds > 0 \tag{6.19}$$

a contradiction. Therefore $I(t) > 0$ for $t \in [0, c)$.

Finally we prove a rather large bound for the entire system. Consider $W = S + I + \frac{1}{bV}$. Then differentiating yields,

$$\begin{aligned}
\frac{dW}{dt} &= \hat{r}\left(1 - \frac{S+I}{\hat{K}}\right)S - \delta I - \frac{d}{b}V - \frac{\beta SV}{b(S+I)} \\
&\leq \hat{r}\left(1 - \frac{S+I}{\hat{K}}\right)S - \delta I - \frac{d}{b}V \\
&= \hat{r}S - \frac{\hat{r}S^2}{\hat{K}} - \frac{\hat{r}SI}{\hat{K}} - \delta I - \frac{d}{b}V \\
&\leq \hat{r}\hat{K} - \delta I - \frac{d}{b}V \\
&= \hat{r}\hat{K} - \delta I - \frac{d}{b}V + (\hat{r}S - \hat{r}S) \\
&= 2\hat{r}\hat{K} - \delta I - \frac{d}{b}V - \hat{r}S \\
&\leq 2\hat{r}\hat{K} - \min\{\delta, d, \hat{r}\}\left(I + S + \frac{1}{b}V\right) \\
&= 2\hat{r}\hat{K} - \alpha_2 W.
\end{aligned}$$

Integrating finally yields us with

$$W(t) \leq W(0)e^{-\alpha t} + \frac{2\hat{r}\hat{K}}{\alpha_3}\left(1 - e^{-\alpha_2 t}\right).$$

Therefore

$$\limsup_{t\to\infty} W(t) \leq \frac{2\hat{r}\hat{K}}{\alpha_2},$$

where $\alpha_2 = \min\{\delta, d, \hat{r}\}$. Thus we may conclude that all state variables are bounded.

The above contradictions together show that components of the solution of system (6.14), subject to initial data (6.16), are nonnegative for all $t \in [0, c)$. This together with the uniform boundedness of $W = S + I + \frac{1}{b}V$ imply that $c = \infty$. This completes the proof of the proposition.

6.6.2 Steady States

There exists two biologically relevant steady states: $\tilde{E}_0 = \left(\hat{K}, 0, 0\right)$ and $\tilde{E}^* = \left(\tilde{S}^*, \tilde{I}^*, \tilde{V}^*\right)$ where,

$$\begin{aligned}
\tilde{S}^* &= \frac{e^{\delta\tau}\hat{K}\left(d\hat{r} + b\beta\delta\right)}{\beta\hat{r}(e^{\delta\tau}-1)(b-e^{\delta\tau})}\left(R^*-1\right) \\
\tilde{I}^* &= \frac{e^{\delta\tau}\hat{K}\left(d+\beta\right)\left(d\hat{r}+b\beta\delta\right)}{\beta d\hat{r}(e^{\delta\tau}-1)(b-e^{\delta\tau})}\left(R^*-1\right)\left(\tilde{R}_0-1\right) \\
\tilde{V}^* &= \frac{e^{\delta\tau}\hat{K}\delta\left(d+\beta\right)\left(d\hat{r}+b\beta\delta\right)}{\beta d^2\hat{r}(e^{\delta\tau}-1)^2}\left(R^*-1\right)\left(\tilde{R}_0-1\right)
\end{aligned} \tag{6.20}$$

and $R^* = \frac{e^{\delta\tau}(\beta\delta + d(\hat{r}+\delta))}{d\hat{r}+b\beta\delta}$. $\tilde{E}_0$ represents a healthy plant and is called the disease free steady state. $\tilde{E}^*$ is the called the endemic equilibrium and represents chronic infection of the plant by the virus.

Turning our attention to the endemic equilibrium, it's easy to see that it exists exactly when the following conditions hold:

$$R^* = \frac{e^{\delta\tau}\left(\beta\delta + d\left(\hat{r}+\delta\right)\right)}{d\hat{r}+b\beta\delta} > 1 \qquad b > e^{\delta\tau} \text{ and} \qquad \tilde{R}_0 > 1. \tag{6.21}$$

We focus on the stability of the biologically relevant equilibria. With this in mind we compute the characteristic polynomial,

$$h(\lambda) = \det(\lambda\mathbb{I} - P - e^{-\delta\tau}e^{-\lambda\tau}Q). \tag{6.22}$$

The matrices P and Q are given by,

$$P = \begin{pmatrix}
\frac{\hat{r}(-I+\hat{K}-2S)}{\hat{K}} - \frac{IV\beta}{(I+S)^2} & -\frac{\hat{r}S}{\hat{K}} + \frac{SV\beta}{(I+S)^2} & -\frac{S\beta}{I+S} \\
\frac{IV\beta}{(I+S)^2} & -\frac{SV\beta}{(I+S)^2} - \delta & \frac{S\beta}{I+S} \\
-\frac{IV\beta}{(I+S)^2} & \frac{SV\beta}{(I+S)^2} & -d - \frac{S\beta}{I+S}
\end{pmatrix}$$

and

$$Q = \begin{pmatrix} 0 & 0 & 0 \\ -\frac{IV\beta}{(I+S)^2} & \frac{SV\beta}{(I+S)^2} & -\frac{S\beta}{I+S} \\ \frac{bIV\beta}{(I+S)^2} & -\frac{bSV\beta}{(I+S)^2} & \frac{bS\beta}{I+S} \end{pmatrix}.$$

At $\tilde{E}_0$ the characteristic equation is

$$(\hat{r}+\lambda)(\delta+\lambda)(\beta b e^{(-\lambda\tau-\delta\tau)} - (d+\beta) - \lambda) = 0$$

where the roots are $\lambda_1 = -\hat{r}$, $\lambda_2 = -\delta$ and λ_3 satisfies $\lambda_3 = b\beta e^{-(\lambda_3+\delta)\tau} - (d+\beta)$. Setting $\lambda_3 = 0$ and rearranging this we obtain,

$$1 = \frac{b\beta e^{-\delta\tau}}{d+\beta} =: \tilde{R}_0,$$

the basic reproduction number for the standard incidence model. This further confirms our definition of the basic reproduction number since it determines the stability of the disease free steady state.

Proposition 6.6.2 *$\tilde{E}_0$ is asymptotically stable if $\tilde{R}_0 < 1$.*

Proof For $\tilde{E}_0$ to exists, $\tilde{S}^* > 0$ and therefore $\hat{K} > 0$. By the above discussion the roots of the characteristic polynomial are given by $\lambda_1 = -\hat{r}$, $\lambda_2 = -\delta$ and $\lambda = b\beta e^{-(\lambda+\delta)\tau} - (d+\beta)$. Hence, the first two roots have negative real part. Thus the stability of the disease free equilibrium depends on the roots of,

$$g(\lambda) = \lambda + (d+\beta) - b\beta e^{-(\lambda+\delta)\tau}.$$

We prove that $g(\lambda)$ cannot have a root with nonnegative real part when $\tilde{R}_0 < 1$. By way of contradiction, assume $\tilde{R}_0 < 1$, but there does exist a root with nonnegative real part, $\lambda = x + iy$ where $x \geq 0$ and $x, y \in \mathbb{R}$. Setting $g(\lambda) = 0$ yields $\lambda = b\beta e^{-\delta} e^{-\lambda\tau} - (d+\beta)$ and substituting $\lambda = x + iy$ we obtain,

$$\begin{aligned} x + iy &= b\beta e^{-\delta\tau} e^{-(x+iy)\tau} - (d+\beta) \\ &= b\beta e^{-\delta\tau} e^{-x\tau} e^{-iy\tau} - (d+\beta) \\ &= b\beta e^{-\delta\tau} e^{-x\tau} \left(\cos(y\tau) - i\sin(y\tau)\right) - (d+\beta) \\ &= b\beta e^{-\delta\tau} e^{-x\tau} \cos(y\tau) - (d+\beta) - ib\beta e^{-\delta\tau} \sin(y\tau). \end{aligned} \tag{6.23}$$

