AN ESTIMATE OF THE BASIC REPRODUCTIVE NUMBER, R_0, FOR THE 2015-2016 ZIKA VIRUS OUTBREAK IN PUERTO RICO

By

Félix M. Pabón-Rodríguez

A **thesis** submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

MATHEMATICAL STATISTICS

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS

2018

Approved by:

Edgardo Lorenzo-González, Ph.D. Member, Graduate Committee	Date
Dámaris Santana-Morant, Ph.D. Co-President, Graduate Committee	Date
Karen Ríos-Soto, Ph.D. President, Graduate Committee	Date
José Cedeño-Maldonado, Ph.D. Representative of Graduate Studies	Date
Victor Ocasio-González, Ph.D. Chairperson of the Department	Date

Abstract of Dissertation Presented to the Graduate School
of the University of Puerto Rico in Partial Fulfillment of the
Requirements for the Degree of Master of Science

AN ESTIMATE OF THE BASIC REPRODUCTIVE NUMBER, R_0, FOR THE 2015-2016 ZIKA VIRUS OUTBREAK IN PUERTO RICO

By

Félix M. Pabón-Rodríguez

May 2018

Advisor: Dr. Karen Ríos-Soto
Major Department: Department of Mathematical Sciences

In recent years, vector-borne diseases are taking serious attention from researchers and health specialist across the world. The emergence of vector-borne diseases, such as Chikungunya and Zika, coupled with outbreaks of both diseases in the Americas are of great interest to the scientific community since there is still very much to learn about their transmission, risks and effects.

The Zika virus (ZIKV) is primarily transmitted by infected females Aedes aegypti mosquitoes, but there is also confirmed evidence that it can be transmitted directly (human to human) by sexual contacts and from mother to fetus. The apparent effects in the neurological system through the Guillain-Barré syndrome and the neonate microcephaly are of great concern. In Puerto Rico, 66% of the confirmed cases were females, and since the Zika disease is usually asymptomatic, pregnant women may not even know that they have the virus. At the end of December 2016, the Puerto Rico Department of Health estimated 37,500 cases.

In this work, we focus on the 2015-2016 Zika virus (ZIKV) outbreak in Puerto Rico and use the data of confirmed Zika cases by laboratory obtained from the weekly

reports published by the Puerto Rico Department of Health. To analyze the behavior of the Zika virus in Puerto Rico, a mathematical model that takes into account vector and sexual transmission is considered. Using the data and the epidemic model, the initial exponential growth rate of the epidemic is estimated, (defined as the force of infection), by different statistical methods, in order to estimate the basic reproductive number (R_0) of the Zika epidemic in Puerto Rico. In addition, the Ordinary Least Squares (OLS) and Generalized Least Squares (GLS) methods were considered to minimize the point-by-point distances between the predicted data by the mathematical model and the observed data for the Zika epidemic in Puerto Rico. The optimization procedure was performed to estimate the transmission rates β_h and κ (that are unknown) in order to estimate and generate a distribution for R_0, using the parameters from the model through a sampling process.

Keywords: Zika virus, estimation, data analysis, statistical methods, Zika epidemic model, basic reproductive number, force of infection, sexual transmission, Bayesian inference, Negative Binomial distribution

Resumen de Disertación Presentado a Escuela Graduada
de la Universidad de Puerto Rico como requisito parcial de los
Requerimientos para el grado de Maestría en Ciencias

UN ESTIMADO DEL NÚMERO REPRODUCTIVO BÁSICO, R_0, PARA EL BROTE DEL VIRUS DE ZIKA 2015-2016 EN PUERTO RICO

Por

Félix M. Pabón-Rodríguez

Mayo 2018

Consejero: Dr. Karen Ríos-Soto
Departamento: Departamento de Ciencias Matemáticas

En la actualidad, las enfermedades transmitidas por vectores están tomando mucha atención por parte de los investigadores y especialistas de la salud, de distintas partes del mundo. Enfermedades emergentes como Chikungunya y Zika, y situaciones donde ambos virus circulan al mismo tiempo, en las Américas, son de gran interés para la comunidad científica, dado que aún hay mucho por aprender acerca de la transmisión, riesgos y efectos.

El virus del Zika (ZIKV) es transmitido principalmente, por la picada de un mosquito hembra, del tipo *Aedes Aegypti*, pero existe evidencia que confirma que el virus también puede ser transmitido de forma directa, es decir, de humano a humano, como por ejemplo, por transmisión sexual o de la madre al feto, en caso de un embarazo. Los efectos aparentes en el sistema neurológico por medio del Síndrome de Guillain-Barré y la Microcefalia en los bebés, causa mucha preocupación. En Puerto Rico, durante la epidemia del 2016, 66% de los casos confirmados fueron féminas, y debido a que el virus del Zika es usualmente asintomático, las mujeres embarazadas podrían tener el virus, sin estar conscientes de que lo tienen. A finales

de diciembre 2016, el Departamento de Salud de Puerto Rico estimó una incidencia acumulada aproximada de 37,500 casos.

En este trabajo, nos enfocamos en el brote de Zika ocurrido entre 2015 y 2016, en Puerto Rico, utilizando los datos obtenidos de los reportes del Departamento de Salud de Puerto Rico. Para analizar el comportamiento del virus del Zika en Puerto Rico, se considera un modelo matemático que toma en cuenta la transmisión vectorial y de forma directa (sexual). Utilizando los datos y el modelo de la epidemia, la tasa de crecimiento exponencial inicial de la epidemia (definida como la fuerza de infección) es estimada por diversos métodos estadísticos, con el propósito de estimar el número reproductivo básico, R_0, para la epidemia 2015-2016 del Zika en Puerto Rico. En adición, para minimizar las distancias punto-a-punto, entre los valores que predice el modelo matemático y los datos observados para la epidemia de Zika en Puerto Rico, los métodos de Mínimos Cuadrados Ordinarios y Mínimos Cuadrados Generalizados fueron considerados. Este procedimiento de optimización permitirá obtener estimados para las tasas de transmisión β_h y κ (que son desconocidas) con el propósito de estimar y generar una distribución de R_0, utilizando los parámetros del modelo, a través de un proceso de muestreo.

Palabras claves: Virus del Zika, estimación, análisis de datos, métodos estadísticos, modelo de epidemia del Zika, número reproductivo básico, fuerza de infección, transmisión sexual, inferencia Bayesiana, distribución Negativa Binomial

DEDICATION

To my parents, Lourdes Rodríguez Gutierrez and Osvaldo Pabón Quiñones, and to my brother, Osvaldo Pabón Rodríguez, who have given me the love and support to always move forward.

ACKNOWLEDGMENTS

First, I want to thank God, who is guiding me with peace and showing me the right way to success.

To my family. They are my reason to keep going.

To Karen Ríos-Soto and Dámaris Santana-Morant, for all the advice they gave me during the process of this research. For sharing with me your knowledge and academic experiences. For helping me to be prepared for what I may face in the future.

I want to say thanks to all my friends for the support, especially to Jennifer Guzmán and Ángel Figueroa, for always being present and for helping me to keep fighting and never give up.

Also, I want to thanks Dr. Reyes Ortiz-Albino and Dr. Olgamary Rivera, because they always found a way to motivate me to achieve what I want. I admire you both. They are my model to follow, as leaders, researchers, and educators.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The Zika virus is a vector-borne disease which is caused primarily by the bite of an infected female *Aedes Aegypti* mosquito. The first case was identified in Uganda, Africa, in 1947, when a group of scientists were conducting routine surveillance for yellow fever, in a sentinel rhesus monkey. Years later, in 1952, the first Zika case in humans was confirmed, also in Uganda [77]. According to the World Health Organization [77], outbreaks of Zika virus disease have been recorded in Africa, the Americas, Asia and the Pacific.

Figure 1–1, shows the Zika virus spread around the world, starting from Africa and moving to U.S., the U.K., Denmark and recently to Germany. The first large outbreak of disease caused by Zika infection, according to the Worlf Health Organization [77], was reported from the Island of Yap in 2007, where the 73% of the total population got the virus. Some recent Zika outbreaks are presented in Table 1–1, while Figure 1–2 shows the affected areas before 2015 and during the 2015-2016 Zika epidemics worldwide.

Year	Region	Cumulative Incidence	Population
2013-2014	French Polynesia	30,000	277 thousands
2015	Barranquilla, Colombia	65,750	1.2 million
2015-2016	Puerto Rico	37,500	3.7 million

Table 1–1: Recent Zika virus outbreaks

To understand how Zika is spreading around the world, Lancaster University and the World Health Organization (WHO) [79] created a map illustrating the flow

of the Zika infections. As we know, the first case of Zika virus in humans was identified in Uganda, Africa, in 1952, marked as 1 in Figure 1–1. Years later, a human case was detected in Nigeria, Africa (2 in Figure 1–1). In 1966, a first case was confirmed in South East Asia, and in the late 1970s it was documented in Pakistan, India, Malaysia and Indonesia (3 in Figure 1–1). The first official Zika epidemic was on the isolated island of Yap, Micronesia, in 2007 (4 in Figure 1–1), affecting 73% of the total population. The Zika virus hit French Polynesia in 2013, with a huge outbreak (5 in Figure 1–1) and later in 2014, Zika arrived in northern Brazil (6 in Figure 1–1).

It spread slowly through Brazil for around a year, before the WHO reported the first outbreak outside of Colombia in October 2015. Other countries then followed quickly, with transmission reported in Colombia, Suriname, El Salvador, and Guatemala. The disease spread to Mexico for the first time (raising concerns in the U.S.) as well as to Paraguay, Venezuela and Panama. In 2015, Honduras, Puerto Rico, French Guyana and Martinique, all in the Caribbean, had cases. Cases have also been confirmed in Guyana, Barbados, Ecuador, Bolivia, Haiti and the Dominican Republic. Some cases have been reported locally in the U.S., the U.K., Denmark and recently in Germany (7 in Figure 1–1). Lancaster University and WHO (which gave us a brief historical background about the spreading of the Zika virus) concluded that the Zika virus has circumnavigated the globe and then it could legitimately be described as a pandemic.

Zika virus is primarly transmitted to people through the bite of a female infected mosquito, but direct transmission of the virus is also possible, for example, by sexual transmission. Scientists from the Centers for Disease Control and Prevention (CDC) have concluded that Zika virus is a cause of Microcephaly and other severe fetal brain defects [10]. Links to other neurological complications are also being investigated by the World Health Organization [77] and other health agencies.

Figure 1–1: How the Zika virus spread around the world. The image was obtained from WHO and Lancaster University. [79]

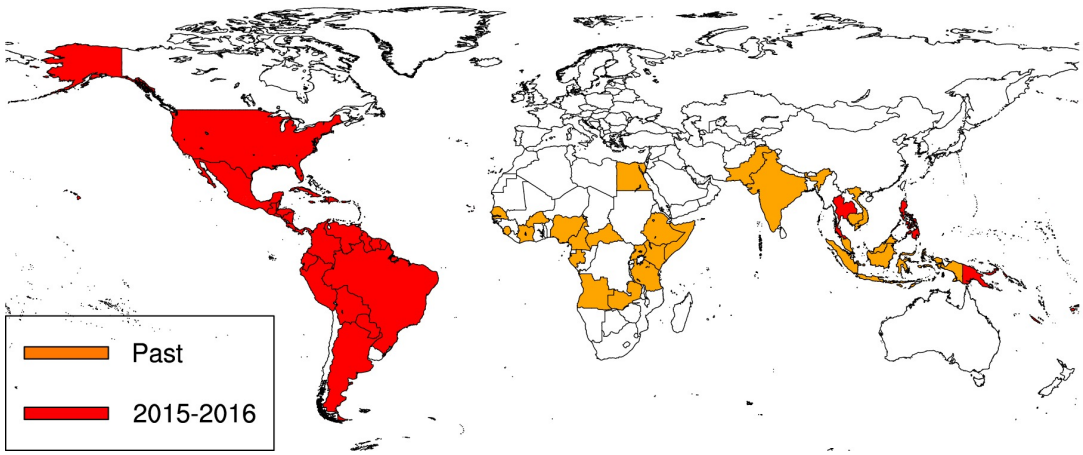


Figure 1–2: Zika: Affected areas before 2015 (orange) and affected areas between 2015 and 2016 (red). The Zika outbreak in Puerto Rico is marked in red, because it happened during 2015-2016. The image was obtained from the Diagram Collection: World Map Of Zika Cases. [73]

Figure 1–3 provides an illustration of the transmission cycle for the Zika infection, where vector transmission is from vector to human and from human to vector; and the sexual transmission is from human to human.

The main objective of this research is to estimate the basic reproductive number (R_0) of the Zika virus outbreak in Puerto Rico from November 2015 to December 2016 using weekly incidence data. The basic reproductive number is defined by Anderson and May [2] as the average number of secondary infections that results when

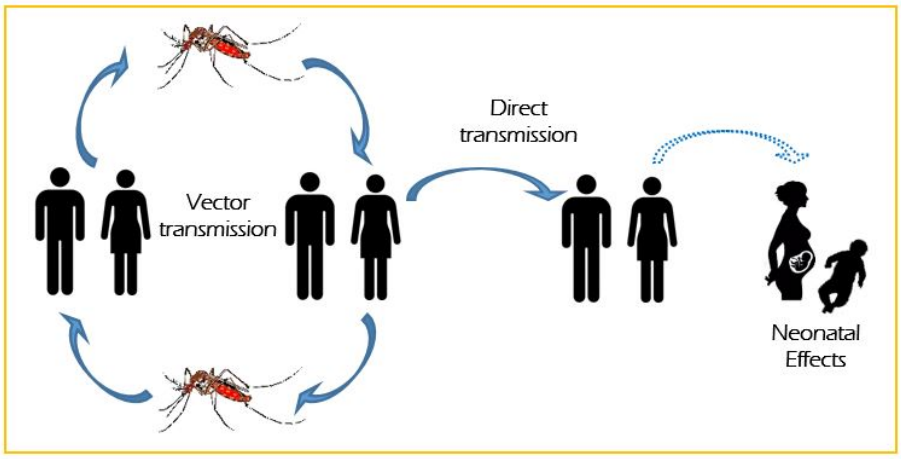


Figure 1–3: The infection cycle and effects.

a single infectious individual is introduced into an entirely susceptible population. Since vector as well as sexual transmission of the Zika virus will be considered, R_0 might help to determine the amount of effort needed to avoid or eliminate a Zika epidemic.

In order to achieve this goal, a mathematical model describing the transmission dynamics of Zika need to be build or adopt. Since the Zika virus can be transmitted through vectors as well as directly from human to human, a mathematical model that includes this two mechanism of transmission will be used. With a mathematical model of ordinary differential equations and a statistical model to account for the number of Zika cases per week (weekly incidence), the following contributions will be attained:

1. An estimate of the initial exponential growth rate (ρ) and the basic reproductive number (R_0), with different statistical methods. So far we understand that this would be the first estimate of R_0 for the Zika 2015-2016 epidemic in Puerto Rico, as the Zika virus is an emergent disease on the island. A posterior distribution for ρ and R_0 will also be provided.
2. An estimate of the reporting rate (r), due to asymptomatic cases and individuals that do not seek medical care, using a Bayesian approach.

3. The Ordinary Least Squares (OLS) and Generalized Least Squares (GLS) methods were considered to minimize the point-by-point distances between the predicted data by the mathematical model and the observed data for the Zika epidemic in Puerto Rico. The optimization procedure was performed to estimate the transmission rates β_h and κ (that are unknown) in order to estimate and generate a distribution for R_0, using the parameters from the model through a sampling process.

1.1 Main objective

Estimate the basic reproductive number (R_0) of the Zika virus outbreak in Puerto Rico from November 2015 to December 2016 using weekly incidence data.

1.2 Specific objectives

To achieve the main goal of this work, we established some specifics objectives:

1. Identify an epidemiological model of ordinary differential equations that represent the Zika dynamics in Puerto Rico, to determine an adequate expression of the basic reproductive number (R_0), with and without the exponential growth rate.
2. Obtain (by approximation) the dataset of the weekly incidence from the public report provided by the Puerto Rico Department of Health, by using a Plot Digitizer to extract the data.
3. Determine the exponential phase of the incidence curve of the Zika epidemic and estimate the initial exponential growth rate, ρ, by several methods, such as:
    - Visual method
        - Linear regression to the logarithm of the cumulative incidence
        - Bayesian approach
    - Favier et al.'s method (see [27])
    - Chowell et al.'s method (see [15])

4. Estimate the reporting rate of the Zika cases in Puerto Rico due to the asymptomatic cases and due to the people that do not seek medical assistance.
5. Estimate the transmission rates β_h and κ, in the mathematical model, that provide the best fitted curve to the observed data, through an optimization process.
6. Establish a comparison of our estimates of the initial exponential growth rate (ρ), basic reproductive number (R_0) and reporting rate (r) from Puerto Rico Zika outbreak with other results presented in the literature for Zika epidemics around the world.

CHAPTER 2
MATHEMATICAL MODEL AND DERIVATION OF THE BASIC REPRODUCTIVE NUMBER

2.1 Mathematical models of infectious diseases

Through the years, from the Paleolithic era to the present, mathematical models has been used in order to explain several phenomena that occur in our universe. Most of the processes that occur in our environment are dynamical processes, which are time dependent or change depending on time.

A question of concern is: Which mathematical tools can be employed to study public health phenomena, such as infectious disease? Carl Boyer, on his book "A History of Mathematics" [6], indicated that "*modeling* is a tool that is responsible for representing or simulating real-life situations, using mathematical equations, to determine or predict behavior". A mathematical model allows us to simplify the study of a dynamic phenomenon, since most of the situations involve non-linear behavior, something that, without modeling, could be difficult to analyze.

According to Anderson and May (in epidemiological context) [2], mathematical models have been, and still are, a very important tool that helps us to understand this area. Moreover, is the principal tool in order to study the behavior of an epidemic. The goal of a mathematical model of an infectious disease is to identify the causes of the outbreaks, the spread and the behavior of the disease, and most importantly, how to control it.

In this work, a mathematical model consisting of a system of ordinary differential equations will be used to study the spread of the Zika virus in Puerto Rico,

that is by estimating the basic reproductive number of the outbreak from confirmed by laboratory weekly cases. The interest of this work is on the number of infected individuals, taking into account the mechanism of how the individual got the virus (vector vs. sexual transmission). Since the transmission of the Zika follows a process in order to produce the infection, the entire population needs to be separated into compartments that usually describe the infectious state. One of the first and famous model of vector-borne disease is the Ross-MacDonald model (1957) [2]. This model was employed to study malaria transmission in Africa. The model is given by the following system:

$$\begin{aligned} Y' &= abI\frac{H-Y}{H} - \epsilon Y, \\ I' &= ac(V-I)\frac{Y}{V} - \delta I \end{aligned} \tag{2.1}$$

where the total mosquito (V) and human population sizes (H) are constant. We have that mosquitoes can be susceptible or infectious (I), humans are either susceptible or infectious (Y), there are no incubation periods, a is the mosquito biting rate, b is the mosquito to human transmission probability per bite, c is the human to mosquito transmission probability per bite , ϵ is the human recovery rate and δ is the mosquito death rate.

As indicated early, the main objective of this work is to estimate the basic reproductive number, R_0, for the 2015-2016 Zika outbreak in Puerto Rico, but we need to learn details about this epidemiological measure, for example, how it is defined, how to calculate it, what kind of data is needed, and more important, what can be concluded from it. The basic reproductive number of a infectious disease, denoted by R_0, is the average number of secondary infectious that result if a single infectious individual is introduced into an entirely susceptible population [2, 7, 12, 13, 23, 25, 27, 30, 39, 40, 64, 65, 67, 72, 74].

Ross (1911) defined R_0, for System 2.1, as the product of the basic reproductive number of humans infected by a mosquito and the basic reproductive number of mosquitoes infected from a person, that is,

$$R_0 = R_0^{HV} R_0^{VH}. \tag{2.2}$$

Typically, if $R_0 < 1$ every infected individual can produce on average less than one new infected individual, and it is possible to predict that the infection will disappear from the population. On the contrary, if $R_0 > 1$, then the infection is able to invade the susceptible population and the disease can persist and increase. The analysis of this threshold is important and useful aspect in studying a disease because it allows to decide which control action (how and when to apply them) would be most effective in reducing R_0 below one [2] [7].

2.2 Epidemics modeling of vector borne infectious diseases

Between 2013 and 2014, more than 30,000 cases of Zika virus disease were estimated in French Polynesia. This outbreak is known as the second largest outbreak on worldwide history, while the 2007 Zika outbreak in Yap Island is known as the largest because 73% of the total population got infected [77]. Kucharski et al. [40] use a mathematical model of vector-borne infections to examine the transmission dynamics on the six archipelagos of French Polynesia. In Kucharski's model, both people and mosquitoes were modeled using a susceptible-exposed-infectious-removed (SEIR) framework. This model incorporated delays as a result of the intrinsic (human) and extrinsic (vector) incubation periods.

The complete model is given by:

$$
\begin{aligned}
dS^H/dt &= -\beta_H S^H I^V, \\
dE^H/dt &= \beta_H S^H I^V - \alpha_H E^H, \\
dI^H/dt &= \alpha_H E^H - \gamma I^H, \\
dR^H/dt &= \gamma I^H, \\
dC/dt &= \alpha_H E^H, \\
dS^V/dt &= \delta - \beta_V S^V \frac{I^H}{N} - \delta S^V, \\
dE^V/dt &= \beta_V S^V \frac{I^H}{N} - (\delta + \alpha_V) E^V, \\
&\text{and} \\
dI^V/dt &= \alpha_V E^V - \delta I^V.
\end{aligned}
\tag{2.3}
$$

In the model, S^H represents the number of susceptible people, E^H is the number of people currently in their incubation period, I^H is the number of infectious people, R^H is the number of people that have recovered, C denotes the cumulative number of people infected, and N is the human population size, all at time t. Similarly, S^V represents the proportion of mosquitoes currently susceptible, E^V the proportion in their incubation period, and I^V the proportion of mosquitoes currently infectious, all of them at time t. Since the mean human lifespan is much longer than the outbreak duration, they omitted human births and deaths. Through the mathematical analysis, Kucharski et al. [40] derived the basic reproductive number for the Zika epidemic, defined as the product of the average number of mosquitoes infected by the typical infectious human, and vice versa, or,

$$
R_0 = \frac{\beta_V}{\gamma} \cdot \frac{\alpha_V}{\delta + \alpha_V} \frac{\beta_H}{\delta},
\tag{2.4}
$$

similar to the R_0 in Ross model.

This model only includes one mechanism of transmission, and is because by the date of publication, the sexual transmission was not yet discovered or considered for Zika virus outbreaks. Even though the model does not include the sexual transmission from human to human, the analysis of the basic reproductive number and the statistical process provided information about Zika dynamics. The median estimates of R_0 ranged between 2.6 and 4.8. Kucharski et al. [40] also studied the problem of the under-reporting number of cases caused by the asymptomatic individuals (mostly), and in addition, showed a demographic model that will provide the potential of the Zika virus to cause future outbreaks.

Gao et al. [30] studied a mathematical model for Zika virus transmission that included mosquito-borne and sexual transmission as the two mechanism of infection. The difference between Gao's and Kucharski's model are the mechanisms of transmission and the stratification of the human population. As we specified before, the Zika virus can be asymptomatic and because of this, is very uncertain to identify the exact number of infected individuals. Gao et al. [30] presented a model in which the human population is divided into six classes: susceptible ($S_h(t)$), exposed ($E_h(t)$), symptomatically infected ($I_{h1}(t)$), convalescent ($I_{h2}(t)$), asymptomatically infected ($A_h(t)$), and recovered ($R_h(t)$), all at times $t > 0$. In Puerto Rico, we know there is a problem with the reporting rate because of the asymptomatic cases. Even when Gao et al. [30] estimated $R_0 = 2.055$ (95% CI: 0.52-6.30) for the Zika epidemic in Brazil, Colombia and El Salvador (together), by taking into account the asymptomatic and symptomatic individuals as two classes of infected inndividuals, they did not provided information on how they obtained the data about the asymptomatic cases neither if they had the numbers of asymptomatic cases or if they were assumed. For this reason, in our work we will not analyzed a model similar to the one proposed by Gao et al. [30], because we cannot identified the weekly number of infected individuals that were asymptomatic in the epidemic.

Favier et al. [27], constructed a mathematical model for dengue fever disease (a vector-borne disease) of the type SI for the human and mosquito population. Their work rely on a statistical analysis and different methods to calculate the basic reproductive number, by using early epidemic curves for vector-borne diseases. Their emphasis was on the estimation of the initial exponential growth rate of the epidemic. As they specified on the article, the methods were: (1) a direct estimation of R_0 from the definition: the evaluation of the number of hosts potentially infected by one single infectious case, (2) the correspondence between R_0 and the final prevalence (better known as the final size relation), (3) transmission chains, and (4) an estimate of R_0 from the slope of the initial exponential growth of the cumulative number of cases, called the force of infection. The work on this thesis will consider the last method provided by Favier et al. [27], which used the initial exponential growth of the epidemic (ρ) as a way to estimate the basic reproductive number (R_0). This consideration is very important in our work because it has the advantage of relaying on the available data, which in our case is the weekly incidence and weekly cumulative incidence of the 2015-2016 Zika epidemic in Puerto Rico (see next chapter for more details) [27].

We also reviewed the work of Brauer et al. [7], where models with and without sexual transmission for the study of the dynamics of Zika virus were proposed. The work of Brauer et al. [7] is divided into three parts which we discuss in the next section.

2.3 Expression of R_0 using parameters in the model

From Chapter 1, we learned that the Zika virus has two mechanism of transmission, vectorial throught mosquitoes and directly throught sexual contacts. In order to select an appropriate mathematical model to estimate the basic reproductive number R_0 for the ZIKV epidemic in Puerto Rico, we need to adopt a model that includes both mechanism of transmission. In the next subsections, we will discuss in

details the work of Brauer et al. [7], the work that was the first mathematical model that described Zika virus dynamics, including vector as well as sexual transmission of the virus.

2.3.1 Vector transmission only

For a vector-borne disease, the model explained by Brauer et al. includes compartments corresponding to susceptible, exposed, infected and recovered humans, and susceptible, exposed, and infected mosquitoes, known as the SEIR/SEI model. By definition, the transmission rates, β_h and β_v, is the product of the number of mosquitoes bites per unit of time, times the probability of infection given the contact (bite) between the human and the mosquito. Because the probability is unknown, this parameter is very difficult to estimate. Therefore, instead of using the transmission rates, an expression for the basic reproductive number that depend on the initial exponential growth rate, ρ , of the epidemic, will be used [7] [72]. In order to obtain an expression of R_0 in terms of ρ, Brauer et al. [7], used a linearization of the system around the disease-free equilibrium (DFE) and then solve a characteristic equation for the initial exponential growth rate ρ.

The dynamics of the epidemic considering only vector transmission starts in a population of $N_h(t)$ humans and $N_v(t)$ adult female mosquitoes, at time t, where the susceptible adult female mosquitoes at time t, $S_v(t)$, upon biting an infectious human, incubate the virus for an average period of time, $1/\alpha_v$, and then progresses to the infectious compartment at time t, $I_v(t)$. The mosquitoes die after an average of $1/\delta$ days. The transmission rate from an infectious human to a susceptible mosquito is β_v, which is the transmission rate given that the contact between the infectious human and the susceptible mosquito is effective.

On the other hand, the susceptible humans at time t, $S_h(t)$, can be infected by being bitten by an infectious female mosquitoes or through direct contact with

another infected human (for instance, by sexual contact). In this case, the human incubates the virus for an average period of time, $1/\alpha_h$ days, before becoming infectious at time t, $I_h(t)$. After an average of $1/\gamma$ days, the human then recovers progressing to the immune compartment at time t, $R_h(t)$. This dynamics is well explained in [7].

The model presented in [7] is described by the seventh coupled, nonlinear ordinary differential equations given in System 2.5. The population state variables and parameters are presented in Tables 2–1 and 2–2, respectively.

$$\begin{aligned} S_h' &= -\beta_h S_h \frac{I_v}{N_v}, \\ E_h' &= \beta_h S_h \frac{I_v}{N_v} - \alpha_h E_h, \\ I_h' &= \alpha_h E_h - \gamma I_h, \\ R_h' &= \gamma I_h, \\ S_v' &= -\beta_v S_v \frac{I_h}{N_h} + \delta N_v - \delta S_v, \\ E_v' &= \beta_v S_v \frac{I_h}{N_h} - (\delta + \alpha_v) E_v, \end{aligned} \tag{2.5}$$

and

$$I_v' = \alpha_v E_v - \delta I_v,$$

where $N_h = S_h + E_h + I_h + R_h$ and $N_v = S_v + E_v + I_v$ are the human and mosquito population sizes, respectively.

Population	**Definition**
$S_h(t)$	Number of susceptible humans
$E_h(t)$	Number of infected humans who are not yet infectious (exposed)
$I_h(t)$	Number of infectious humans
$R_h(t)$	Number of recovered humans
$S_v(t)$	Number of susceptible mosquitoes
$E_v(t)$	Number of infected mosquitoes who are not yet infectious (exposed)
$I_v(t)$	Number of infectious mosquitoes

Table 2–1: Description of the human and mosquito populations state variables, all at times t.

Parameter	Definition	Units
$1/\alpha_v$	Extrinsic incubation period (mosquitoes)	Time
$1/\alpha_h$	Intrinsic incubation period (humans)	Time
$1/\gamma$	Human infectious period	Time
$1/\delta$	Mosquito lifespan	Time
β_h	Vector to Human transmission rate	Dimensionless
β_v	Human to Vector transmission rate	Dimensionless

Table 2–2: Description of the parameters in the model.

The basic reproductive number was heuristically obtained from the equations. As Brauer et al. explained in [7], there are two stages of the infection process. First, the infected human infects mosquitoes, at a rate $\beta_v \frac{N_h}{N_v}$ over an average time $1/\gamma$. This produces $\beta_v \frac{N_h}{N_v \gamma}$ infected mosquitoes, of whom a fraction $\frac{\alpha_v}{\alpha_v+\delta}$ proceed to become infectious. The second stage is that the infected mosquitoes infect humans at a rate $\beta_h \frac{N_v}{N_h}$ for an average time $1/\delta$, producing $\beta_h \frac{N_v}{N_h \delta}$ infected humans per mosquito. The net result of these two stages is

$$\left(\beta_v \frac{N_h}{N_v \gamma}\right)\left(\frac{\alpha_v}{\alpha_v+\delta}\right)\left(\beta_h \frac{N_v}{N_h \delta}\right) = \beta_v \beta_h \frac{\alpha_v}{\delta\gamma(\alpha_v+\delta)}$$

infected humans. Therefore,

$$R_0 = \left(\frac{\beta_v}{\gamma}\right)\left(\frac{\alpha_v}{\alpha_v+\delta}\right)\left(\frac{\beta_h}{\delta}\right). \tag{2.6}$$

By using the Next Generation Matrix process, discussed in [74], the same expression for R_0 can also be obtained.

2.3.2 Sexual transmission only

The model of an infectious disease with sexual transmission only can be defined as

$$\begin{aligned} S_h' &= -\kappa S_h \frac{I_h}{N_h}, \\ E_h' &= \kappa S_h \frac{I_h}{N_h} - \alpha_h E_h, \\ I_h' &= \alpha_h E_h - \gamma I_h, \end{aligned} \tag{2.7}$$

and

$$R_h' = \gamma I_h,$$

where $N_h = S_h + E_h + I_h + R_h$ is the human population size.

If there is sexual transmission, the basic reproductive number is independent of the host-vector interaction [7]. In this case, the interaction produces κ cases in an average time of $1/\gamma$ days. Therefore,

$$R_0 = \frac{\kappa}{\gamma}. \tag{2.8}$$

2.3.3 Vector and sexual transmission

The model in this subsection will be similar to the model discussed in the first subsection, System 2.5. Brauer et al. [7] add the term $\kappa S_h \frac{I_h}{N_h}$ into the second and third differential equation, where κ is the transmission rate from human to human. Figure 2–1 shows a flowchart for the mathematical model with sexual and vector transmission. Brauer et al. presented this model (System 2.9 below) as the first mathematical model to study the dynamics of a Zika virus epidemic. The population state variables remains as stated, but the table of parameters includes now the sexual transmission rate, from human to human (see Table 2–3).