Equating real parts yields,

$$\begin{aligned} x &= b\beta e^{-\delta\tau} e^{-x\tau} \cos(y\tau) - (d+\beta) \\ &= (d+\beta)\left(\frac{b\beta e^{-\delta\tau}}{d+\beta} e^{-x\tau}\cos(y\tau) - 1\right) \\ &= (d+\beta)\left(\tilde{R}_0 e^{-x\tau}\cos(y\tau) - 1\right) < 0. \end{aligned} \tag{6.24}$$

A contradiction. Thus, any root of $g(\lambda)$ must have negative real part. □

Proposition 6.6.3 *$\tilde{E}_0$ is unstable if $\tilde{R}_0 > 1$.*

Proof As in the previous proof, for $\tilde{E}_0$ to exists, $\tilde{S}^* > 0$ and therefore $\hat{K} > 0$. Futhermore, stability depends on,

$$g(\lambda) = \lambda + (d+\beta) - b\beta e^{-(\lambda+\delta)\tau}.$$

Assume $\tilde{R}_0 > 1$, then $g(0) = d + \beta - b\beta e^{\delta\tau} < 0$. Furthermore, for $\lambda \geq 0$ we have,

$$g'(\lambda) = 1 + \tau b\beta e^{-\delta\tau} e^{-\lambda\tau} > 0.$$

Finally, $\lim_{\lambda\to\infty} g(\lambda) = \infty$. Since $g(\lambda)$ is a continuous function that is negative at $\lambda = 0$ and increases to $+\infty$ as $\lambda \to +\infty$, it must cross the cross the λ-axis. This proves the existence of a positive real root. Therefore, the disease free steady state is unstable. □

It is notable to mention that there are other ways to prove the stability of the disease free steady state. Indeed, exercise 4.9 from (Smith, 2010) or theorem 3.1 from (Beretta and Kuang, 2002) along with theorem 1.4 from (Kuang, 1993) are other ways to prove the theorems. We decided to include the proofs that we did, because of their simple and intuitive arguments.

6.7 Numerical Simulations

As we did with the nutrient model, we estimated $\hat{r}$, $S(0)$ and $\hat{K}$ using Equation 6.13 ($\beta = 0$) by fitting it to healthy plant cell data. We then used these parameters as the initial guesses when fitting the full model to the virion data. Our decision not to keep $\hat{r}$ and $\hat{K}$ fixed, but used as the intial guesses is to allow for stochastic changes that can occur to the plant parameters when the virus is introduced. We did not do this in the nutrient model since the limiting nutrient ($Q(t)$)) should change, not the model parameters. For example, when the virus is introduced into the nutrient model, model 6.3, the specific growth rate is what changes due to the now lower $Q(t)$ (from virion uptake), not the maximum specific growth rate, μ_m.

Nonlinear least squares was use again with an analogous error function as Equation 6.12.

We present model fits to virion data in Figure 6.6 and fitted parameters in Table 6.7.

Parameter	Fitted (CTRL)	Fitted (+N)	Fitted (+P)	Units
$S(0)$	1.12e+8	1.34e+8	1.24e+08	cells
$\hat{r}$	0.13	0.11	0.19	day^{-1}
$\hat{K}$	8.05e+08	8.21e+08	3.40e+08	cells
β	0.14	0.71	0.33	cells virion^{-1} day^{-1}
d	0.36037	0.24	0.19	day^{-1}
δ	1/13	1/13	1/13	day^{-1}
b	40	5.84	12.89	virions cell^{-1} day^{-1}
τ	9.4	9.7	10.95	days

Table 6.2: Parameter values used in Fig. 6.6

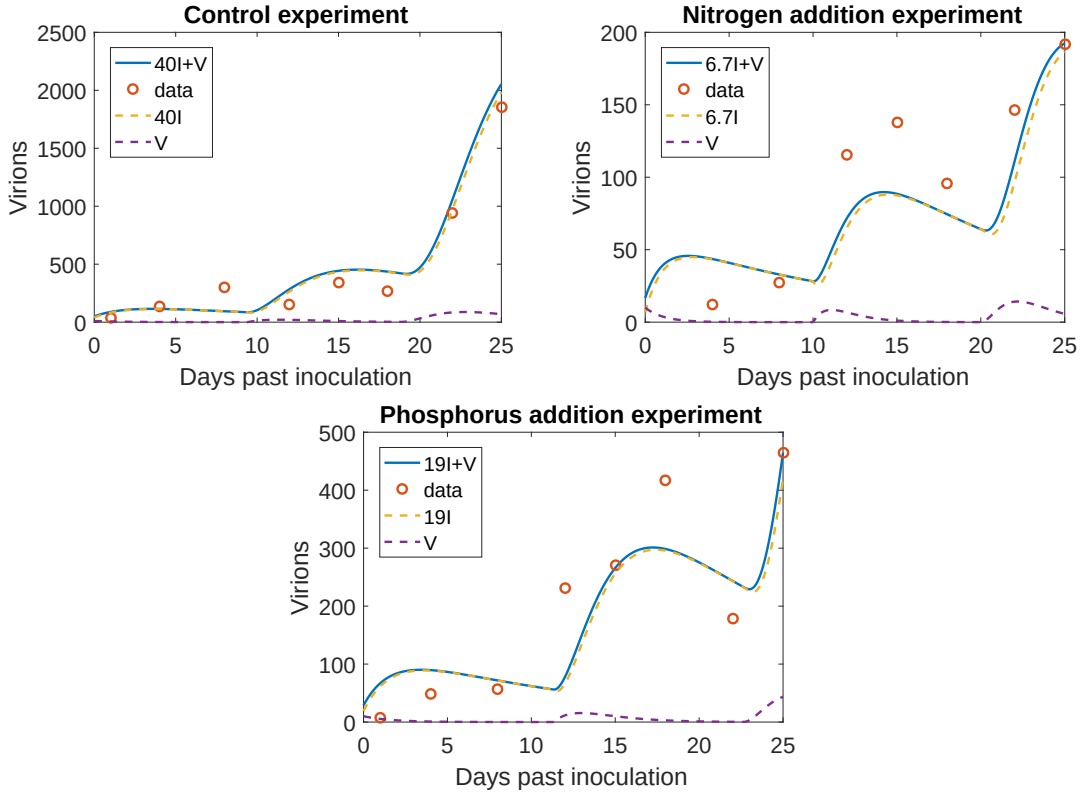

Figure 6.6: Model fits to control, nitrogen addition and phosphorus addition experiments.

6.8 Conclusion and Discussion

In this last chapter we formulated two mathematical models to investigate the virion dynamics under different nutrient regimens. Both of the models were moti-

vated and guided by the nature of the data. Our first model was motivated by the virus dynamics and the hypothesis that nutrient conditions inside the infected cells controlled the virion production. With this in mind, we designed a mathematical model that was based off (Fuhrman *et al.*, 2011) that related virus production to nutrient concentration in its host cells. By assuming that virus production can be approximated by the Droop equation (Equation 6.1), we showed that the behavior of the virion data of Kendig et al., can be explained by nutrient limited growth of the virus population. In particular, the model predicts that virion production is not limited in the control and nitrogen addition experiments. In contrast, the virion dynamics during the +P and +NP regimens can be explained by nutrient limited growth.

Our second modeling approach was designed to investigate the plausibility that the different nutrient regimens delayed the virion production within infected cells. Model fits for the control, nitrogen addition and phosphorus addition experiments suggest that delayed virus production is a plausible mechanism for explaining the dynamics seen in the virion data. In particular, the best fit solutions from our models uncover patterns that otherwise would be hard to detect when viewing the data by itself. Moreover, this modeling approach results in at least two new questions: what mechanisms are capable of delaying virus production and what roles do phosphorus and nitrogen play in these mechanisms?