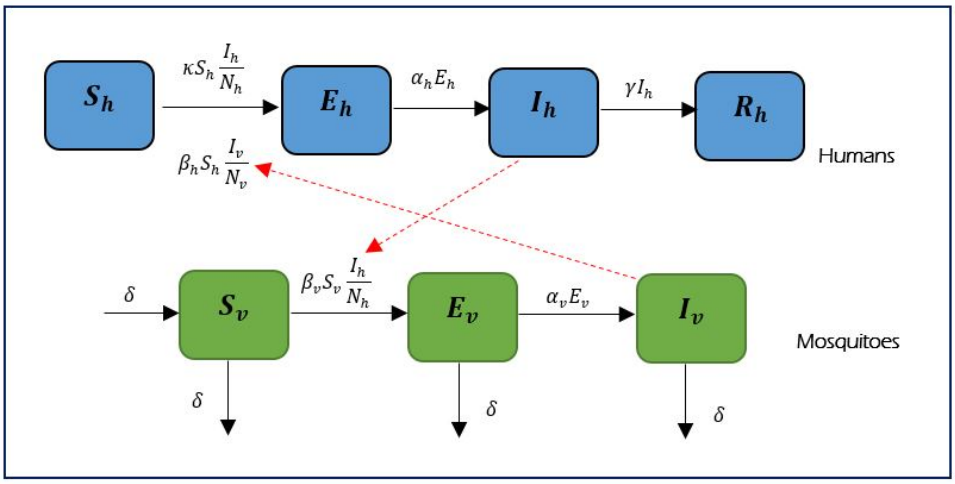


Figure 2–1: Compartmental flowchart of the vector and sexual transmission model of Brauer et al. [7]. Red lines indicate the transmission terms between an infected moquito (I_v) with a susceptible human (S_h) or between an infected human (I_h) with a susceptible mosquito (S_v).

$$
\begin{aligned}
S_h' &= -\beta_h S_h \frac{I_v}{N_v} - \kappa S_h \frac{I_h}{N_h}, \\
E_h' &= \beta_h S_h \frac{I_v}{N_v} + \kappa S_h \frac{I_h}{N_h} - \alpha_h E_h, \\
I_h' &= \alpha_h E_h - \gamma I_h, \\
R_h' &= \gamma I_h, \\
S_v' &= -\beta_v S_v \frac{I_h}{N_h} + \delta N_v - \delta S_v, \\
E_v' &= \beta_v S_v \frac{I_h}{N_h} - (\delta + \alpha_v) E_v,
\end{aligned}
\tag{2.9}
$$

and

$$
I_v' = \alpha_v E_v - \delta I_v,
$$

where $N_h = S_h + E_h + I_h + R_h$ and $N_v = S_v + E_v + I_v$ are the human and mosquito population sizes, respectively.

Due to the consideration of the vector and sexual transmission, there are three transmission rates that are difficult to estimate. Brauer et al. [7] performed the same

Parameter	Definition	Units
$1/\alpha_v$	Extrinsic incubation period (mosquitoes)	Time
$1/\alpha_h$	Intrinsic incubation period (humans)	Time
$1/\gamma$	Human infectious period	Time
$1/\delta$	Mosquito lifespan	Time
β_h	Vector to Human transmission rate	Dimensionless
β_v	Human to Vector transmission rate	Dimensionless
κ	Human to Human transmission rate	Dimensionless

Table 2–3: Description of the parameters in the Zika model. Since our data were extracted from a weekly report, we take time in weeks.

analysis as before, a linearization of the system around the disease-free equilibrium (DFE).

For the model in System 2.9, Brauer et al. [7] determined that the basic reproductive number is given by

$$\begin{aligned} R_0 &= R0_{vector} + R0_{sex} \\ &= \beta_v \beta_h \frac{\alpha_v}{\delta\gamma(\alpha_v + \delta)} + \frac{\kappa}{\gamma}, \end{aligned} \tag{2.10}$$

where it is resulted from the Next Generation Matrix, proposed by Van den Driessche and Watmough in [74].

2.4 Expression of R_0 using the initial exponential growth rate (ρ)

In most cases, some parameters from the mathematical model are unknown, leading to take a different approach to compute R_0. In this case, the notion of an exponential growth will be consider in order to encounter an expression for the basic reproductive number that does not depend of the unknown parameters, which in this case are the transmission rates of the epidemic. This section will explained the linearization of a mathematical model to compute R_0.

2.4.1 Vector transmission only

Using the mathematical model in System 2.5 (presented in [7]), a linearization procedure around the disease-free equilibrium (DFE) will be employ, where $S_h = N_h$, $E_h = I_h = 0$, $S_v = N_v$ and $E_v = I_v = 0$ will be explained. For the process, Brauer et al. [7] assume that $y = N_h - S_h$ and $z = N_v - S_v$, in order to obtain the linearization.

$$\begin{aligned} y = N_h - S_h &\longrightarrow S_h = N_h - y \\ &\longrightarrow S_h' = -y' \\ &\longrightarrow y' = -S_h' \end{aligned}$$

and

$$\begin{aligned} z = N_v - S_v &\longrightarrow S_v = N_v - z \\ &\longrightarrow S_v' = -z' \\ &\longrightarrow z' = -S_v' \end{aligned}$$

Therefore, the System 2.5 can be written in the form calculated by Brauer et al. on their article (see System 2.11). Notice that the equation of the evolution of the recovered humans, $R_h(t)$, decouples from the system, as it is determined by computing the population value of the infected human equation [67].

$$\begin{aligned} y' &= \beta_h N_h \frac{I_v}{N_v}, \\ E_h' &= \beta_h S_h \frac{I_v}{N_v} - \alpha_h E_h, \\ I_h' &= \alpha_h E_h - \gamma I_h, \\ z' &= \beta_v N_v \frac{I_h}{N_h} - \delta z, \\ E_v' &= \beta_v S_v \frac{I_h}{N_h} - (\delta + \alpha_v) E_v, \end{aligned} \tag{2.11}$$

and

$$I_v' = \alpha_v E_v - \delta I_v.$$

The characteristic equation of the System 2.11 is

$$det\begin{bmatrix} \lambda & 0 & 0 & 0 & 0 & \beta_h\frac{N_h}{N_v} \\ 0 & -(\lambda+\alpha_h) & 0 & 0 & 0 & \beta_h\frac{N_h}{N_v} \\ 0 & \alpha_h & -(\lambda+\gamma) & 0 & 0 & 0 \\ 0 & 0 & \beta_v\frac{N_v}{N_h} & -(\lambda+\delta) & 0 & 0 \\ 0 & 0 & \beta_v\frac{N_v}{N_h} & 0 & -(\lambda+\delta+\alpha_v) & 0 \\ 0 & 0 & 0 & 0 & \alpha_v & -(\lambda+\delta) \end{bmatrix} = 0$$

Then, reducing the characteristic equation Columns 1 and 4,

$$\lambda(\lambda+\delta)\cdot det\begin{bmatrix} -(\lambda+\alpha_h) & 0 & 0 & \beta_h\frac{N_h}{N_v} \\ \alpha_h & -(\lambda+\gamma) & 0 & 0 \\ 0 & \beta_v\frac{N_v}{N_h} & -(\lambda+\delta+\alpha_v) & 0 \\ 0 & 0 & \alpha_v & -(\lambda+\delta) \end{bmatrix} = 0$$

The initial exponential growth rate, ρ, is the largest root of the following fourth degree polynomial (see details in [7])

$$g(\lambda) = (\lambda+\alpha_h)(\lambda+\gamma)(\lambda+\delta+\alpha_v)(\lambda+\delta) - \beta_h\beta_v\alpha_h\alpha_v. \tag{2.12}$$

From Brauer et al. [7], Equation 2.12 has a unique positive root when $g(\lambda) = 0$ and this is the initial exponential growth rate. The initial exponential growth rate may be measured experimentally, and if the measured value is ρ, then from Equation 2.12 it is obtained:

$$(\lambda+\alpha_h)(\lambda+\gamma)(\lambda+\delta+\alpha_v)(\lambda+\delta) = \beta_h\beta_v\alpha_h\alpha_v = R_0\alpha_h\delta\gamma(\delta+\alpha_v). \tag{2.13}$$

Therefore,

$$\hat{R}_0 = \frac{(\rho+\alpha_h)(\rho+\gamma)(\rho+\delta)(\rho+\delta+\alpha_v)}{\gamma\alpha_h\delta(\delta+\alpha_v)}. \tag{2.14}$$

Notice that $\hat{R}_0$ does not depend on β_h and β_v, the transmission rates.

2.4.2 Sexual transmission only

By using the same technique, a linearization around the disease-free equilibrium (DFE) will be perform, where $S_h = N_h$, $E_h = I_h = 0$, and taking $y = N_h - S_h$. With this information, the System 2.7 can be written as a system of three differential equations, that is

$$\begin{aligned} y' &= \kappa I_h, \\ E_h' &= \kappa I_h - \alpha_h E_h, \\ &\text{and} \\ I_h' &= \alpha_h E_h - \gamma I_h. \end{aligned} \tag{2.15}$$

For this system, the characteristic equation is given by

$$det \begin{bmatrix} -\lambda & 0 & \kappa \\ 0 & -(\lambda + \alpha_h) & \kappa \\ 0 & \kappa & -(\lambda + \gamma) \end{bmatrix} = 0.$$

Reducing the equation, we have then

$$\lambda \cdot det \begin{bmatrix} -(\lambda + \alpha_h) & \kappa \\ \kappa & -(\lambda + \gamma) \end{bmatrix} = 0.$$

Therefore, by the same assumption about the initial exponential growth rate,

$$\rho\alpha_h\kappa - \rho(\rho + \alpha_h)(\rho + \gamma) = 0. \tag{2.16}$$

Leaving the parameter κ in one side of the equation and dividing by γ, we obtain that

$$\frac{\kappa}{\gamma} = \frac{(\rho + \alpha_h)(\rho + \gamma)}{\alpha_h\gamma}. \tag{2.17}$$

By using the Equation 2.8, we finally obtain that the basic reproductive number, for a model with only sexual transmission, in term of ρ, is

$$\hat{R}_0 = \frac{(\rho + \alpha_h)(\rho + \gamma)}{\alpha_h \gamma}. \tag{2.18}$$

2.4.3 Vector and sexual transmission

After the linearization of the System 2.9 around the disease-free equilibrium (DFE), Brauer et al. obtained the following equation:

$$(\lambda + \alpha_h)(\lambda + \gamma)(\lambda + \delta)(\lambda + \delta + \alpha_v) - \kappa\alpha_h(\lambda + \delta)(\lambda + \delta + \alpha_v) - \beta_h\beta_v\alpha_h\alpha_v = 0 \tag{2.19}$$

Since the initial exponential growth rate, ρ, is the largest root of the previous equation [7],

$$(\rho + \alpha_h)(\rho + \gamma)(\rho + \delta)(\rho + \delta + \alpha_v) - \kappa\alpha_h(\rho + \delta)(\rho + \delta + \alpha_v) - \beta_h\beta_v\alpha_h\alpha_v = 0 \tag{2.20}$$

Equation 2.10 can be written as

$$\beta_h\beta_v\alpha_h\alpha_v = \delta\gamma\alpha_h(\delta + \alpha_v)R_0 - \kappa\delta\alpha_h(\delta + \alpha_v) \tag{2.21}$$

Finally, by replacing Equation 2.21 into Equation 2.20, the following expression for the estimation of R_0 is obtained:

$$\hat{R}_0 = R0_{sex} + \left[\frac{(\rho + \alpha_h)(\rho + \gamma)}{\gamma\alpha_h} - R0_{sex}\right]\frac{(\rho + \delta)(\rho + \delta + \alpha_v)}{\delta(\delta + \alpha_v)}. \tag{2.22}$$

All the expressions for the basic reproductive number (2.14, 2.18, and 2.22) makes sense because if it is assume, for instance, that $\rho = 0$, all of them will be equal to 1, which is the appropriate threshold behavior. In [72], Towers et al. used Equation 2.22, to estimate R_0 for the Zika epidemic occurred in Barranquilla, Colombia. For this outbreak, they estimated R_0 to be 3.8 [2.4, 5.6] with $\rho = 0.076$ $days^{-1}$ (or 0.532 $weeks^{-1}$).

In order to study the 2015-2016 Zika virus outbreak in Puerto Rico, we adopt the mathematical model defined by Brauer et al. [7], which is the System 2.9. This model is the only one we found that includes the two types of transmission (vector and sexual), that incorporates the initial exponential growth rate (ρ) into the expression for the basic reproductive number (R_0). To estimate R_0, the expression obtained by Towers et al. [72] is also used (Equation 2.22). It is important to understand that Equation 2.22 requires the estimation of the initial exponential growth rate (ρ), or the force of infection as defined by Favier et al. [27].

CHAPTER 3
STATISTICAL METHODS

3.1 Derivation of the initial exponential growth rate (ρ)

The expression for the basic reproductive number, Equation 2.22, requires the estimation of the initial exponential growth rate (ρ), or the force of infection as defined by Favier et al. [27]. Questions that resulted in the process of this research were: Is the initial exponential growth rate equivalent to the force of infection? and if this relationship holds, how can the force of infection be defined?

When an epidemic occurs (typically if $R_0 > 1$), the observed data from the epidemic, in particular, the increase in the incidence, have an exponential form, during its initial phase. This suggest that we can model the initial phase of an epidemic as

$$y(t) = y_0 e^{\rho t}, \tag{3.1}$$

where ρ is the initial exponential growth rate and $y(0) = y_0$ is the initial condition of the epidemic, which means, that y_0 is the initial number of infected individuals (incidence at time zero). If we think, for example, in the ZIKV epidemic, assuming that X is a continuous random variable representing the time of exposure of an individual in the population, with probability density function (p.d.f.), $f(t)$, and cumulative distribution function (c.d.f.),

$$F(t) = P(X < t), \tag{3.2}$$

giving the probability of acquiring the Zika virus at time t, then

$$S(t) = P(X \geq t) = 1 - F(t) = \int_t^{\infty} f(x)dx, \tag{3.3}$$

gives the probability of being susceptible to the ZIKV at time t [62]. In this case, $S(t)$ is called the survival function. We can describe the random variable X by using the hazard function, that is

$$\rho(t) = \lim_{dt \to 0} \frac{Prob(t \leq X \leq t + dt \mid X \geq t)}{dt}. \tag{3.4}$$

By following the procedure in [34, 62], Equation 3.4 simplifies to

$$\rho(t) = \frac{f(t)}{S(t)}. \tag{3.5}$$

Integrating from 0 to t, and assuming that all the people has survived at time zero, $(S(0) = 1)$, we have that

$$S(t) = \exp\left(-\int_0^t \rho(x)dx\right). \tag{3.6}$$

At the beginning of the exponential behaviour, a constant risk of infection over time can be assumed [2], then

$$\begin{aligned} \int_0^t \rho(x)dx &= \int_0^t \rho \cdot dx \\ &= [\rho x]_0^t \\ &= \rho[t - 0] \\ &= \rho t. \end{aligned} \tag{3.7}$$

Therefore,

$$S(t) = \exp(-\rho t), \tag{3.8}$$

where ρ is called the force of infection, or infection hazard. In [34], if t is the time of exposure of the population, the expected proportion of the initial population that

still susceptible at time t, will have decayed away exponentially according to Equation 3.8. In this case, $\exp(-\rho t)$ can be define as the probability that a susceptible individual becomes infected. This background of the survival function gives us the reason to state that the force of infection is equivalent to the initial exponential growth rate, because the force of infection shows similar characteristics, that we can identify from Equation 3.1. With this thought in mind, we can now proceed to discuss the different statistical methods we have considered in order to identify the exponential phase and how to estimate the initial exponential growth.

3.2 Determining the exponential phase and estimate of ρ

Before estimating ρ, we need to identify what is called the exponential phase. The exponential phase is a J shaped growth curve that represent the increase of the number of reported cases over time, until the maximum number of cases is observed. This phase will end before the curve start rapidly decreasing. Our work will consider several methods to identify the exponential phase of the 2015-2016 Zika virus epidemic in Puerto Rico.

3.2.1 Visual method (VM) for exponential phase

In general, if a incidence curve of an epidemic is considered (see Figure 3–1 for instance), an exponential form of this curve can be noticed. To determine the exponential phase of the epidemic, the maximum number of cases need to be identified, which in this case correspond to the time 400, where 1200 cases were observed. Therefore, the exponential phase will be the region from time 0 to 400, marked with a double arrow (purple) in Figure 3–2, while the exponential growth is the green arrow. The orange line represents what it is known, in most cases, by t_{peak}, which is the time where the maximum number of cases is observed. To estimate the initial exponential growth rate, two statistical methods will be used. First, a linear regression to the logarithm of the cumulative incidence curve will be

employ considering exponential phase. Second, a Bayesian approach will be perform taking into consideration the exponential phase and the distribution associated to the observed data.

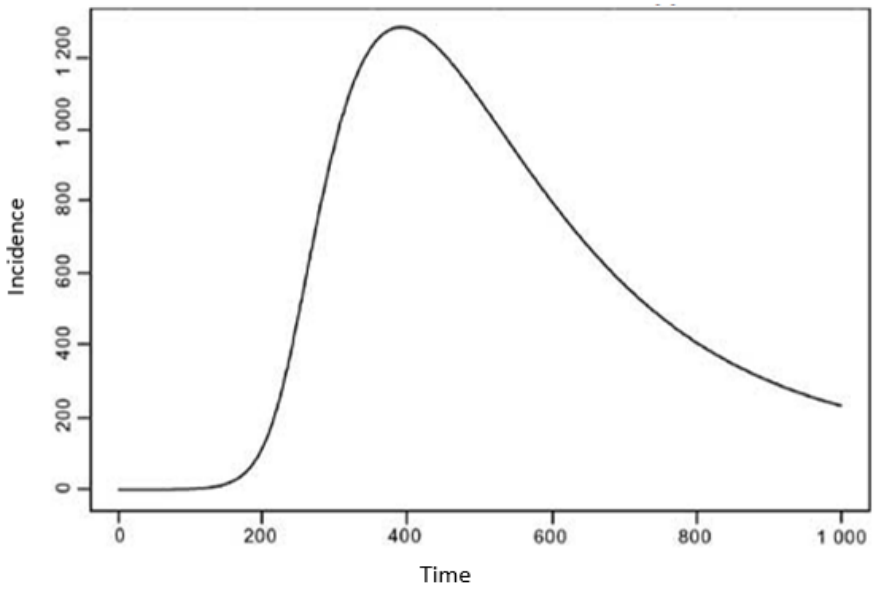


Figure 3–1: Example of a incidence curve of an epidemic.

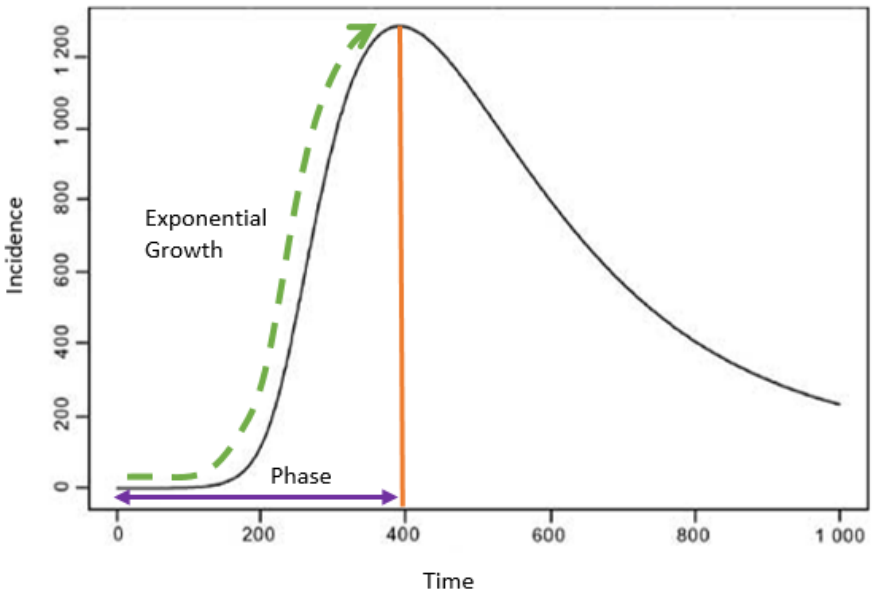


Figure 3–2: Example of a incidence curve of an epidemic with the exponential growth and phase.

3.2.2 Favier et al.'s method

The work of Favier et al. [27] evaluated a method of deriving the basic reproductive number (R_0) for vector-borne diseases from the early epidemic curves, in particular, for vector-borne diseases with incubations in the vectors and in the hosts. A statistical model will be applied to several dengue epidemics in different climatic regions of Brazil. In order to estimate R_0, the force of infection (initial exponential growth rate) was linked to the expression of R_0.

According to Favier et al., the first step was estimating the force of infection. The method suggest that the mean number of new cases by unit of time (the day or the week for instance) needs to be plotted against the cumulative number of cases. The phase of exponential growth of the cumulative number of cases will be evidenced by a linear growth of the curve and the slope correspond to the force of infection. An estimation of the force of infection is computed by a least-square linear fit of this linear phase. For this, they determined when the exponential phase ends, that is when the curvature of the curve overtakes the stochastic noise around the initial linear trend.

To obtain the exponential phase and the estimation of the force of infection, Favier et al. proceeds with the following steps in accordance with Figure 3–3:

1. The number of cases declared daily is plotted against the total number of hosts ever infected.
2. Linear fits are performed, each time using one more data point.
3. The evolution of the goodness-of-fit of these linear regressions and of the slope of the fitted line will be plotted simultaneously. The slope will be the estimation of the force of infection. In the beginning of process, the slopes fluctuates greatly. In an intermediate phase, it oscillates around a stationary value, and finally, it decreases slowly. The best estimation of the force of infection lies in the intermediate phase and the cut-off of the initial linear phase should lie inside of that phase (or region).

4. They warned that choosing the minimal goodness-of-fit as a cut-off is not satisfactory, as it lies in the third phase and therefore leads to an underestimation of the force of infection.
5. The best estimation of the force of infection will lie in the second phase. To obtain the estimate of the force of infection, the 80th percentile of the slopes left to the minimum goodness-of-fit.

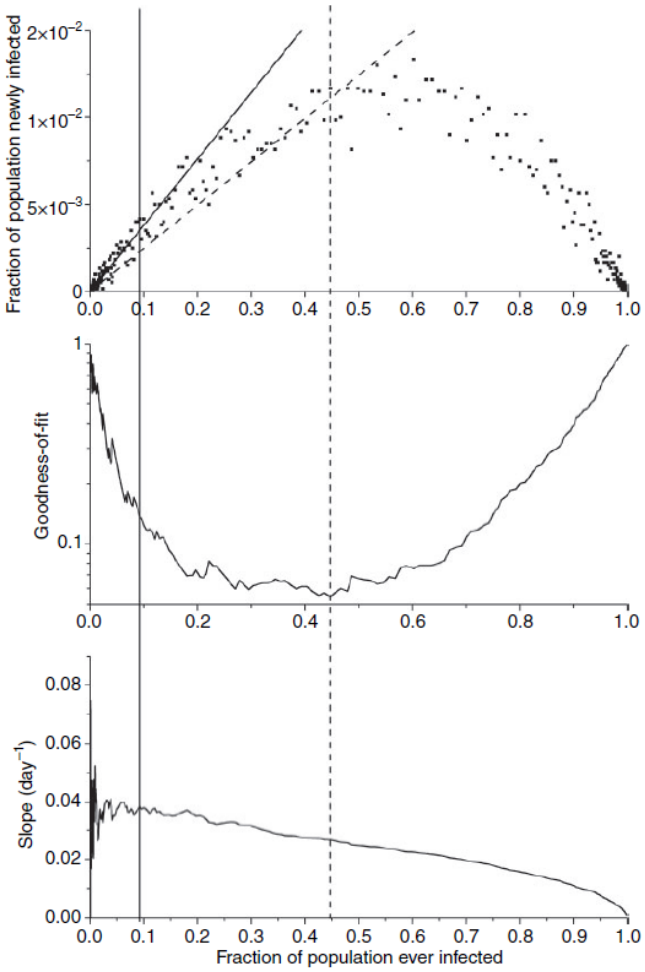


Figure 3–3: Favier et al.'s methodology. Figure obtained from [27]

This statistical method will be use to obtain the exponential phase and the estimation of the initial exponential growth rate (ρ) for the 2015-2016 Zika epidemic in Puerto Rico.

3.2.3 Chowell et al.'s method

The third method we considered to determine the exponential phase and the estimation of ρ is proposed by Chowell et al., in [15]. They analyzed the spread pattern of the 2009 A/H1N1 pandemic across 15 regions of Chile based on daily hospitalizations. They also estimated the reproduction number based on the growth rate of the exponential pandemic phase. We will use the procedure they performed in order to determine the initial exponential phase of the Zika outbreak in Puerto Rico.

In the early stages of an epidemic, when the effect of increasing incidence on the depletion of susceptibles is small, the growth of the epidemic is exponential in nature, with rate ρ [2, 27]. For the procedure, they assumed the classical SEIR (susceptible-exposed-infectious-recovered) transmission model, where the reproduction number (R_0) is determined from

$$R_0 = \left(1 + \frac{\rho}{\gamma}\right)\left(1 + \frac{\rho}{\kappa}\right), \tag{3.9}$$

where $1/\gamma$ and $1/\kappa$ are the latent and infectious periods, respectively.

In general, the standard deviation width of an epidemic curve consisting of N incidence measurements, y_j^{data}, at t_i different time points ($j = 1, \ldots, N$) is given by

$$\sigma_t = \sqrt{\frac{\sum_{j=1}^{N}(t_j - \bar{t})^2 y_j^{data}}{\sum_{j=1}^{N} y_j^{data}}}, \tag{3.10}$$

where

$$\bar{t} = \frac{\sum_{j=1}^{N} t_j y_j^{data}}{\sum_{j=1}^{N} y_j^{data}}. \tag{3.11}$$

Chowell et al. suggest that the exponential rise portion of a epidemic curve will be identified as the incidence data points at the beginning of the epidemic that are sufficiently many standard deviations away from the time of peak incidence (denoted by t_{peak}).

The exponential rise region is thus the region where

$$t_j < (t_{peak} - f \cdot \sigma_t). \tag{3.12}$$

In order to determine the optimal cut off value, denoted by f, Chowell et al. performed exponential rise fits to simulated data incidence curves from an SEIR (susceptible-exposed-infectious-recovered) model. Therefore, they choose the value of f that provides unbiased estimates of the true exponential rise. They found that $f = 1$. For further details about how the optimal value of f was obtained, see the suplementary material of reference [15].

The initial exponential growth rate, ρ, can be estimated from the exponential growth phase of the pandemic, using a Poisson Maximum Likelihood Method, which is explained in the supplementary document of [15]. Instead of using a Maximum Likelihood Method, a Bayesian approach will be used as indicated in [12–15, 40].

3.3 Inverse Problem Estimation

A mathematical model of nonlineal differential equations involves a vector of parameters and initial conditions that allow us to explain (or simulate) the phenomenon under study. If all the parameters and initial conditions associated with mathematical model are known, then the procedure would be to assign these values to the model and solve the system of differential equations in order to obtain the numerical solutions. However, in real situations, there are parameters associated to the model that are unknown. In this case, is necessary to estimate those parameters or the unknown initial conditions using an available data, leading to an inverse problem or to estimate the parameters in the model [4, 16]. In order to estimate the parameters, we need the following components: a mathematical model, a statistical model, a data set (observed data) and the optimization method [4].

In this section, the Ordinary Least Squares (OLS) and Generalized Least Squares (GLS) methods will be explained, which were methods implemented to perform the

optimization process of the parameters in the mathematical model. For an extensive discussion on linear and generealized linear models and least squares, see [1], [60], and [37]. For a discussion of inverse problems in epidemiological models see [4], [16], and [31].

3.3.1 Ordinary Least Squares (OLS) method

Let consider a system of nonlinear ordinary differential equations (in our case, System 2.9) given by

$$\frac{d\vec{z}}{dt} = \vec{g}(t, \vec{z}(t), \vec{\theta}), \tag{3.13}$$

where $\vec{\theta}$ is the vector of parameters to be estimated and the observational process is given by

$$\vec{y}(t) = C\vec{z}(t; \vec{\theta}). \tag{3.14}$$

A discrete process of the data can be assumed in which N longitudinal observations are obtained as follows

$$\vec{y}(t_j) = C\vec{z}(t_j; \vec{\theta}), \quad j = 1, \ldots, N, \tag{3.15}$$

where C is a $m \times n$ matrix that depends on the observed data. In the matrix, m represent the types of data and n represent the number of state variables in the model. For our case, $m = 1$, because we only have one type of data, which correspond to the weekly incidence for the ZIKV epidemic in Puerto Rico and the model proposed by Brauer et al. have $n = 7$ state variables.

In general, the data $\{y_j\}$ is not exactly $\vec{y}(t_j)$, $j = 1, \ldots, N$, because by the characteristic of the problem may produce errors between the observed data and the predicted data by the mathematical model.

The statistical model, for the case of the Ordinary Least Squares (OLS) method, is given by

$$\vec{Y_j} = \vec{f}(t_j; \vec{\theta_0}) + \vec{\varepsilon_j}, \quad j = 1, \ldots, N, \tag{3.16}$$

where $\vec{f}(t_j;\vec{\theta}) = C\vec{z}(t_j;\vec{\theta})$ for $j = 1,\ldots,N$, will represent the numerical solutions that can be computed from the mathematical model, at time t_j, under the unknown vector of parameters, $\vec{\theta_0}$. In the statistical model, the vector $\vec{\theta_0}$ is considered as the parameters that produced the observed data $\vec{Y_j}$ for $j = 1,\ldots,N$. In addition, the terms $\vec{\varepsilon_j}$ are independent random variables that represent the point-by-point distances between the predicted data from the mathematical model and the observed data. This distances can be defined as errors.

Since this point-by-point distances are unknown, it is assumed that ε_j can be generated from a probability distribution with mean $E(\vec{\varepsilon_j}) = 0$ and variance given by

$$V_0 = Var(\vec{\varepsilon_j}) = diag(\sigma_{0,1}^2, \sigma_{0,2}^2, \ldots, \sigma_{0,m}^2)_{m\times m},$$

where $\sigma_{0,i}^2$ are the unknown variances for each type of data, for $i = 1,\ldots,m$. It can be drawn that $\vec{Y_j}$ is a random variable with $E(\vec{Y_j}) = \vec{f}(t_j;\vec{\theta_0})$ and

$$Var(\vec{Y_j}) = diag(\sigma_{0,1}^2, \sigma_{0,2}^2, \ldots, \sigma_{0,m}^2)_{m\times m},$$

because $\vec{Y_j}$ depends on $\vec{\varepsilon_j}$, which is a random variable.