The second modeling approach suggests that virion production delay is another reasonable mechanism for explaining 3 out of the 4 data sets (CTRL, +N and +P). Since virion estimations from the experiment were taken from leaf tissue samples, virus particles that move to the roots to replicate and then return to the leaves (Eweida *et al.*, 1988) could explain the bumps in the data. Another mechanism for describing the data could be that the nutrient regimens are changing the

Experiment	Delay model		Nutrient model	
	MSE	MAPE	MSE	MAPE
CTRL	19524	0.52	9275	.42859
+N	1027	0.38	1677	.55992
+P	5894	0.52	8372	.4108

Table 6.3: Error comparison table: Mean absolute and mean square errors from model 6.3 and model 6.14.

host's innate immune system (Jones and Dangl, 2006). RNA silencing is one such defense a plant has against infectious particles and perhaps this mechanism is eventually overcome by viral silencing suppressors (Csorba *et al.*, 2015), allowing the virus population to increase as seen in the CTRL data from figure 6.6. Ultimately, since some of the data points were collected after a nutrient addition, it is plausible that the patterns in the data could be the results from the increased supply of nutrient.

In all of our simulations, we find that model fits highly depend on the virus particles within the infected cells. Indeed, Figure 6.4, Figure 6.5 and Figure 6.6 show that free virus particles are almost nonexistent, but are driving the infection in the host. This behavior is expected and can be explained by the tightly packed structure of the phloem cells and the fact that virus particles pass from cell to cell via the plasmodesmata. Another future direction for both models would be to incorporate the intracellular virus population.

Table 6.3 shows mean square error and mean absolute percentage error (MAPE) for both models for experiments: CTRL, +N and +P. The delayed nutrient model outperformed the nutrient model in experiment +N and has a smaller MSE value in experiment +P. In contrast, the nutrient model performed better than the delayed model in the CTRL experiment and had a smaller MAPE in +P.

Although, we did parameterize the model to the best of our abilities, there is still much that can be done. In particular, both of our modeling approaches do not explicitly model the virus particles within each infected cell, and instead we make the

assumption that there are b_m (b in model 6.14) viruses per infected cell. This was done, due to the lack of data and the need for first iteration models. Ultimately, this does not hinder our goals, which are to show that mathematical models can be used to explain the behavior seen in the virus data.

As in chapter 1, parameter identifiability and sensitivity analysis could guide the direction for future model iterations and experiment design. Sensitivity analysis could be implemented by various methods such as the Morris and Sobol' methods, Latin hypercube sampling-partial rank correlation coefficient and the sensitivity heat map method (Wu *et al.*, 2013), while parameter identifiability could be implemented by the Differential Algebra Identifiability of Systems (DASIY) and other such methods discussed in (Saccomani *et al.*, 2003; Karlsson *et al.*, 2012; Raue *et al.*, 2009, 2014).

"Center for disease control", URL http://www.cdc.gov/vhf/ebola/transmission/index.html (2014).

Agnandji, S. T., A. Huttner, M. E. Zinser, P. Njuguna, C. Dahlke, J. F. Fernandes, S. Yerly, J.-A. Dayer, V. Kraehling, R. Kasonta *et al.*, "Phase 1 trials of rVSV Ebola vaccine in Africa and Europe–preliminary report", New England Journal of Medicine (2015).

Agusto, F. B., M. I. Teboh-Ewungkem and A. B. Gumel, "Mathematical assessment of the effect of traditional beliefs and customs on the transmission dynamics of the 2014 Ebola outbreaks", BMC medicine **13**, 1, 96 (2015).

Ali, M., S. Hameed and M. Tahir, "Luteovirus: insights into pathogenicity", Archives of virology **159**, 11, 2853–2860 (2014).

Althaus, C., N. Low, E. Musa, F. Shuaib and S. Gsteiger, "Ebola virus disease outbreak in nigeria: Transmission dynamics and rapid control", Epidemics **11**, 80 – 84, URL http://www.sciencedirect.com/science/article/pii/S1755436515000341 (2015a).

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

Althaus, C. L., N. Low, E. O. Musa, F. Shuaib and S. Gsteiger, "Ebola virus disease outbreak in Nigeria: transmission dynamics and rapid control", Tech. rep., PeerJ PrePrints (2015b).

Anderson, R. M. and R. M. May, *Infectious diseases of humans*, vol. 1 (Oxford university press Oxford, 1991).

Arnold, D., *Colonizing the body: State medicine and epidemic disease in nineteenth-century India* (Univ of California Press, 1993).

Atangana, A. and E. F. D. Goufo, "On the mathematical analysis of ebola hemorrhagic fever: deathly infection disease in West African countries", BioMed research international **2014** (2014).

Bacaër, N., "Daniel bernoulli, d'alembert and the inoculation of smallpox (1760)", in "A short history of mathematical population dynamics", pp. 21–30 (Springer, 2011).

Bacaër, N., "The model of kermack and mckendrick for the plague epidemic in bombay and the type reproduction number with seasonality", Journal of mathematical biology **64**, 3, 403–422 (2012).

Benedictow, O. J., "The black death: The greatest catastrophe ever", URL http://www.historytoday.com/ole-j-benedictow/black-death-greatest-catastrophe-ever (2005).

Beretta, E. and Y. Kuang, "Geometric stability switch criteria in delay differential systems with delay dependent parameters", SIAM Journal on Mathematical Analysis **33**, 5, 1144–1165 (2002).

Bernoulli, D. and S. Blower, "An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it", Reviews in medical virology **14**, 5, 275–288 (2004).

Bickel, P. J. and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics, volume I*, vol. 117 (CRC Press, 2015).

Browne, C., H. Gulbudak and G. Webb, "Modeling Contact Tracing in Outbreaks with Application to Ebola", arXiv preprint arXiv:1505.03821 (2015).

Camacho, A., A. Kucharski, Y. Aki-Sawyerr, M. A. White, S. Flasche, M. Baguelin, T. Pollington, J. R. Carney, R. Glover, E. Smout *et al.*, "Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: a Real-time Modelling Study", PLoS currents **7** (2015).

Carrigan, L., H. Ohm and J. Foster, "Barley yellow dwarf virus translocation in wheat and oats", Crop science **23**, 4, 611–612 (1983).

Cenciarelli, O., S. Pietropaoli, A. Malizia, M. Carestia, F. DAmico, A. Sassolini, D. Di Giovanni, S. Rea, V. Gabbarini, A. Tamburrini *et al.*, "Ebola virus disease 2013-2014 outbreak in west africa: an analysis of the epidemic spread and response", International journal of microbiology **2015** (2015).

Center for Disease Control, "Outbreaks Chronology: Ebola Virus Disease", Website, URL `http://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html` (2015).

Chowell, G., N. W. Hengartner, C. Castillo-Chavez, P. W. Fenimore and J. Hyman, "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 (2004).

Chowell, G., D. Hincapie-Palacio, J. Ospina, B. Pell, A. Tariq, S. Dahal, S. Moghadas, A. Smirnova, L. Simonsen and C. Viboud, "Using phenomenological models to characterize transmissibility and forecast patterns and final burden of zika epidemics", PLOS Currents Outbreaks (2016a).

Chowell, G., J. M. Hyman, L. M. Bettencourt and C. Castillo-Chavez, *Mathematical and statistical estimation approaches in epidemiology* (Springer, 2009).

Chowell, G. and H. Nishiura, "Transmission dynamics and control of Ebola virus disease (EVD): a review", BMC medicine **12**, 1, 196 (2014).

Chowell, G. and H. Nishiura, "Characterizing the transmission dynamics and control of ebola virus disease", PLoS Biol **13**, 1, e1002057 (2015).