Given a set of observations $\vec{Y} = (\vec{Y_1}, \vec{Y_2}, \ldots, \vec{Y_N})$, the main idea of the OLS method, is to achieve good estimates by finding the minimizer $\vec{\theta}_{OLS}$ of the function

$$J_{OLS}(\vec{\theta}) = \sum_{j=1}^{N} \Big[\vec{Y_j} - \vec{f}(t_j;\vec{\theta})\Big]^T V_0^{-1} \Big[\vec{Y_j} - \vec{f}(t_j;\vec{\theta})\Big], \tag{3.17}$$

where $\vec{Y_j} - \vec{f}(t_j;\vec{\theta}) = \vec{\varepsilon_j}$ for $j = 1,\ldots,N$. By doing the appropriate matrix multiplication on the right side of Equation 3.17, the minimizer $\vec{\theta}_{OLS}$, is now given by

$$\vec{\theta}_{OLS}(\vec{Y}) = arg\,\min_{\vec{\theta}\in\Theta} \sum_{j=1}^{N} \Big[\sum_{i=1}^{m} \frac{1}{\sigma_{0,i}^2}\Big(Y_j(t_j) - f_i(t_j;\vec{\theta})\Big)^2\Big]. \tag{3.18}$$

If $\{\vec{y}_j\}_{j=1}^{N}$ is a the realization of the random process $\left\{\vec{Y}_j\right\}_{j=1}^{N}$, and solving for Equation 3.18, we can find the OLS estimate, $\hat{\theta}_{OLS}$, for $\vec{\theta}_{OLS}$, given by

$$\hat{\theta}_{OLS} = arg\,\min_{\vec{\theta}\in\Theta} \sum_{j=1}^{N} \Big[\sum_{i=1}^{m} \frac{1}{\sigma_{0,i}^2} \Big(y_i(t_j) - f_i(t_j;\vec{\theta}) \Big)^2 \Big], \tag{3.19}$$

By definition of the variance, we can write V_0 as

$$V_0 = diag\, E\Big(\frac{1}{N} \sum_{j=1}^{N} \Big[\vec{Y}_j - \vec{f}(t_j;\vec{\theta_0}) \Big] \cdot \Big[\vec{Y}_j - \vec{f}(t_j;\vec{\theta_0}) \Big]^T \Big)_{m\times m}. \tag{3.20}$$

For the estimation of $\hat{\theta}_{OLS}$, the matrix V_0 (which is unknown) is needed, and for the estimate of V_0, the parameter $\vec{\theta_0}$, which is also unknown, are necessary. In this case, for the computation of $\vec{\theta_0}$ and V_0, can be use

$$\vec{\theta}_{OLS} \approx arg\,\min_{\vec{\theta}\in\Theta} \sum_{j=1}^{N} \Big[\sum_{i=1}^{m} \frac{1}{\sigma_{0,i}^2} \Big(y_i(t_j) - f_i(t_j;\vec{\theta}) \Big)^2 \Big]$$

and

$$V_0 \approx \hat{V} = diag\Big(\frac{1}{N-p} \sum_{j=1}^{N} \Big[y_i(t_j) - f(t_j;\vec{\theta_0}) \Big] \cdot \Big[y_i(t_j) - f(t_j;\vec{\theta_0}) \Big]^T \Big)_{m\times m},$$

where p is the number of parameters to be optimize and the approximation for V_0 is an unbiased estimator. According to Agresti [1] and Cowan [20], when the N increase, the vector $\vec{\theta}_{OLS}$ follow an asymptotic property:

$$\theta_{OLS} \sim N(\vec{\theta_0}, \Sigma_0^N), \tag{3.21}$$

where Σ_0^N is the covariance matrix, which it can be approximated as

$$\Sigma_0^N \approx \left(\sum_{i=1}^{N} \chi_i^T(\vec{\theta_0}) V_0^{-1} \chi_i(\vec{\theta_0}) \right)_{p\times p}^{-1}, \tag{3.22}$$

where $\chi_j(\vec{\theta_0})$ is the sensitivity matrix defined as

$$\chi_j(\vec{\theta}) = \begin{bmatrix} \frac{\partial f_1(t_j;\vec{\theta})}{\partial \theta_1} & \frac{\partial f_1(t_j;\vec{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_1(t_j;\vec{\theta})}{\partial \theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m(t_j;\vec{\theta})}{\partial \theta_1} & \frac{\partial f_m(t_j;\vec{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_m(t_j;\vec{\theta})}{\partial \theta_p} \end{bmatrix}_{m\times p}, \tag{3.23}$$

where $j = 1, 2, \ldots, N$. The sensitivity matrix will describe the variance of the parameters from the model (p is the number of parameters). Finally, by using the approximation of V_0 and $\vec{\theta}_{OLS}$ that

$$\theta_{OLS} \sim N(\vec{\theta_0}, \Sigma_0^N) \approx N(\hat{\theta}_{OLS}, \hat{\Sigma}_0^N),$$

where

$$\hat{\Sigma}_0^N = \left(\sum_{i=1}^{N} \chi_i^T(\hat{\theta}_{OLS}) \hat{V}^{-1} \chi_i(\hat{\theta}_{OLS}) \right)_{p\times p}^{-1}, \tag{3.24}$$

and V is the approximation of V_0. It is possible to approximate the covariance matrix as

$$\hat{\Sigma}_0^N \approx \left(X^T(\hat{\theta}_{OLS}) \hat{V}^{-1} X(\hat{\theta}_{OLS}) \right)_{p\times p}^{-1}, \tag{3.25}$$

where the $X(\vec{\theta})$ is a $N \times p$ matrix given by

$$X_i(\hat{\theta}) = \begin{bmatrix} \frac{\partial f_i(t_1;\vec{\theta})}{\partial \theta_1} & \frac{\partial f_i(t_1;\vec{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_i(t_1;\vec{\theta})}{\partial \theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_i(t_N;\vec{\theta})}{\partial \theta_1} & \frac{\partial f_i(t_N;\vec{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_i(t_N;\vec{\theta})}{\partial \theta_p} \end{bmatrix}.$$

As a last procedure, the computation of the standard error for the k-th element of $\hat{\theta}_{OLS}$ can be calculated by the squared root of the (k, k) position of the matrix Σ_0^N, which is given by

$$SE_k(\hat{\theta}_{OLS}) \approx \sqrt{\hat{\Sigma}_{k,k}(\hat{\theta}_{OLS})}. \tag{3.26}$$

3.3.2 Generalized Least Squares (GLS) method

The OLS method provided in previous section assumed that the variances associated with the epidemic observations were longitudinally constant and not dependent on the values of the observations. According to Khan et al. [38], this may not be a realistic assumption especially if the epidemic data is influenced by a source of non-constant systematic error such as under-reporting.

It can be assumed that the mathematical model (System 2.9), together with a particular choice of parameters (θ_0) and initial conditions, describes the epidemic process, but the N observations $\vec{Y_j}$ are affected by random deviations from the process. Therefore, it is assumed that

$$\vec{Y_j} = F(t_j; \vec{\theta_0}) + F(t_j; \vec{\theta_0})^{\omega} \cdot \vec{\varepsilon_j}, \quad j = 1, \ldots, N, \tag{3.27}$$

where

$$F(t_j; \vec{\theta_0}) = diag\Big(f_1(t_j; \vec{\theta_0}), f_2(t_j; \vec{\theta_0}), \ldots, f_m(t_j; \vec{\theta_0})\Big).$$

In our case, since $m = 1$, then $F(t_j; \vec{\theta_0}) = f(t_j; \vec{\theta_0})$, which represent the weekly incidence that can be computed from the solution of the mathematical model, at time t_j, under the unknown vector of parameters, θ_0. The errors are assumed to be independent and identically distributed (*i.i.d.*) random variables with zero mean ($E(\vec{\varepsilon_j}) = 0$), representing the deviation of the model predictions from the observed data, and variance

$$V_0 = Var(\vec{\varepsilon_j}) = diag\Big(\sigma^2_{0,1}, \sigma^2_{0,2}, \ldots, \sigma^2_{0,N}\Big),$$

where $\sigma^2_{0,j}$, for $j = 1, 2, \ldots, N$, are unknown.

According to these assumptions, the observation mean is equal to model prediction,

$$E(\hat{Y}_j) = \begin{bmatrix} f_1(t_j;\vec{\theta}_0) \\ f_2(t_j;\vec{\theta}_0) \\ \vdots \\ f_m(t_j;\vec{\theta}_0) \end{bmatrix}_{m\times 1},$$

while the variance in the observations is a function of the time point, denoted as

$$Var(\hat{Y}_j) = diag\Big(\sigma^2_{0,1} f_1(t_j;\vec{\theta}), \sigma^2_{0,2} f_2(t_j;\vec{\theta}), \ldots, \sigma^2_{0,m} f_m(t_j;\vec{\theta})\Big)_{m\times m}.$$

As defined by Cintron et al. in [16], the GLS estimator for a set of observations $\vec{Y} = (\vec{Y}_1, \vec{Y}_2, \ldots, \vec{Y}_N)$ is the solution of the normal equations

$$\sum_{j=1}^{N} W_j \cdot \Big[\vec{Y}_j - f(t_j;\vec{\theta})\Big]^T \nabla_{\vec{\theta}} f(t_j;\vec{\theta}) = 0, \tag{3.28}$$

where the W_j are a set of non-negative weights, assuming $m = 1$, defined as

$$W_j = \frac{1}{f(t_j;\vec{\theta}_0)^{2\omega}}. \tag{3.29}$$

According to Banks et al. [4], if we assume $\omega = 1$ in Equation 3.29, we have that the weights are taken to be inversely proportional to the square of the predicted incidence. On the other hand, if $\omega = \frac{1}{2}$ then the weights are proportional to the reciprocal of the predicted incidence. The assumption of $\omega = 0$ leads to the standard Ordinary Least Squares (OLS) approach as described in previous section.

The rest of the analysis about the GLS method is similar to the method outlined in the previous section. For more details, we refer the reader to the work presented by Cintron et al. in [16], Khan et al. [38] and Banks et al. [4].

3.4 Markov Chain Monte Carlo (MCMC) method

During the twentyfirst century, the use of Markov chain Monte Carlo (MCMC) method has grown dramatically [75]. This method is useful for estimating posterior distributions under a Bayesian approach. It is a computer-driven sampling method that allows to characterize a distribution without knowing a precise information about the parameter or measure that we intend to estimate. A particular charactiristic of the MCMC process is that it can be used to draw samples from distributions even when all that is known about the distribution is how to calculate the density for various samples [18, 29].

In this project, the MCMC procedure is considered to generate a probability distribution for the initial exponential growth rate (ρ) and reporting rate (r) of the 2015-2016 Zika epidemic in Puerto Rico. In addition, a Monte Carlo simulation will also be used such that a sampling from random parameter distributions can produce a probability distribution for the basic reproductive number (R_0) of the epidemic.

In order to understand the process involved in the MCMC method, we have to analyze two properties: Monte Carlo and Markov chain. According to Gamerman and Lopes [29] and Ravenzwaaij et al. [75], the "Monte Carlo" is the process of estimating the properties of a distribution by inspecting random samples from other distributions. The advantage of the Monte Carlo method is to calculate the mean of a large sample of values, and this can be much easier than calculating the mean directly from the distribution equations. This advantage is most marked when random samples are easy to draw, and when the distribution equations are hard to work with in other ways [75].

In the other hand, the *Markov chain* property of MCMC have the idea that the random samples that are being generated are coming from sequential processes. By definition, each random sample is used as a stepping to the next random sample, and this notion created the *chain*. According to references [18, 29, 75],a special

characteristic of the chain is that, while each new sample depends on the one before it, new samples do not depend on any samples before the previous one, and this is the *Markov* property.

MCMC is a useful statistical tool in Bayesian inference because of the focus on posterior distributions [20]. Sometimes, the posterior distribution of a particular event (or process) is hard to find, and here is when the MCMC is helpful. MCMC can allow the user to approximate aspects of posterior distributions that cannot be directly calculated [18, 29, 75].

Bayesian inference uses the information provided by the observed data about a set of parameters, formally the *likelihood*, to update the *prior* state of beliefs about the set of parameters. The *prior* state of beliefs should be a guess probability distribution, according to previous research about possible values or characteristic of the set of parameters. Formally, the Bayesian inference is linked to the Bayes rule, which is defined ([20, 51, 62]) as

$$p(\theta|Y) \propto p(Y|\theta) \cdot p(\theta) \tag{3.30}$$

where θ indicates the set of parameters of interest and Y indicates the observed data, $p(\theta|Y)$ indicates the posterior distribution or the probability of θ given the data, $p(Y|\theta)$ indicates the likelihood or the probability of the data given the set of parameters θ, and $p(\theta)$ indicates the prior or the a-priori probability of the set of parameters θ.

The general idea about Bayesian inference is when an analytical expression for the likelihood is available, it can be combined with the prior information to derive a posterior distribution analytically [20, 51, 62]. However, if there is not access to an analytical expression, in Bayesian inference, this problem can be solved by implementing a MCMC procedure. We can draw a sequence of samples from the posterior and then examining their mean, range, and other statistical measures.

Some applications of the MCMC method are: Bayesian model comparison, memory retention, signal detection theory, extrasensory proccesing trees, risk taking, and primate decision making [75].

3.4.1 RJAGS Package: Bayesian graphical models using MCMC

In this work, a MCMC procedure (Bayesian approach) will be employ on several statistical analysis, for example, to estimate the initial exponential growth rate (ρ), the basic reproductive number (R_0) and the reporting rate (r) for the 2015-2016 Zika epidemic in Puerto Rico.

For the simulations through this approach, the R language and environment for statistical computation and graphics will be use with the "rjags" package. This package was created by Plummer et al. [56] in 2016, such that the package provides an interface from R to the JAGS library for Bayesian data analysis. JAGS, defined as "Just Another Gibbs Sampler" uses MCMC to generate a sequence of dependent samples from the posterior distribution of the parameters. A previous work presented by Martyn Plummer in 2013 as also considered, where a JAGS manual helps the users to understand the basics of modelling with JAGS [54].

It is important to mention that JAGS is a clone of "Bayesian analysis Using Gibbs Sampling (BUGS)". According to Casella and George [9], in BUGS and JAGS, a Gibbs Sampler is used, which is a technique for generating random variables from a marginal distribution indirectly, without having to calculate the density.

In general, Plummer et al. [56] indicates that the analysis using the "rjags" package proceeds with the following steps:

1. Define the model using the BUGS language in a separate file.
2. Read in the model file using the jags.model function. This creates an object of class "jags".

3. Update the model using the update method for "jags" objects. This constitutes a burn-in period.
4. Extract samples from the model object using the coda.samples function. This creates an object of class "mcmc.list" which can be used to summarize the posterior distribution. The "coda" package also provides convergence diagnostics to check that the output is valid for analysis.

For further details, see the R news from 2006 [55], where Plummer and Murell created a compendium of articles dedicated to Bayesian inference and MCMC simulation.

3.5 Negative Binomial (NB) distribution

The data used in this work will be linked to the NB distribution in order to perform some of the statistical analysis such as the estimation of the reporting rate (r) and initial exponential growth rate (ρ), through a Bayesian approach.

Let $\{y_j\}$ be the weekly incidence data of a Zika epidemic, for $j = 1, \ldots, N$, that follows a Negative Binomial distribution, with an over-dispersion parameter, ϕ, and probability parameter, $0 \leq p \leq 1$, then

$$P\left(y_j \mid \phi, p\right) = \frac{\Gamma(y_j + \phi)}{y_j! \cdot \Gamma(\phi)} p^{\phi}(1-p)^{y_j}. \tag{3.31}$$

Assuming that the over-dispersion parameter, ϕ, is the same for every y_j, then the expected value of each y_j is:

$$E[y_j] = \frac{\phi(1-p)}{p} \quad \longrightarrow \quad p = \frac{\phi}{\phi + E[y_j]}. \tag{3.32}$$

Finally, using $y(t) = y_0 e^{\rho t}$ and Equation 3.32, the Negative Binomial distribution (NBD) is written as

$$P\left(y_j \mid \phi, \rho\right) = \frac{\Gamma(y_j + \phi)}{y_j! \cdot \Gamma(\phi)} \left(\frac{\phi}{\phi + y_0 e^{\rho t}}\right)^{\phi} \left(\frac{y_0 e^{\rho t}}{\phi + y_0 e^{\rho t}}\right)^{y_j}, \tag{3.33}$$

where the vector of parameters is given by $(\phi, \rho)^T$. If $\phi \to \infty$, the Negative Binomial process approaches a Poisson process [62].