Chowell, G., H. Nishiura and L. M. Bettencourt, "Comparative estimation of the reproduction number for pandemic influenza from daily case notification data", Journal of the Royal Society Interface **4**, 12, 155–166 (2007).

Chowell, G., L. Simonsen, C. Viboud and Y. Kuang, "Is West Africa approaching a catastrophic phase or is the 2014 Ebola epidemic slowing down? Different models yield different answers for Liberia", PLoS currents **6** (2014a).

Chowell, G., C. Viboud, J. M. Hyman and L. Simonsen, "The western africa ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates", arXiv preprint arXiv:1411.7364 (2014b).

Chowell, G., C. Viboud, L. Simonsen and S. Moghadas, "Characterizing the reproduction number of epidemics with early sub-exponential growth dynamics", arXiv preprint arXiv:1603.01216 (2016b).

Clasen, J. L. and J. J. Elser, "The effect of host Chlorella NC64A carbon:phosphorus ratio on the production of Paramecium bursaria Chlorella Virus-1", Freshw. Biol. **52**, 1, 112–122 (2007).

Commission, P. R. *et al.*, "The epidemiological observations made by the commission in bombay city", Journal of Hygiene **7**, 724–98 (1907).

Csorba, T., L. Kontra and J. Burgyn, "viral silencing suppressors: Tools forged to fine-tune host-pathogen coexistence", Virology **479480**, 85 – 103, URL `http://www.sciencedirect.com/science/article/pii/S0042682215000744`, 60th Anniversary Issue (2015).

D'Arcy, C. J., P. A. Burnett *et al.*, *Barley yellow dwarf: 40 years of progress.* (American Phytopathological Society (APS Press), 1995).

Davison, A. C. and D. V. Hinkley, *Bootstrap methods and their application*, vol. 1 (Cambridge university press, 1997).

Drake, J. M., R. B. Kaul, L. W. Alexander, S. M. ORegan, A. M. Kramer, J. T. Pulliam, M. J. Ferrari and A. W. Park, "Ebola cases and health system demand in Liberia", PLoS Biology **13**, 1 (2015).

Droop, M., "Nutrient limitation in osmotrophic protista", American Zoologist **13**, 1, 209–214 (1973a).

Droop, M., "Some thoughts on nutrient limitation in algae1", Journal of Phycology **9**, 3, 264–272 (1973b).

Droop, M., "The nutrient status of algal cells in continuous culture", Journal of the Marine Biological Association of the United Kingdom **54**, 04, 825–855 (1974).

Duncan-Jones, R. P., "The impact of the antonine plague", Journal of Roman Archaeology **9**, 108–136 (1996).

Efron, B. and R. J. Tibshirani, *An introduction to the bootstrap* (CRC press, 1994).

Eisenberg, M. C., J. N. 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", arXiv preprint arXiv:1501.05555 (2015).

Erion, G. G. and W. E. Riedell, "Barley yellow dwarf virus effects on cereal plant growth and transpiration", Crop Science **52**, 6, 2794–2799 (2012).

Everett, R., *Applications of the Droop Cell Quota Model to Data Based Cancer Growth and Treatment Models*, Ph.D. thesis, URL `http://hdl.handle.net/2286/R.A.150557` (2015).

Everett, R., J. Nagy and Y. Kuang, "Dynamics of a data based ovarian cancer growth and treatment model with time delay", Journal of Dynamics and Differential Equations pp. 1–22 (2015).

Everett, R., A. Packer and Y. Kuang, "Can mathematical models predict the outcomes of prostate cancer patients undergoing intermittent androgen deprivation therapy?", Biophysical Reviews and Letters **9**, 02, 173–191 (2014).

Eweida, M., P. Oxelfelt and K. Tomenius, "Concentration of virus and ultrastructural changes in oats at various stages of barley yellow dwarf virus infection", Annals of applied biology **112**, 2, 313–321 (1988).

Faye, O., P.-Y. Boëlle, E. Heleze, O. Faye, C. Loucoubar, N. Magassouba, B. Soropogui, S. Keita, T. Gakou, L. Koivogui *et al.*, "Chains of transmission and control of ebola virus disease in conakry, guinea, in 2014: an observational study", The Lancet Infectious Diseases **15**, 3, 320–326 (2015).

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

Fisman, D. N., T. S. Hauck, A. R. Tuite and A. L. Greer, "An idea for short term outbreak projection: nearcasting using the basic reproduction number", PloS one **8**, 12, e83622 (2013).

Fraser, C., "Estimating individual and household reproduction numbers in an emerging epidemic", PLoS One **2**, 8, e758 (2007).

Frieden, T. R., I. Damon, B. P. Bell, T. Kenyon and S. Nichol, "Ebola 2014 — New Challenges, New Global Response and Responsibility", New England Journal of Medicine **371**, 13, 1177–1180 (2014).

Frieden TR, D. I., "Ebola in west africacdcs role in epidemic detection, control, and prevention.", (2015).

Fuhrman, K., G. Pinter and J. Berges, "Dynamics of a virus–host model with an intrinsic quota", Mathematical and Computer Modelling **53**, 5, 716–730 (2011).

Funk, D. J. and A. Kumar, "Ebola virus disease: an update for anesthesiologists and intensivists", Canadian Journal of Anesthesia/Journal canadien d'anesthésie **62**, 1, 80–91 (2015).

Funk, S., M. Salathé and V. A. Jansen, "Modelling the influence of human behaviour on the spread of infectious diseases: a review", Journal of the Royal Society Interface **7**, 50, 1247–1256 (2010).

Gage, K. L., T. R. Burkot, R. J. Eisen and E. B. Hayes, "Climate and vectorborne diseases", American journal of preventive medicine **35**, 5, 436–450 (2008).

Gao, D. and S. Ruan, "A multipatch malaria model with logistic growth populations", SIAM journal on applied mathematics **72**, 3, 819–841 (2012).

Gatacre, W., "Report: Bubonic plague in bombay for 1896-97", (1897).

Geddes, A. M., "The history of smallpox", Clinics in dermatology **24**, 3, 152–157 (2006).

Gilliam, J. F., "The plague under marcus aurelius", The American Journal of Philology **82**, 3, 225–251 (1961).

Gomes, M. F., A. P. y Piontti, L. Rossi, D. Chao, I. Longini, M. E. Halloran and A. Vespignani, "Assessing the international spreading risk associated with the 2014 West African Ebola outbreak", PLOS Currents Outbreaks **1** (2014).

Gourley, S. A., Y. Kuang and J. D. Nagy, "Dynamics of a delay differential equation model of hepatitis b virus infection", Journal of Biological Dynamics **2**, 2, 140–153 (2008).

Haeser, H., *Lehrbuch der Geschichte der Medicin und der epidemischen Krankheiten*, vol. 2 (Hermann Dufft, 1881).

Hsieh, Y. and Y. Cheng, "Real-time forecast of multiphase outbreak", Emerging Infectious Diseases **12**, 1, 122 (2006).

Hsieh, Y.-H. and C. Chen, "Turning points, reproduction number, and impact of climatological events for multi-wave dengue outbreaks", Tropical Medicine & International Health **14**, 6, 628–638 (2009).

Jones, J. D. and J. L. Dangl, "The plant immune system", Nature **444**, 7117, 323–329 (2006).

Karlsson, J., M. Anguelova and M. Jirstrand, "16th ifac symposium on system identification an efficient method for structural identifiability analysis of large dynamic systems*", IFAC Proceedings Volumes **45**, 16, 941 – 946, URL `http://www.sciencedirect.com/science/article/pii/S1474667015380745` (2012).

Keeling, M. and C. Gilligan, "Bubonic plague: a metapopulation model of a zoonosis", Proceedings of the Royal Society of London B: Biological Sciences **267**, 1458, 2219–2230 (2000a).

Keeling, M. and C. Gilligan, "Metapopulation dynamics of bubonic plague", Nature **407**, 6806, 903–906 (2000b).