CHAPTER 4
ESTIMATION OF R_0 THROUGH THE INITIAL EXPONENTIAL GROWTH RATE

The basic reproductive number (R_0) is a threshold condition used to measure the transmission potential of a disease and it is defined by Anderson and May, in [2], as the average number of secondary individuals infected by one primary case in a complete susceptible population. According to Rothman et al. [65], R_0 can be thought as the number of secondary infections produced by a typical case of an infection in a population that is totally susceptible. It is important to note that R_0 is a dimensionless number and not a rate. In general,

$$R_0 \propto \frac{infection}{contact} \cdot \frac{contact}{time} \cdot \frac{time}{infection}, \tag{4.1}$$

or

$$R_0 = \tau \cdot \bar{c} \cdot d, \tag{4.2}$$

where τ is the transmissibility (for example, the probability of infection given contact between a susceptible and infected individual), $\bar{c}$ is the average rate of contact between susceptible and infected individuals, and d is the duration of infectiousness [65]. In terms of vector-borne diseases, the transmission can be given from a vector to a human when an infected vector is placed into a entirely susceptible human population, or from a human to a vector, otherwise. We always can identify the R_0 of a disease as the product of the R_0's of every way of transmission, depending on the disease [2, 65].

The basic reproductive number is affected by several factors, such that the rate of contacts in the host population, the probability of infection being transmitted during contact, the duration of infectiousness, among others, depending on the disease and the analysis. Remember that, in general, if $R_0 > 1$, a chain reaction starts, allowing that the infection invade the susceptible population, leading to a possible large outbreak, while $R_0 < 1$ means that the disease will produce less than one new infected individual on average, leading the disease to its disappearance from the population.

In Subsection 2.3.3, a mathematical model, established by Brauer et al. [7], was chosen to model the behaviour of Zika virus epidemic in Puerto Rico. For ZIKV, Towers et al. [72] derived an expression (Equation 2.22) for the basic reproductive number (R_0), depending on the initial exponential growth rate (ρ) of the epidemic, which is given by

$$\hat{R_0} = R0_{sex} + \left[\frac{(\rho + \alpha_h)(\rho + \gamma)}{\gamma\alpha_h} - R0_{sex}\right] \frac{(\rho + \delta)(\rho + \delta + \alpha_v)}{\delta(\delta + \alpha_v)}. \tag{4.3}$$

To estimate the probability distribution for R_0, a ten thousand Monte Carlo Iterations [50] will be perform, given the probability distributions of the parameters in the model ($\alpha_h, \alpha_v, \gamma, \delta$), and a probability distribution for ρ and $R0_{sex}$. For these parameters, the serial interval for the Zika virus was considered. Towers et al. [72], stated the serial interval of the Zika virus as the sum of the incubation periods plus the mean of the infectious periods [69], which is provided by:

$$T = \frac{1}{\alpha_h} + \frac{1}{\alpha_v} + \frac{1}{2\delta} + \frac{1}{2\gamma}. \tag{4.4}$$

For the simulations of this work, the time serial interval of 10 to 23 days (or 10/7 to 23/7 weeks, in our case) was used, in accordance to Majumder et al. [46].

From Equation 4.3, it can be noticed that the expression for R_0 depends on different parameters (see Table 2–3 for more details). To estimate R_0, distributions

to those parameters needs to be assigned, taking into account the values from the literature [12, 13, 15, 24, 27, 30, 40, 67, 72, 76], with an exception of $R0_{sex}$ and ρ. Currently, there are no estimates of $R0_{sex}$ in the literature. According to Towers et al. [72], we can assume lack of sustained ZIKV (Zika virus) transmission in areas free of the *Aedes Aegypti* vector, then $0 < R0_{sex} < 1$. This means that, with only direct transmission through sexual contact between humans, the epidemic cannot persist in the population. In this case, we can assume that $R0_{sex} \sim Unif(0, 1)$.

The other unknown parameter is the initial exponential growth rate, ρ. This parameter can be estimated from the incidence data and it allows to disregard the need of values for the transmission rates, β_h, β_v, and κ, which depends on the number of infected female mosquitoes bites and the probability of infection given the contact between the human and the mosquito, which are difficult to estimate. If the estimation of ρ cames from a Bayesian approach, then the posterior distribution can be taken as the probability distribution for the Monte Carlo iterations. On the other hand, for Non-Bayesian approach, the statistical analysis used will not produce a probability distribution, but it will estimate a single value (point estimate). In this case, Towers et al. [72] mentioned that it is recommended assign in a normal distribution around the point estimate, using the standard deviation that resulted from the estimation.

The probability distributions for α_h, α_v, γ, and δ were derived from the ranges that we found on the literature for these epidemiological quantities, by assuming a Uniform probability distribution over the range. Mahumder et al. [46] estimate the intrinsic latent period $(1/\alpha_h)$ to be in the range 3-12 days, the human infectious period $(1/\gamma)$ to be in the range 3-5 days and the extrinsic latent period $(1/\alpha_v)$ to be in the range 4-6 days. On the other hand, Chowell et al. [12, 14] estimated the average mosquito lifespan $(1/\delta)$ to be in the range 6-15 days and Kucharski et

al. [40] estimated $1/\alpha_h \in [2,7]$ days, $1/\gamma \in [3,7]$ days, $1/\alpha_v \in [10,15]$ days, and $1/\delta \in [10,20]$ days.

Table 4–1 shows the probability distributions (in weeks) that were used to estimate R_0 using Monte Carlo simulations, with Equation 4.3. Since the values for the parameters are measured in days, and our time unit is in weeks, an adequate conversion was made.

Parameter	**Probability Distribution**	**Reference**
$1/\alpha_h$	Unif($\frac{2}{7},\frac{12}{7}$) weeks	[46] [40]
$1/\alpha_v$	Unif($\frac{4}{7},\frac{15}{7}$) weeks	[46] [40]
$1/\gamma$	Unif($\frac{3}{7},\frac{7}{7}$) weeks	[46] [40]
$1/\delta$	Unif($\frac{6}{7},\frac{20}{7}$) weeks	[12] [14] [40]
$R0_{sex}$	Unif(0,1)	[72]
$\frac{1}{\alpha_h}+\frac{1}{\alpha_v}+\frac{1}{2\delta}+\frac{1}{2\gamma}$	$\frac{10}{7}$ to $\frac{23}{7}$ weeks	[46]
ρ	N($\hat{\rho}$, $\hat{s}$) or Posterior distribution	[72]

Table 4–1: Probability distributions for the parameters in the expression of R_0 (Equation 2.22).

The simulation consist of sampling (ten thousand iterations) from the probability distributions of every parameters in Table 4–1, taking into consideration the time serial interval for the ZIKV, in order to produce a probability distribution for the basic reproductive number, by using Equation 4.3.

The data set used in this research will be shown as well as the estimation of the basic reproductive number though the initial exponential growth rate. The discussion will be presented in two categories, Non-Bayesian and Bayesian, in order to distinguish between the implementation of the statistical methods presented in Section 3.2. Before we present the results of ρ and R_0, we will start this chapter by

providing information about the available data for the 2015-2016 Zika epidemic in Puerto Rico.

4.1 Data set: 2015-2016 Zika epidemic in Puerto Rico

The data used in this work, consist of the weekly number of cases (incidence) of the Zika virus. The weekly incidence was extracted from the weekly reports published online [57] by the Puerto Rico Department of Health (PRDH) using a Plot Digitizer [61]. Since the PRDH did not provided the data set for the Zika cases per week and only shared weekly reports with graphs, we used the plot digitizer to extract the data from the graphs, at the same time that a new report was published. This technique of extraction of raw data is very useful for the exploration of patterns in a data and to simulate certain phenomena, if the raw data was not accesable, as in our case. The extraction process was performed using a package in R Project [70]. In Figure 4–1, it can be observed the weekly incidence curve (yellow line) presented in the last public report provided by the PRDH. In addition, the same graph shows the weekly incidence for the Chikungunya and Dengue virus, during the same period of time. The number of cases for those viruses are relatively small in comparison with the weekly cases reported for the Zika virus.

At the end of 2016, almost 37,500 people got infected, from which 66% of the cases were females. After the extraction process, with the Plot Digitizer, our data set consist of 58 observations, from November 2015 to December 2016. The weekly incidence curve, and cumulative incidence curve that were obtained, are presented in Figure 4–2 (a) and (b), respectively.

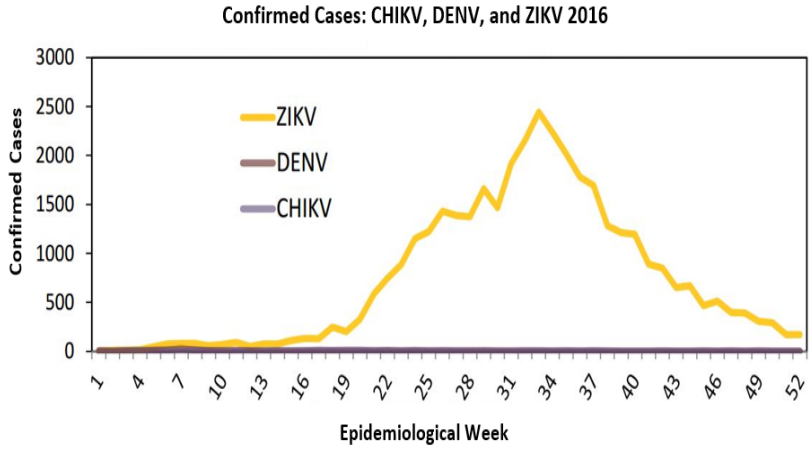


Figure 4–1: Screenshot: Weekly incidence curve from the last public report of 2016, by the Puerto Rico Department of Health (PRDH)[57].

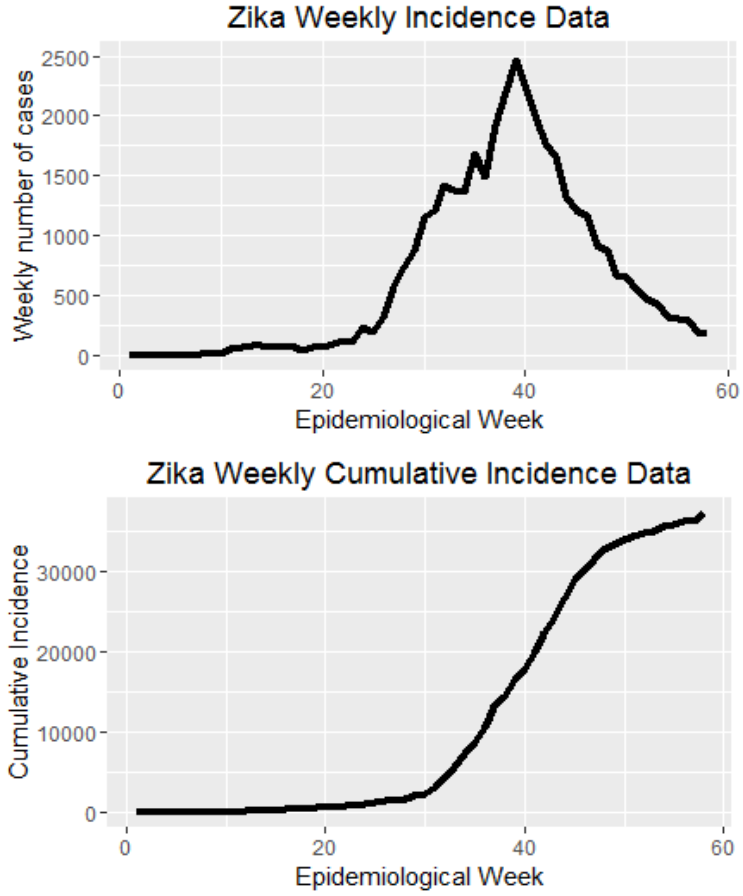


Figure 4–2: (a) Weekly incidence curve after the extraction process with the Plot Digitizer. (b) Weekly cumulative incidence curve using our data set.

Several statistical models can be used to represent the number of infected individuals with the Zika virus, but the Poisson distribution and Negative Binomial distribution are the most common in Epidemiology. In the analysis of infectious diseases, according to Bettencourt and Ribeiro [5], there are many scenarios where the use of a Poisson process is adequate as well as the Negative Binomial distribution. Bettencourt and Ribeiro [5] compared these two statistical models in order to provide guidelines for others researchers, about which distribution is more adequate, depending on the problem that you are facing. If the data shows that the mean and variance are the same, then the Poisson distribution is recommended to model the number of infected individuals. Otherwise, if the variance is bigger than the mean, this create an over-dispersion in the data and is better to use the Negative Binomial distribution.

This comparative is explained in details by Bettencourt and Ribeiro, in the supplementary material of [5]. From [43], we have learned that the Negative Binomial distribution has many applications, such as the analysis of parasite loads, species occurrence, parasitoid attacks, abundance samples, spatial clustering of populations and among others. Lloyd-Smith [43] has explained the importance of running statistical tests to be sure about what kind of data one is working with.

In recent works [40, 67, 72], we noticed the constant use of the Negative Binomial distribution as the statistical model not only to account for the number of new infected individuals (incidence), but also to estimate the initial exponential growth rate (ρ) of the epidemic and consequently, to estimate the basic reproductive number (R_0), through Equation 2.22. Details about this, in our case, will be provided in Chapter 5.

4.1.1 Statistical test: Goodness of fit

Following the advise of Lloyd-Smith [43], we performed a statistical test in order to decide if the data set we extracted from the weekly report, published by the Puerto Rico Department of Health, follows a Poisson distribution or a Negative Binomial distribution. We decided to implement a Goodness of fit test to the observed data ($\vec{Y_j}$ for $j = 1, 2, \ldots, 58$), which is the weekly incidence data for the 2015-2016 Zika virus epidemic in Puerto Rico, as plotted in Figure 4–2.

According to [20, 51, 62], we can use a Goodness-of-fit criteria to select the best distribution of a data set between a variety of models that we assume could fit the data. Using the function $gofstat()$, from the R package "fitdistrplus" [22], we run the R code assuming that our data can possible follow a Poisson distribution or a Negative Binomial distribution. The R output is presented in Figure 4–3.

```
Goodness-of-fit criteria
                                 Poisson   NegBin
Akaike's Information Criterion 40434.07 810.7367
Bayesian Information Criterion 40436.06 814.7147
```

Figure 4–3: R output of the Goodness-of-fit criteria between a Poisson distribution versus a Negative Binomial distribution.

To understand the results, we studied the meaning behind the Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC). For that, we consulted the references [51, 62], that define AIC as "an estimate of a constant plus the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model", so that a lower AIC means a model is considered to be closer to the truth. On the other hand, BIC is "an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup", so that a lower BIC means that a model is considered to be more likely to be the true model. Based on the values for AIC and BIC (see Figure 4–3), the test indicate

that our data can be better describe as a Negative Binomial (NB) distribution than a Poisson distribution. For details about the probability density function of the NB distribution, see Section 3.5. Now, let start with the estimation of R_0 through the estimate of ρ.

4.2 Estimation of R_0 from a Non-Bayesian approach

4.2.1 VM for exponential phase and Linear Regression (LR) for ρ

Previously, in Section 3.2.1, we mentioned that the exponential phase of a epidemic is a J shaped growth curve that represent the increase of the number of reported cases over time, until the maximum number of cases is observed. Then, the phase will end before the curve start rapidly decreasing. In our case, according to the graph of the weekly incidence data, presented in Figure 4–2(a) visually, there are three possible endpoints of the initial exponential phase, identified as the epidemiological weeks 32, 35, and 39. These three epidemiological weeks are marked as blue, green, and red, respectively, in Figure 4–4.

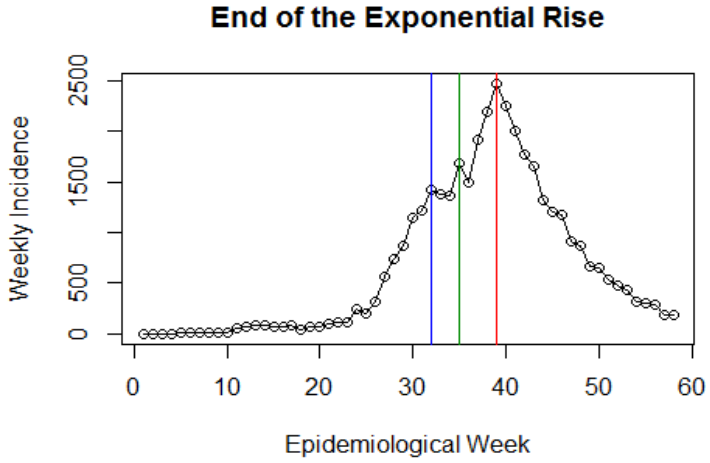


Figure 4–4: Possible endpoints of the exponential phase, visually taken.

Then, we used the cumulative incidence curve presented in Figure 4–2(b), to estimate the initial exponential growth rate, ρ. For this, we use the logarithm of the

cumulative incidence to make the data looks linear and to estimate the slope of the linear regression equation. A useful way to think about the cumulative incidence (incidence proportion) is that it can be asssociated with a probability of developing a disease, or getting an infection, over a given period of time; as such, it is an estimate of risk [2].

Our results consist of three estimates of ρ, one for each endpoint of the exponential phase. The estimate of ρ is the slope of the linear regression equation. This value is provided by the second coefficient of the summary function after we run the linear regression in R version 3.3.1, with the function $lm(X_2 \sim X_1)$, where X_2 is the vector with the logarithm of the cumulative incidence and X_1 is the vector of time (epidemiological weeks).

Through the method discussed in this subsection, we found that the estimate for the initial exponential growth rate is $\hat{\rho} = 0.26$ weeks^{-1} $[0.23, 0.29]$ 95% confidence interval (CI) if week 32 is the end of the exponential rise. If we choose the time where the maximum number of cases occurs, as the end of the exponential phase (week 39), then $\hat{\rho} = 0.24$ weeks^{-1} $[0.24, 0.26]$ within a 95% CI.

If we observe Table 4–2 and Figure 4–5, the estimates of ρ are approximated to each other, no matter what epidemiological week is selected as the exponential end phase. Moreover, we can notice that the estimates are decreasing as the times increase. This behavior was pointed out in their work by Favier et al.. A summary of the results from this method are provided in Table 4–2, showing the estimates and the corresponding 95% confidence interval of the estimation, and Figure 4–5, with the output graph from the analysis in R.

Week	$\hat{\rho}$ **(weeks^{-1})**	**95%CI**
32	0.26	[0.23, 0.29]
35	0.25	[0.23, 0.28]
39	0.24	[0.22, 0.26]

Table 4–2: Estimates of the initial exponential growth rate at week 32, 35, and 39, respectively.

In this scenario, a Normal distribution around the point estimate of ρ was used, and the standard deviation was obtained from the summary of the corresponding linear regression. The standard deviation are 0.0137, 0.0118, and 0.0099, for the exponential phase ending at week 32, 35, and 39, respectively. In addition, for $R0_{sex}$, a Uniform distribution from 0 to 1 was assumed, as defined in [72]. After the Monte Carlo Iterations, the probability distribution of R_0 is presented in Figure 4–6, which is represented by a right-skewed distribution. For the estimated value of the basic reproductive number, we choose the mean of the probability distribution. Therefore, for a ZIKV epidemic in Puerto Rico with an exponential phase ending in the epidemiological week 32, we estimated the basic reproductive number to be $\bar{R}_0 = 1.80$ $[1.40, 2.38]$ 95% credible region (CR). If the exponential phase ends at week 39, then $\bar{R}_0 = 1.72$ $[1.72, 2.23]$ 95% CR. For more information about the resulted mean estimates and probability distributions, see Table 4–3 and Figure 4–6, respectively.

Week	$\bar{R}_0$	**95%CR**
32	1.80	[1.40, 2.38]
35	1.76	[1.38, 2.31]
39	1.72	[1.36, 2.23]

Table 4–3: Mean estimates of the basic reproductive number (R_0) at week 32, 35, and 39, respectively, by using the Visual method: Linear regression to the logarithm of the cumulative incidence data.

Since the estimated values of $\hat{R}_0$ are bigger than one, we can certainly confirm that a Zika virus epidemic occurred in Puerto Rico, no matter when the exponential

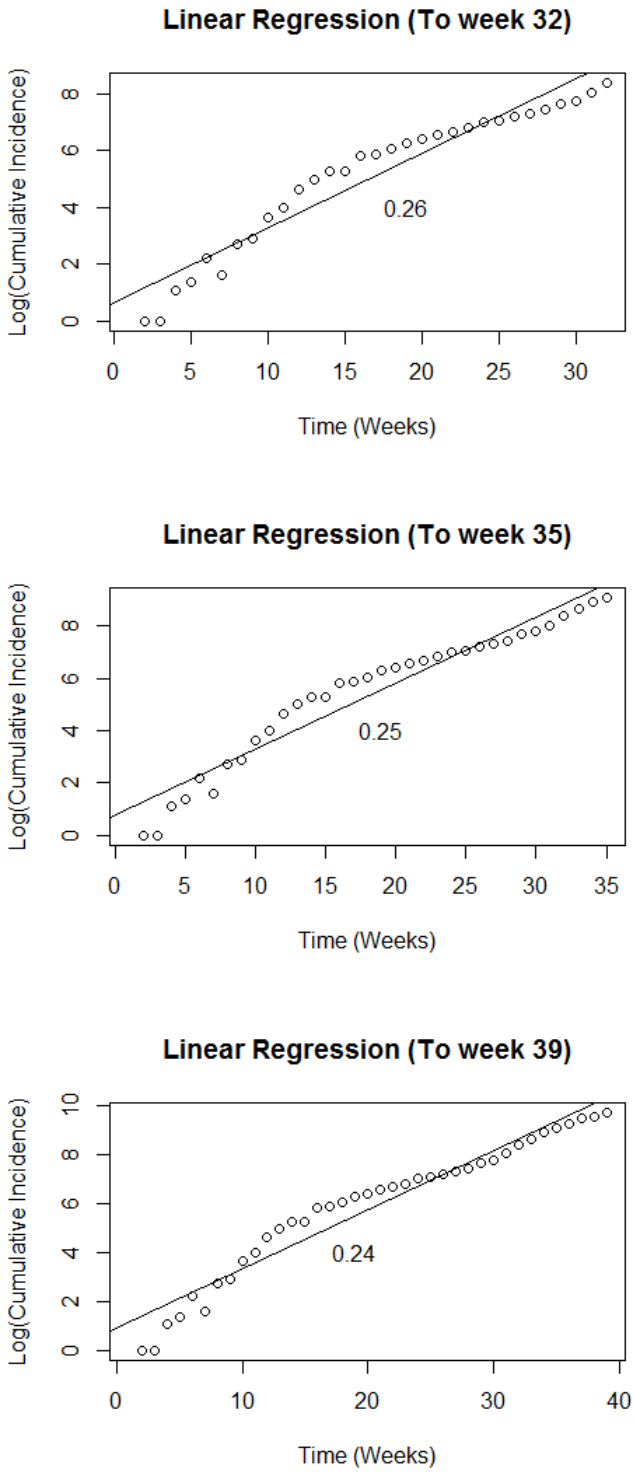


Figure 4–5: Estimates of the initial exponential growth rate by using linear regression to the logarithm of the cumulative incidence, to week 32, 35, and 39, respectively.

phase ended. The R_0 for the 2013-2014 ZIKV outbreak in French Polynesia was estimated by Kucharski et al. [40] to be between 2.6 to 4.8, resulting in a statistical agreement with the estimate of the R_0 obtained by Towers et al. [72], $R_0 = 3.8$. In addition, Gao et al. [30] estimated R_0 for the ZIKV epidemic from Colombia, Brazil and El Salvador, to be 2.055 $[0.523, 6.300]$.

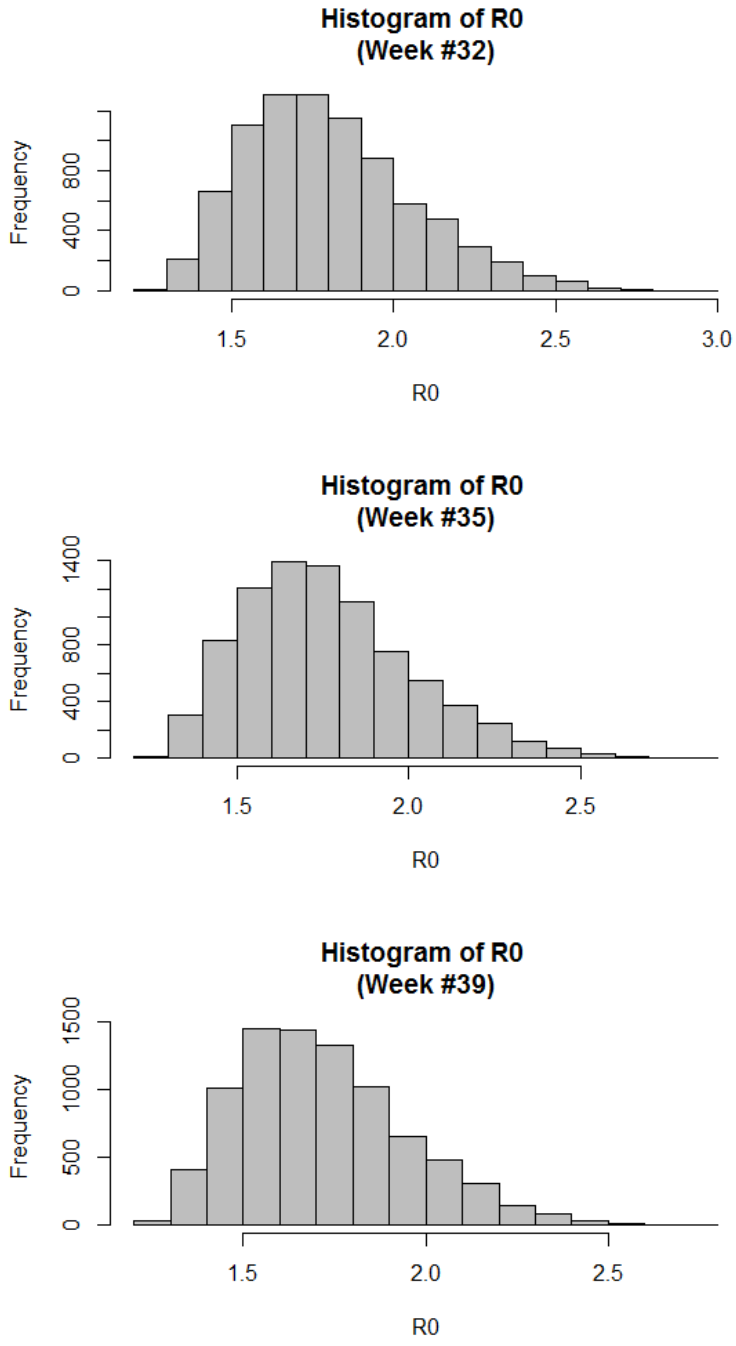


Figure 4–6: Probability distribution of the basic reproductive number (R_0) for ZIKV outbreak in Puerto Rico, if we use linear regression to the logarithm of the cumulative incidence, to week 32, 35, and 39, respectively.

4.2.2 Favier et al.'s method for exponential phase and ρ

For this method, the weekly incidence data against the weekly cumulative incidence is used. Favier et al.'s method [27], which is explained in Subsection 3.2.2, finds the point where the exponential phase ends, and at the same time, obtain an estimate for the initial exponential growth rate, ρ. However, Favier et al. used the term "force of infection" as an equivalent to the initial exponential growth rate of the epidemic. They performed linear fits, each time using one more data point and plotting the evolution of the goodness-of-fit of the corresponding linear regressions. Once we performed the analysis suggested by Favier et al., in [27], with our data set, we obtained the set of graphs in Figure 4–7.

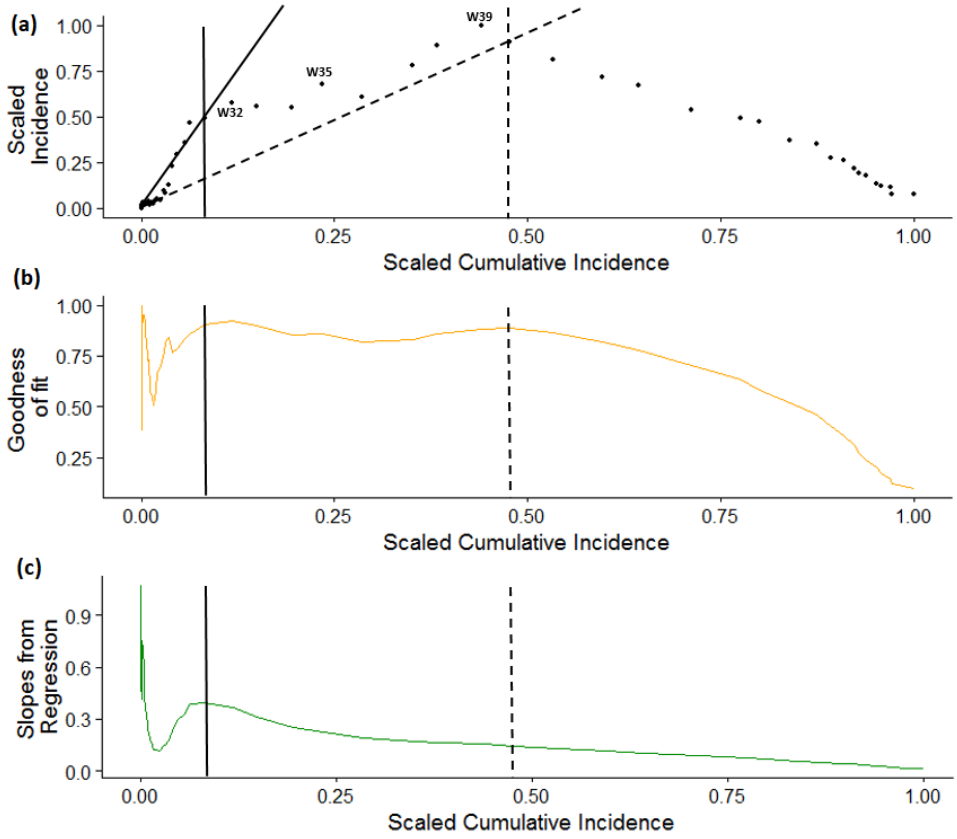


Figure 4–7: (a) Weekly incidence data against the cumulative incidence data of Zika epidemic; (b) Evolution of the goodness-of-fit. We used the R-Squared value, to measure the goodness-of-fit of the linear regression; (c) Evolution of the slopes, which are the estimates of the force of infection.

Favier et al.'s method suggests that choosing the end point given by the maximum goodness-of-fit (we took into consideration the R-Squared value) as an indicator of the end of the exponential phase of the epidemic is not adequate, because it can produce an underestimation of the force of infection (ρ). For instance, see week 40 (vertical dashed line) on the graph, in Figure 4–7(c), in which the estimate will be relatively small. In the same graph, we can observe that in the beginning of the evolution of the slopes, the graph stops fluctuating greatly at the 31st epidemiological week, where it oscillates around a value (vertical solid line). The best estimation of the force of infection comes from the slope of the linear regression equation that lies between the two regression lines (Figure 4–7(a)). Favier et al. [27] suggest to consider the 80th percentile of the slopes to the left of the maximum R-Squared value (Figures 4–7(b) and 4–7(c)), to find a good estimate of ρ. With this information, week 40 cannot be the end of the exponential rise, neither week 31 since in that week the fluctuations stops. We determined, by using the method of Favier et al. [27], that the estimates of initial exponential growth rate corresponds to an estimate between the epidemiological weeks 33 and 34. From the same calculation we did before, using the R Project [70], we determined that the estimates of initial exponential growth rate is $\hat{\rho} = 0.29$ weeks^{-1} $[0.27, 0.35]$ 95% CI, corresponding to an estimate between the epidemiological weeks 33 and 34.

Endpoint	$\bar{\rho}$ **(weeks^{-1})**	**95% CI**
Between 33 and 34	0.29	[0.27, 0.35]

Table 4–4: Estimate of the initial exponential growth rate from Favier et al.'s method.

In this method, we found that the end of the exponential phase lies between the epidemiological week 33 and 34 with an estimate of $\hat{\rho} = 0.29$. After the Monte Carlo iterations, we have that $\bar{R_0} = 1.91$ with [1.45, 2.60] 95% CR. The probability distribution is in Figure 4–8, showing the same shape as before.