Kermack, W. O. and A. G. McKendrick, "A contribution to the mathematical theory of epidemics", in "Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences", vol. 115, pp. 700–721 (The Royal Society, 1927).

Khan, A., M. Naveed, M. Dur-e Ahmad and M. Imran, "Estimating the basic reproductive ratio for the Ebola outbreak in Liberia and Sierra Leone", Infectious diseases of poverty **4**, 1, 13 (2015).

King, A. M., *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*, vol. 9 (Elsevier, 2011).

Kiskowski, M. A., "A three-scale network model for the early growth dynamics of 2014 West Africa Ebola epidemic", PLoS currents **6** (2014).

Klein, I., "Urban development and death: Bombay city, 1870–1914", Modern Asian Studies **20**, 04, 725–754 (1986).

Kuang, Y., *Delay differential equations: with applications in population dynamics*, vol. 191 (Academic Press, 1993).

Kuang, Y., J. Huisman, J. J. Elser *et al.*, "Stoichiometric plant-herbivore models and their interpretation", Mathematical Biosciences and Engineering **1**, 2, 215–222 (2004a).

Kuang, Y., J. D. Nagy and J. J. Elser, "Biological stoichiometry of tumor dynamics: mathematical models and analysis", Discrete and Continuous Dynamical Systems Series B **4**, 1, 221–240 (2004b).

Lacroix, C., E. W. Seabloom and E. T. Borer, "Environmental nutrient supply alters prevalence and weakens competitive interactions among coinfecting viruses", New Phytologist **204**, 2, 424–433 (2014).

Lagarias, J. C., J. A. Reeds, M. H. Wright and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions", SIAM Journal on optimization **9**, 1, 112–147 (1998).

Lee, H. S., *Dates in Infectious Disease: A Chronological Record of Progress in Infectious Diseases over the Last Millennium* (Taylor & Francis, 2002).

Legrand, J., R. Grais, P. Boelle, A. Valleron and A. Flahault, "Understanding the dynamics of Ebola epidemics", Epidemiology and infection **135**, 04, 610–621 (2007).

Leroy, E. M., B. Kumulungui, X. Pourrut, P. Rouquet, A. Hassanin, P. Yaba, A. Délicat, J. T. Paweska, J.-P. Gonzalez and R. Swanepoel, "Fruit bats as reservoirs of ebola virus", Nature **438**, 7068, 575–576 (2005).

Lewnard, J. A., M. L. N. Mbah, J. A. Alfaro-Murillo, F. L. Altice, L. Bawo, T. G. Nyenswah and A. P. Galvani, "Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis", The Lancet Infectious Diseases **14**, 12, 1189–1195 (2014).

Li, Z., Z. Teng, X. Feng, Y. Li and H. Zhang, "Dynamical Analysis of an SEIT Epidemic Model with Application to Ebola Virus Transmission in Guinea", Computational and mathematical methods in medicine **2015** (2015).

Littman, R. J., "The plague of athens: epidemiology and paleopathology", Mount Sinai Journal of Medicine **76**, 5, 456–467 (2009).

Mamo, D. K. and P. R. Koya, "Mathematical Modeling and Simulation Study of SEIR disease and Data Fitting of Ebola Epidemic spreading in West Africa", Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN pp. 3159–0040 (2015).

Marquardt, D. W., "An algorithm for least-squares estimation of nonlinear parameters", Journal of the society for Industrial and Applied Mathematics **11**, 2, 431–441 (1963).

May, R. M., R. M. Anderson *et al.*, "Population biology of infectious diseases: Part ii", Nature **280**, 5722, 455–461 (1979).

McLynn, F., *Marcus Aurelius: warrior, philosopher, emperor* (Random House, 2011).

Meltzer, M. I., 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", MMWR Surveill Summ **63**, suppl 3, 1–14 (2014).

Merler, S., M. Ajelli, L. Fumanelli, M. F. Gomes, A. P. y Piontti, L. Rossi, D. L. Chao, I. M. Longini, M. E. Halloran and A. Vespignani, "Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis", The Lancet Infectious Diseases **15**, 2, 204–211 (2015).

Monecke, S., H. Monecke and J. Monecke, "Modelling the black death. a historical case study and implications for the epidemiology of bubonic plague", International Journal of Medical Microbiology **299**, 8, 582–593 (2009).

Moré, J. J., "The levenberg-marquardt algorithm: implementation and theory", in "Numerical analysis", pp. 105–116 (Springer, 1978).

Neofytou, G., Y. Kyrychko and K. Blyuss, "Mathematical model of plant-virus interactions mediated by rna interference", Journal of theoretical biology **403**, 129–142 (2016a).

Neofytou, G., Y. Kyrychko and K. Blyuss, "Time-delayed model of immune response in plants", Journal of theoretical biology **389**, 28–39 (2016b).

Nielsen, C. F., S. Kidd, A. Sillah, E. Davis, J. Mermin and P. H. Kilmarx, "Improving burial practices and cemetery management during an ebola virus disease epidemic-Sierra Leone, 2014", MMWR Surveill Summ **64**, 1–8 (2015).

Nishiura, H. and G. Chowell, "The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends", in "Mathematical and statistical estimation approaches in epidemiology", pp. 103–121 (Springer, 2009).

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

Pardoe, I. and S. Weisberg, "An Introduction to bootstrap methods using Arc", Unpublished Report available at www. stat. umn. edu/arc/bootmethREV. pdf (2001).

Pell, B., J. Baez, T. Phan, D. Gao, G. Chowell and Y. Kuang, "Patch models of EVD transmission dynamics", in "Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases", edited by Chowell and Hyman (Springer, 2016).

Poole, J. and A. Holladay, "Thucydides and the plague of athens", The Classical Quarterly (New Series) **29**, 02, 282–300 (1979).

Portz, T., Y. Kuang and J. D. Nagy, "A clinical data validated mathematical model of prostate cancer growth under intermittent androgen suppression therapy", Aip Advances **2**, 1, 011002 (2012).

Rachah, A. and D. F. Torres, "Mathematical Modelling, Simulation, and Optimal Control of the 2014 Ebola Outbreak in West Africa", Discrete Dynamics in Nature and Society (2015).

Raue, A., J. Karlsson, M. P. Saccomani, M. Jirstrand and J. Timmer, "Comparison of approaches for parameter identifiability analysis of biological systems", Bioinformatics URL `http://bioinformatics.oxfordjournals.org/content/early/2014/01/29/bioinformatics.btu006.abstract` (2014).

Raue, A., C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmller and J. Timmer, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood", Bioinformatics **25**, 15, 1923–1929, URL `http://bioinformatics.oxfordjournals.org/content/25/15/1923.abstract` (2009).

Raven, P. H., R. F. Evert and S. E. Eichhorn, *Biology of plants* (Macmillan, 2005).

Retief, F. P. and L. Cilliers, "The epidemic of athens, 430-426 bc.", South African medical journal= Suid-Afrikaanse tydskrif vir geneeskunde **88**, 1, 50–53 (1998).

Richards, F., "A flexible growth function for empirical use", Journal of experimental Botany **10**, 2, 290–301 (1959).

Rivers, C. M., 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 currents **6** (2014).

Rogers, L., "The methods and results of forecasting the incidence of cholera, smallpox and plague in india", Transactions of the Royal Society of Tropical Medicine and Hygiene **27**, 3, 217–238 (1933).

Ross, R., "On some peculiar pigmented cells found in two mosquitos fed on malarial blood", British medical journal **2**, 1929, 1786 (1897).

Ross, R., *The prevention of malaria* (Dutton, 1910).

Saccomani, M. P., S. Audoly and L. D'Angi, "Parameter identifiability of nonlinear systems: the role of initial conditions", Automatica **39**, 4, 619 – 632, URL `http://www.sciencedirect.com/science/article/pii/S0005109802003023` (2003).

Schotthoefer, A. M., S. W. Bearden, J. L. Holmes, S. M. Vetter, J. A. Montenieri, S. K. Williams, C. B. Graham, M. E. Woods, R. J. Eisen and K. L. Gage, "Effects of temperature on the transmission of yersinia pestis by the flea, xenopsylla cheopis, in the late phase period", Parasit Vectors **4**, 1, 191 (2011).

Shaman, J., W. Yang and S. Kandula, "Inference and forecast of the current West African Ebola outbreak in Guinea, Sierra Leone and Liberia", PLoS currents **6** (2014).

Smith, H., *An introduction to delay differential equations with applications to the life sciences*, vol. 57 (Springer Science & Business Media, 2010).

Snow, J., *On the mode of communication of cholera* (John Churchill, 1855).

Stenseth, N. C., B. B. Atshabar, M. Begon, S. R. Belmain, E. Bertherat, E. Carniel, K. L. Gage, H. Leirs and L. Rahalison, "Plague: past, present, and future", PLoS Med **5**, 1, e3 (2008).

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

Trębicki, P., R. K. Vandegeer, N. A. Bosque-Pérez, K. S. Powell, B. Dader, A. J. Freeman, A. L. Yen, G. J. Fitzgerald and J. E. Luck, "Virus infection mediates the effects of elevated co2 on plants and vectors", Scientific reports **6** (2016).

Tromas, N., M. P. Zwart, G. Lafforgue and S. F. Elena, "Within-host spatiotemporal dynamics of plant virus infection at the cellular level", PLoS Genet **10**, 2, e1004186 (2014).

Valdez, L., H. H. A. Rêgo, H. Stanley and L. Braunstein, "Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies", arXiv preprint arXiv:1502.01326 (2015).

Van Kinh Nguyen, S. C. B., A. Boianelli, M. Meyer-Hermann and E. A. Hernandez-Vargas, "Ebola virus infection modeling and identifiability problems", Frontiers in microbiology **6** (2015).

Viboud, C., L. Simonsen and G. Chowell, "A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks", Epidemics 15 (2016).

Wayne, R. O., *Plant cell biology: from astronomy to zoology* (Academic Press, 2009).

Weitz, J. S. and J. Dushoff, "Post-death Transmission of Ebola: Challenges for Inference and Opportunities for Control", arXiv preprint arXiv:1411.3435 (2014).

WHO Ebola Response Team, "Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections", N Engl J Med **371**, 16, 1481–95 (2014).

World Health Organization, "Ebola Response Roadmap Situation report 03-05-2015", URL `http://www.who.int/csr/disease/ebola/situation-reports/en/` (2015a).

World Health Organization, "Ebola response: What needs to happen in 2015", URL `http://www.who.int/csr/disease/ebola/one-year-report/response-in-2015/en/` (2015b).

World Health Organization, "Ebola vaccines, therapies, and diagnostics", URL `http://www.who.int/medicines/emp_ebola_q_as/en` (2015c).

World Health Organization, "Guidance for Safe Handling of Human Remains of Ebola Patients in U.S. Hospitals and Mortuaries", URL `http://www.cdc.gov/vhf/ebola/healthcare-us/hospitals/handling-human-remains.html` (2015d).

Wu, J., R. Dhingra, M. Gambhir and J. V. Remais, "Sensitivity analysis of infectious disease models: methods, advances and their application", Journal of The Royal Society Interface **10**, 86, URL `http://rsif.royalsocietypublishing.org/content/10/86/20121018` (2013).

www.ebola-challenge.org, URL `www.ebola{-}challenge.org` (2016).

Yamin, D., S. Gertler, M. L. Ndeffo-Mbah, L. A. Skrip, M. Fallah, T. G. Nyenswah, F. L. Altice and A. P. Galvani, "Effect of Ebola progression on transmission and control in Liberia", Annals of internal medicine **162**, 1, 11–17 (2015).

Yu, H.-L. and G. Christakos, "Spatiotemporal modelling and mapping of the bubonic plague epidemic in india", International journal of health geographics **5**, 1, 1 (2006).

Yu, H.-L., A. Kolovos, G. Christakos, J.-C. Chen, S. Warmerdam and B. Dev, "Interactive spatiotemporal modelling of health systems: the seks–gui framework", Stochastic Environmental Research and Risk Assessment **21**, 5, 555–572 (2007).

APPENDIX A

FORECASTING AND FITTING ERROR TABLES

District	Logistic		Two-Patch (H)		Two-Patch		Three-Patch (H)		Three-Patch (S)		Three-Patch	
	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds
BOMI	3.6147	1.198	1.7439	0.39968	2.23	0.49683	1.9891	0.4395	3.1711	0.52022	3.9724	0.47549
BONG	1.4365	0.54448	1.5116	0.42227	1.6447	0.46937	1.6753	0.5351	2.1265	0.49347	2.9081	0.58443
GBARPOLU	0.71881	0.078303	0.6592	0.087046	1.001	0.0887	1.1162	0.095164	1.1436	0.077254	2.5221	0.11758
GRAND BASSA	1.8093	0.68238	1.1774	0.29884	1.8118	0.41695	1.7176	0.38735	2.2839	0.44523	3.2918	0.58292
GRANDAPE MOUNT	3.333	0.5762	3.7319	0.63049	3.8035	0.65232	3.3711	0.67542	5.0092	0.73396	8.0373	0.83729
GRAND KRU	0.61518	0.17033	0.80231	0.099753	0.84577	0.17673	0.64961	0.12372	1.0854	0.1753	2.3581	0.2528
LOFA	5.2755	0.36486	5.6001	0.26833	6.9035	0.36263	7.9303	0.34862	6.4173	0.27562	11.1918	0.47298
MARGIBI	3.4293	0.73833	4.4185	0.67457	4.3363	0.72399	4.5471	0.72418	5.619	0.79608	7.7291	0.89117
MONTSERRADO	4.2503	8.3487	4.922	2.2296	5.2784	1.8897	5.6934	2.3507	6.7871	1.5	9.1043	2.424
NIMBA	1.2104	0.12813	1.351	0.13701	1.4561	0.14135	1.5792	0.1464	1.9673	0.15578	2.7993	0.17449
RIVER GEE	0.15977	0.01381	0.1785	0.015599	0.17445	0.015513	0.18504	0.014329	0.23398	0.019005	0.3597	0.019446
RIVERCESS	0.68386	0.077873	0.51155	0.093633	0.54034	0.027355	1.1207	0.1403	1.6564	0.10546	3.84	0.20521
SINOE	0.71361	0.074711	0.76264	0.069135	0.92105	0.089601	1.1015	0.095067	1.5029	0.099703	2.8564	0.13226
CONAKRY	2.4851	2.2749	2.4127	2.428	3.1721	2.443	3.4615	2.5474	4.0175	2.6153	7.2811	2.9771
COYAH	2.9429	3.1608	1.6338	3.4047	2.0082	2.9153	3.1834	3.1261	5.2135	3.2436	5.4088	4.6479
DUBREKA	1.8975	1.0283	0.84368	1.0297	2.5879	0.70316	5.2154	1.0243	3.0884	1.0003	5.7692	1.1607
FARANAH	2.1281	0.37701	2.1799	0.42643	2.4097	0.41307	3.8266	0.50515	5.0919	0.53024	25.1464	0.24839
FORECARIAH	1.6364	3.1111	2.0821	2.8257	2.2314	3.3816	1.6743	13.7987	2.9266	3.4089	5.3969	3.8571
KANKAN	1.1965	0.14143	2.3918	0.1566	1.4882	0.1561	3.15	0.49163	3.0055	0.16257	3.7905	0.23459
KINDIA	2.0281	1.136	1.1496	0.91487	1.6464	1.071	1.5116	0.38334	2.5737	1.0592	5.3894	1.2542
KISSIDOUGOU	2.9894	0.40939	2.3094	0.47537	3.0904	0.53037	2.3511	1.1683	5.3364	0.39528	21.0685	0.77644
MACENTA	0.4115	5.2217	1.6022	5.5868	1.8795	5.3856	2.0131	0.62805	4.0898	5.9809	6.7702	6.1607
NZEREKORE	2.5579	0.38891	0.44257	0.40024	0.42032	0.41519	1.0254	5.7162	1.3283	0.49414	0.68875	0.53331
SIGUIRI	0.44141	0.30196	3.1856	0.32252	3.5061	0.25171	3.8516	0.4314	5.0128	0.35819	8.8683	0.41559
TELIMELE	0.10555	0.55761	0.42019	0.593	0.4517	0.62469	0.65173	0.34477	0.75666	0.6742	1.8585	0.72794
BO	5.0701	2.2253	0.12915	2.4011	0.13449	2.0631	0.175	0.61298	0.15866	2.4455	0.24843	2.6633
BOMBALI	8.8538	9.0596	1.9431	3.2113	1.7743	4.476	4.0898	2.2312	5.8781	3.8964	6.1361	4.9203
KAILAHUN	13.4441	2.3216	3.0016	0.67132	3.8811	0.66044	4.9654	3.3201	7.8443	2.2252	8.5021	1.2032
KAMBIA	3.0924	2.8205	5.1435	1.6991	3.718	1.1592	7.6597	1.0176	11.6425	1.4946	13.382	1.5634
KENEMA	10.8483	2.4777	1.3423	1.2226	2.1451	0.83164	3.4772	1.4973	4.8896	0.85422	4.7483	1.0399
KOINADUGU	2.18	0.38465	8.1016	0.25254	4.9844	0.44878	6.3207	1.2681	15.0373	0.35363	6.2745	0.40977
KONO	5.3272	3.2919	2.7568	3.3897	3.1007	3.3525	1.7213	0.4554	4.7125	3.2659	7.5391	3.6874
MOYAMBA	2.18	0.38465	6.7662	0.25254	6.8282	0.44878	7.4665	3.6681	9.9778	0.35363	15.8279	0.40977
PORT LOKO	15.4465	15.1163	2.7568	2.2217	3.1007	6.4625	1.7213	0.4554	4.7125	14.2096	7.5391	2.7181
PUJEHUN	0.22985	0.02164	16.4268	0.017028	11.313	0.017084	13.17	3.6144	12.6916	0.017823	17.257	0.048621
TONKOLILI	11.6003	3.3944	0.18359	1.598	0.29053	1.9964	0.2574	0.016134	0.42985	4.0286	0.62986	1.8315
PUJEHUN	0.22985	0.02164	7.3516	0.017028	4.5375	0.017084	4.5832	1.6187	6.898	0.017823	10.2676	0.048621
WESTERN AREA RURAL	8.5967	11.4656	0.18359	3.0142	0.29053	5.9014	0.2574	0.016134	0.42985	7.0712	0.62986	4.0112
WESTERN AREA URBAN	5.639	13.5544	5.7848	8.6758	6.0817	8.771	6.0987	3.1066	7.3591	10.3357	8.545	16.817

Table A.1: Fitting errors for all models. Models were trained on one-third and two-thirds of each district data set.

APPENDIX B

EBOLA MODEL REVIEW TABLE

Summary/Result/Citation	Data	Deterministic or Stochastic	Spatial or non-spatial	Phenom.* or Mechanistic
Statistical model: a global reproduction number and expected number of additional cases were calculated from the expected value of EVD incidence. **Results**: R_t consistently lies above one for three countries and below two for Sierra Leone and Liberia from June to August 2014. (Nishiura and Chowell, 2014)	Spline-fitted of data reported by WHO.	Stochastic: assumes EVD incidence is Poisson distributed and generation time is exponentially distributed.	Spatial: transmission matrix for between and within countries.	Mechanistic: models transmission route and mixing of individuals.
SEIR model: the average reproduction number, number of assessment/treatment beds and a forecast of addition cases were estimated based on a Bayesian fit of the collected data. **Results**: R_0 is near the threshold of 1 in December 2014. Bed capacity lags behind, but since transmission decreases, control measures were able to meet the demand in early 2015. (Camacho *et al.*, 2015)	WHO patient database and daily situation reports from August 10, 2014 to January 18, 2015.	Stochastic: assumes incubation period is Erlang distributed and contains other over-dispersion parameters.	Non-spatial: Although, the infectious group is separated into community and hospital.	Mechanistic: Although, the bed demand is estimated based on the separation of the infectious group in a phenomeno-logical way.
SEIRD/SEIR model: points out the problem with identifiability of estimating epidemiological parameters from early-stage epidemic data. **Results**: neglecting post-death transmission (PDT) leads to underestimating R_0. PDT is a key factor to modeling and control EVD.(Weitz and Dushoff, 2014)	Cumulative patient data (WHO) of the first 9 months of the epidemic.	Stochastic: the exposed period for the latently infected is gamma distributed.	Non-spatial	Mechanistic: models the post death transmission along with its effect on R0.
An expanded version of the SEIR model which includes hospitalization and post-death transmission (SEIHFR). **Results**: reducing mobility can delay, but has no effect on the overall control of EVD. Safe burials and hospitalization are keys to control.(Valdez *et al.*, 2015)	Cumulative patients data from March to August of 2014 (WHO).	Stochastic (using Gillespie algorithm): with a quasi-deterministic representation (ODEs) that takes over after a threshold.	Spatial: takes into account of traveling between 15 counties of Liberia.	Mechanistic: transmission between counties is accounted along with post-death transmission, hospitaliza-tion and others.

Continued on next page

Continued from previous page

Summary/Result/Citation	Data	Deterministic or Stochastic	Spatial or non-spatial	Phenom.* or Mechanistic
SIRD model: natural born/death is accounted for. **Results**: even with a small initial number of infected people, without proper prevention, the disease can be spread to the entire population. (Atangana and Goufo, 2014)	No data. Simulations with parameters from literature.	Deterministic	Non-spatial	Mechanistic: examine transmission mechanism of infectious individuals.
SEIHFR model: investigated the impact of increased contact tracing and pharmaceutical interventions. **Results**: contact tracing and infection control are beneficial, but not sufficient to control EVD. Pharmaceutical intervention reduces mortality, but not much of the overall size. (Rivers *et al.*, 2014)	Cumulative patient data up to Oct. 5, 2014.	Deterministic: Model fitted.	Non-spatial	Mechanistic: examines the effect of several mitigation strategies on the epidemic.
Stage-structure SEIRF: model parameters are estimated within a realistic range, transmission and interventions are investigated stage by stage. **Results**: transmission is significant at any stage (early, late, or burial). (Eisenberg *et al.*, 2015)	Cumulative patients data (WHO) up to October 1, 2014.	Stochastic: the LH sampling method is used instead of the usual likelihood-based approaches.	Non-spatial	Mechanistic: intervention methods during infection stage and transmissions by recovered individuals are examined.
SIR model: adjusted to account for vaccination rate (constant) and optimal control (vaccination rate that changes with time). **Results**: vaccination should be used as part of an optimal control strategy to quickly control EVD. (Rachah and Torres, 2015)	Cumulative patients data (WHO) up to Oct. 26, 2014.	Deterministic: data fitting.	Non-spatial	Mechanistic: investigate the effect of vaccination at different rate and stage of the epidemics.

Continued on next page

Continued from previous page

Summary/Result/Citation	Data	Deterministic or Stochastic	Spatial or non-spatial	Phenom.* or Mechanistic
SIRD model: under different assumptions and the effect of increasing treatment centers, case ascertainment or allocating protective kits, the averted number of cases by Dec 15, 2014 is estimated. **Results**: R_0 is 2.49 for Montserrado. Protective kits can help with prevention of EVD, but not sufficient to control it; hospital beds are crucial for control. (Lewnard *et al.*, 2014)	Cumulative patients data (WHO) from June 14, 2014 to Sept. 23, 2014 for Montserrado.	Stochastic: Bayesian Markov Chain Monte Carlo sampling is used to obtain uncertainty of parameters.	Non-spatial	Mechanistic: examine the effect of increasing treatment centers, case ascertainment and allocating protective kits on averting EVD cases, individually or together.
SEIT model: stability analysis of the disease free and endemic equilibrium with parameter estimation by least squares method. R_0 is estimated (4.6) along with its uncertainty. **Results**: R_0 is sensitive to the contact rate, especially in the latent and exposed period. (Li *et al.*, 2015)	Cumulative patients data (WHO) from March 22, 2014 to January 25, 2015 for Guinea.	Deterministic: fitting the model to the data.	Non-spatial	Mechanistic: examine transmission mechanism of infectious individuals.
The Global Epidemic and Mobility (GEM) Model: integrates real demographics to identify over 3,300 subpopulations areas from about 220 countries. An SEIHFR model is coupled to study the transmission dynamics of the EVD within each subpopulation. **Results**: EVD is more likely to spread within Africa, which can increase the international risk over a long period of time. (Gomes *et al.*, 2014)	Cumulative patients data (WHO) and census population and global air travel databases.	Stochastic: aside from GEM model, Monte Carlo simulations, LH sampling and other stochastic methods are used to simulate the spreading of EVD locally and globally.	Spatial: taking into account the travel between each census areas.	Mechanistic: examine extensively the transmission of EVD by air between different regions of the world from the three host countries.

Continued on next page

Continued from previous page

Summary/Result/Citation	Data	Deterministic or Stochastic	Spatial or non-spatial	Phenom.* or Mechanistic
SEIRD model: R_0 is estimated from the individual reproduction number, then used to estimate risk of an outbreak by an undetected case. **Results**: quick intervention (within one serial interval) can greatly limit the transmission of EDV. (Althaus *et al.*, 2015b)	Daily incidence data	Stochastic: Gamma distribution for the incubation period and Poisson for incident and death cases.	Non-spatial	Mechanistic: examine the effect of control measures on an early epidemic.
SEIR: incorporates contact tracing and derives a novel approach to calculating R_e directly from contact tracing and reported case data. **Results**: R_e can be estimated solely based on the tracing quantities. (Browne *et al.*, 2015)	Cumulative patients data and traced cases for Guinea and Sierra Leone (WHO)	Stochastic: maximum likelihood approach is used to estimate R_e with Weibull distributed serial interval.	Non-spatial	Mechanistic: investigate various uses of contact tracing data
SEIRX-EAKF (Extended Adjusted Kalman Filter) model: incorporates the deceased population for data assimilation purpose. EAKF alters R_e and other parameters. **Results**: Forecasts 6 weeks into the future projects some slowing down of the current outbreak. (Shaman *et al.*, 2014)	Cumulative patients data (WHO) Oct., 2014.	Stochastic: uniform distribution for initial data, and multiple stochastic parameters.	Non-spatial: stochasticity and mixed construction may implicitly show some.	Mechanistic: investigate transmission dynamics of EVD.
Target cell-limited model: practical identifiability aspect is explored by examining the parameter space and re-optimization. **Results**: EBOV infection time is much slower compared to influenza virus, but this is compensated by its high replication rate. (Van Kinh Nguyen *et al.*, 2015)	Experimental data of EBOV kinetics	Stochastic: bootstrap method.	Non-spatial	Mechanistic: provides the quantitative details of the interaction between Ebola virus and the host cells.

Continued on next page

Continued from previous page

Summary/Result/Citation	Data	Deterministic or Stochastic	Spatial or non-spatial	Phenom.* or Mechanistic
SEIR model: highly expanded to incorporate many aspect of the EVD epidemic such as demographics. **Results**: the belief system can affect greatly the value of R_0. Increasing health care workers, limiting hospital visitation and educating people on their burial practice help with EVD control. (Agusto *et al.*, 2015)	Cumulative patients data (WHO) for Guinea in 2014.	Deterministic	Spatial: the transmission between different settings (hospital and community).	Mechanistic: examine various details of the transmission of EVD under different public health control hypothesis.
SEIHR model: the susceptible population is divided into low and high-risk groups. Possible methods to terminate the current epidemics are examined. **Results**: significant level of isolation and reduction of high risk population can help bring R_0 below one.(Khan *et al.*, 2015)	Cumulative patients data (WHO) and corrected data (CDC) from May 1, 2014 to Oct. 1, 2014	Deterministic	Non-spatial	Mechanistic: investigate the effect of altering the high risk susceptible population to transmission of EVD.
Data-driven stochastic model: considers 3 stages of infection: incubation, early symptomatic and late symptomatic. **Results**: R_0 is calculated to be between 1 and 2. Isolation of critically infected individual early can eliminate the disease. (Yamin *et al.*, 2015)	Incidence and case fatality from July 07, 2014 to Sept. 22, 2014. Contact-tracing from Aug. 7, 2014 to Aug. 26, 2014.	Stochastic: many distributions built from empirical data are used.	Non-spatial	Mechanistic: examine the impact of disease progression and case fatality and targeted intervention on EVD transmission.
SEIR model: a mathematical survey of SEIR models and their expanded versions is done. Parameters and R_0 are estimated through data fitting. **Results**: the simulations with various values of R_0 are used to confirm the known dynamics of the outbreak when R_0 is below or above unity. (Mamo and Koya, 2015)	Patient data (WHO) up to Dec. 14, 2014.	Deterministic.	Non-spatial	Mechanistic: looks at the transmission and case fatality dynamics of EVD.

Continued on next page

Continued from previous page

Summary/Result/Citation	Data	Deterministic or Stochastic	Spatial or non-spatial	Phenom.* or Mechanistic
Multi-type branching model: forecasts are done under different time-varying interventions. A novel method for model parameterization is introduced. **Results**: hospital capacity is crucial in controlling the epidemics. If hospital capacity is low, then even a 99% hospitalization rate may not contain the outbreak. (Drake *et al.*, 2015)	Time series of operational Ebola Treatment Units and capacity is estimated. Patient data (WHO) up to September 2014.	Stochastic: Poisson distribution is assumed various infection assumptions.	Spatial: incorporates spatial heterogeneity in transmission dynamics.	Mechanistic: analyzes the effect of increasing the number of hospital beds and safe burial to the transmission dynamics of EVD.
SIR model: accounts for the incubation period, burial practice, underreporting and the number of hospital beds. **Results**: without increase in treatment, the number of patients will double to 8,000 by Sept. 30, 2014. 70% of patients being treated by late Dec. 2014 is a threshold to eliminate the disease. (Meltzer *et al.*, 2014)	Patient data (WHO) up to Aug. 28, 2014 and up to Sept. 13, 2014. Past information is used to estimate the mean incubation and infection period.	Stochastic: distributions of incubation and infection period are used.	Non-spatial: However, different settings are considered.	Mechanistic: quantitatively show the cost of delaying control, the condition for effectively eliminate the disease.
SEIHFR model: sensitivity analysis is carried out to obtain the most key factor in controlling EVD. **Results**: R_0 was estimated to be 2.7. The time until interventions take effect and hospitalization rate is crucial to control EVD. (Legrand *et al.*, 2007)	Patient data from WHO and various papers for parameters values.	Stochastic: weekly incidence is assumed to be Poisson distributed. 95% confidence intervals are computed using LHS.	Non-spatial: however, different settings are considered.	Mechanistic: compare the various aspects of EVD in two different outbreaks, Democratic Republic of Congo (1995) and Uganda (2000).

Table B.1: Literature review of modeling efforts of EVD.
* Phenomenological.

Concluded