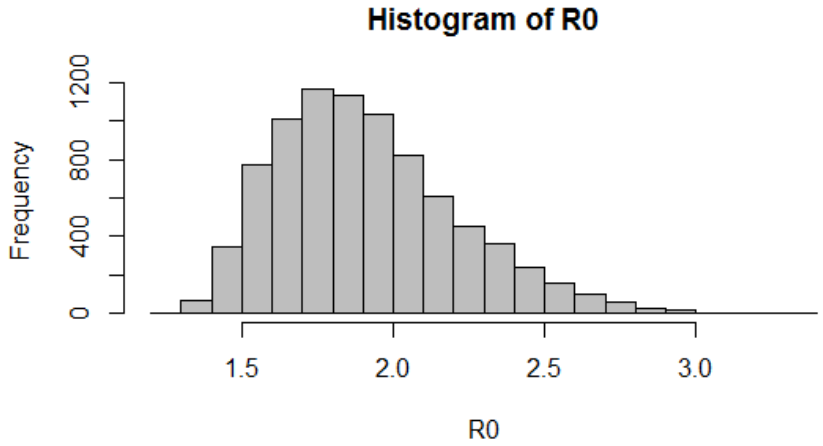


Figure 4–8: Probability distribution of the basic reproductive number (R_0) if we use the method defined by Favier et al. [27] to estimate ρ.

4.3 Estimation of R_0 from a Bayesian approach

4.3.1 VM for exponential phase and MCMC for ρ

In this method, we will mix two ideas: (1) The use of the Negative Binomial distribution, and (2) a Bayesian approach, in order to estimate the initial exponential growth rate (ρ) of the ZIKV epidemic in Puerto Rico. We take a close attention to Kucharski et al.'s work in [40], where a Markov Chain Monte Carlo (MCMC) simulation was performed to calculate the reporting rate (r). Here we used their idea of MCMC, to obtain a posterior distribution of ρ for the Zika epidemic in Puerto Rico. In Subsection 4.1.1, we showed that our data seems to follows a Negative Binomial distribution and the definition of this probability distribution is given by

$$P\left(y_i \mid \phi, \rho\right) = \frac{\Gamma(y_i + \phi)}{y_i! \cdot \Gamma(\phi)} \left(\frac{\phi}{\phi + y_0 e^{\rho t}}\right)^{\phi} \left(\frac{y_0 e^{\rho t}}{\phi + y_0 e^{\rho t}}\right)^{y_i}, \tag{4.5}$$

where ρ is the initial exponential growth rate and ϕ the over-dispersion parameter. To implement this Bayesian approach, we considered Equation 4.5 and prior distributions for ρ and ϕ. From the results in [7], also for ZIKV, we can assume a $Uniform(0, 1)$ distribution as a prior for the initial exponential growth rate (ρ). For

the over-dispersion parameter (ϕ), we choose a $Uniform(0, 100)$ prior distribution, instead of a $Uniform(0, \infty)$. The reason of our selection is because of the implementation procedure of the MCMC in the statistical software R (the next paragraph will address this). Given that the prior distributions for parameters have been assessed, the next procedure is to combine the likelihood function with priors to make a Bayesian inference. This allows us to simulate via Markov Chain Monte Carlo (MCMC) [18, 50, 56], to obtain a posterior distribution of ρ. For purpose of the simulation, we run 8 replicates of 25,000 iterations, each one with a burn-in period of 5,000 iterations, using the package "rjags" [56] in the statistical program R. The MCMC simulation provides a posterior distribution of ρ and ϕ, by using the current information about the parameters that comes from the literature.

In order to perform the MCMC procedure in the statistical software R, all the prior distributions in model (*jags* model) need to be *proper* distributions [56]. By definition, an unnormalized density, $f(\theta)$, is *proper* if $\int f(\theta)d\theta < \infty$, and otherwise it is *improper* [18]. In addition, the references [18, 50] indicate the following Theorem: "*If the prior is proper and the data is discrete, then the posterior is proper.*" Since our prior distributions for ρ and ϕ, and the data used satisfy the hypothesis of the theorem, we can conclude that the posterior distribution will be a proper distribution.

The results of this method is a posterior distribution for the initial exponential growth rate (ρ) and over-dispersion parameter (ϕ), corresponding to the three possible endpoints of the exponential phase (weeks 32, 35, and 39, respectively). In Figure 4–9, we can observe the updated information (posterior distribution) of the initial exponential growth rate, which is approximately normally distributed. Figure 4–9 can be described as the histogram of every possible estimated value for the parameter and its frequencies, according to the iterations performed in order to update the information about the vector of parameters.

For the estimation of ρ, we selected the mean of the posterior distribution of ρ. In the case of a exponential phase ending at week 32, we estimated $\bar{\rho} = 0.24$ weeks^{-1} $[0.23, 0.26]$ with a 95% credible region (CR), or $\bar{\rho} = 0.23$ weeks^{-1} $[0.22, 0.24]$ with a 95% credible region (CR) if the exponential rise ends at the epidemiological week 39. See Table 4–5 for further information about the estimated values of this approach.

Week	$\bar{\rho}$ **(weeks^{-1})**	**95% Credible Region**
32	0.24	[0.23, 0.26]
35	0.24	[0.23, 0.25]
39	0.23	[0.22, 0.24]

Table 4–5: Mean estimate of the initial exponential growth rate at week 32, 35, and 39, respectively.

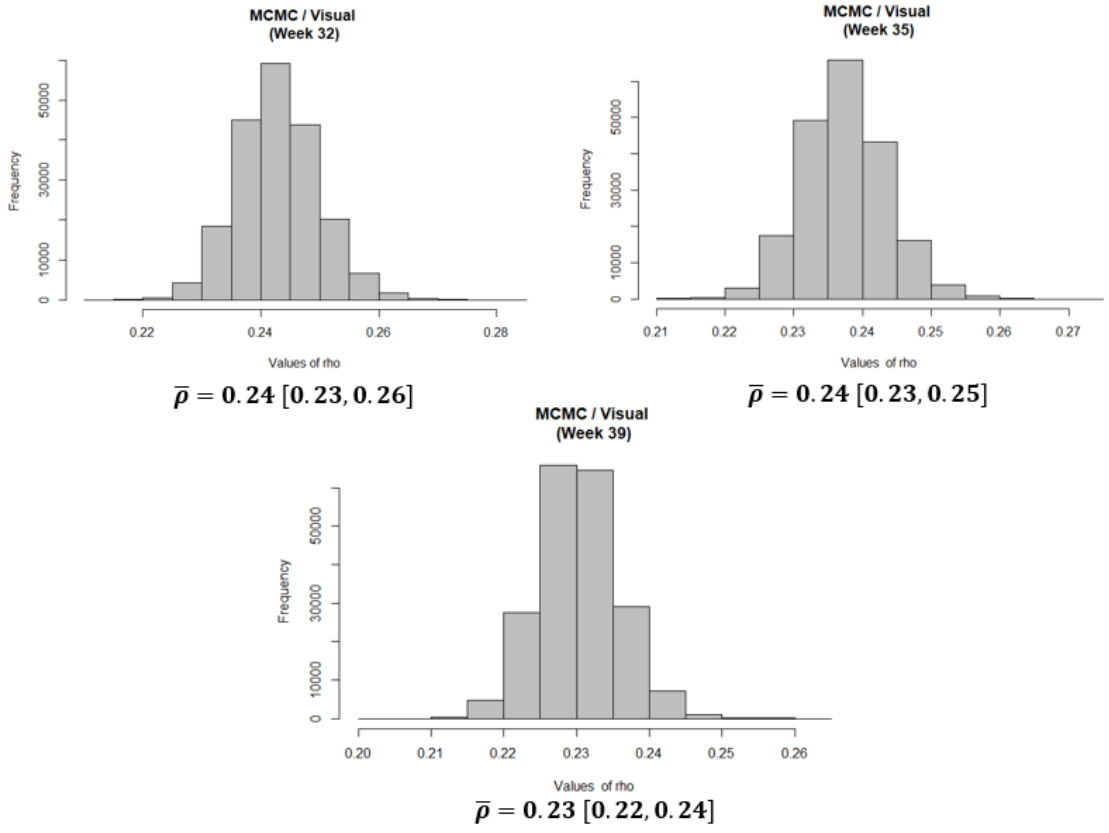


Figure 4–9: Posterior distributions for the initial exponential growth rate.

For this method, the estimates for the basic reproductive number are $\bar{R}_0 = 1.72$ $[1.37, 2.23]$ 95% CR, $\bar{R}_0 = 1.70$ $[1.35, 2.20]$ 95% CR, and $\bar{R}_0 = 1.67$ $[1.34, 2.15]$ 95% CR, corresponding to an exponential phase ending at the epidemiological weeks 32, 35, and 39, respectively. Again, the probability distribution of the basic reproductive number, given in Figure 4–10, shows a right-skewed distribution.

Week	$\bar{R}_0$	**95%CR**
32	1.72	[1.36, 2.23]
35	1.70	[1.35, 2.19]
39	1.67	[1.34, 2.14]

Table 4–6: Mean estimates of the basic reproductive number (R_0) at week 32, 35, and 39, respectively, by using a Bayesian approach to estimate ρ.

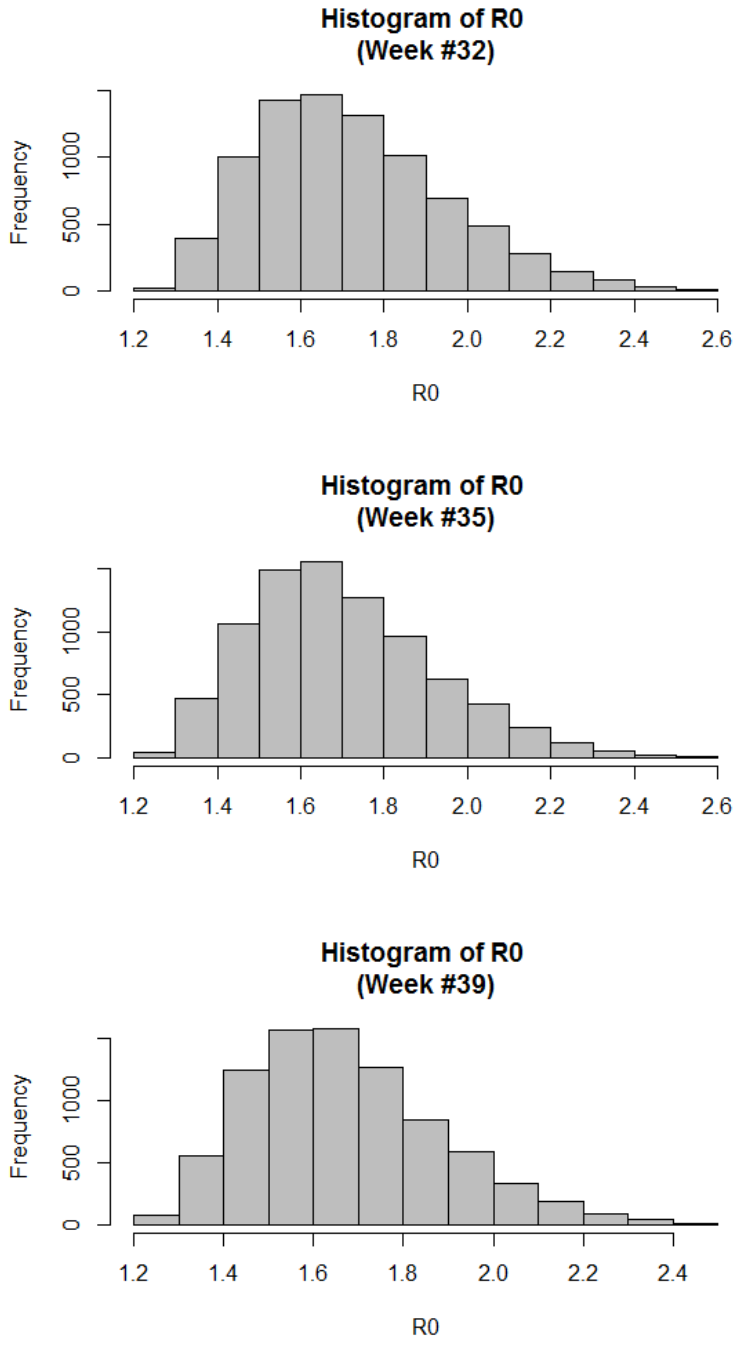


Figure 4–10: Probability distribution of the basic reproductive number (R_0) for ZIKV outbreak in Puerto Rico, through the Visual Method using a Bayesian approach.

4.3.2 Chowell et al.'s method for exponential phase and MCMC for ρ

The exponential growth rate, ρ, can be estimated from the exponential growth phase of the pandemic, using a Poisson Maximum Likelihood Method, which is explained in the supplementary document of [15]. Instead of using a Maximum Likelihood Method, we use a Bayesian approach with the Negative Binomial distribution as indicated in [12–15, 40]. In this case, Chowell et al.'s method will give us an specific week to where the exponential phase ends.

If y_j^{data} are the weekly incidence of the Zika epidemic in Puerto Rico, at t_j time points, where $j = 1, ..., 58$, the standard deviation width of the epidemic curve is given by

$$\sigma_t = \sqrt{\frac{\sum_{j=1}^{58}(t_j - \bar{t})^2 y_j^{data}}{\sum_{j=1}^{58} y_j^{data}}}, \tag{4.6}$$

where

$$\bar{t} = \frac{\sum_{j=1}^{58} t_j y_j^{data}}{\sum_{j=1}^{58} y_j^{data}}. \tag{4.7}$$

The exponential rise portion of the epidemic curve consist of incidence data points at the beginning of the epidemic that are sufficiently many standards deviations away from the time of peak incidence (denoted by t_{peak}) [15]. The exponential rise is thus the region where

$$t_j < (t_{peak} - f \cdot \sigma_t). \tag{4.8}$$

According to Zika weekly incidence data, the t_{peak} occur in the epidemiological week 39. Thus, by using R, an estimate of the values from Equation 4.6 and Equation 4.7 can be calculated. We only need to identify the value of f that provides unbiased estimates of the true exponential rise. As Chowell et al. found on their work, we also determined that $f \geq 1$, which means, that the exponential phase of the Zika epidemic in Puerto Rico, consist of data points that are one standard deviation away from the epidemic peak. We also run the mathematical model, System 2.9, to ensure

that the process coincide with the results from our data set. From the calculations with Equation 4.8, assuming $f = 1$, we found that the exponential phase ends in the epidemiological week 31.

From Chowell et al.'s method and the MCMC process, we found that the estimate of the initial exponential growth rate is given by the mean of the posterior distribution, presented in Figure 4–11, while Table 4–7 shows the estimate of ρ and the 95% CR. We determined that $\bar{\rho} = 0.24$ weeks^{-1} $[0.23, 0.26]$ 95% CR. The estimate for the basic reproductive number is $\bar{R_0} = 1.73$ with [1.37, 2.24] 95% CR. The probability distribution for R_0 is in Figure 4–12.

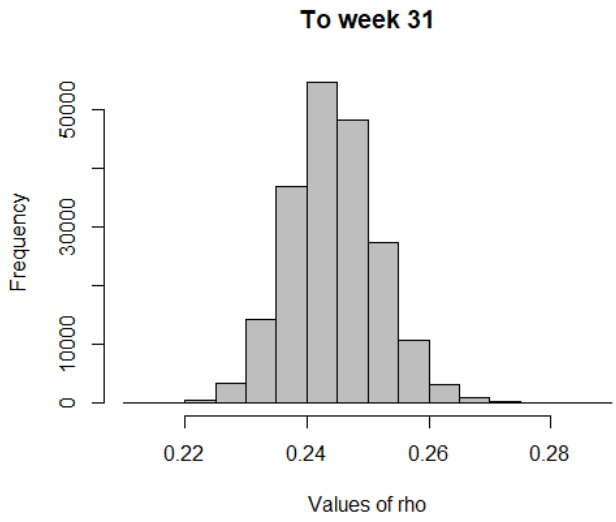


Figure 4–11: Posterior distribution for the initial exponential growth rate at week 31, by using Chowell et al.'s method.

Time peak	**Endpoint**	$\bar{\rho}$ **(weeks^{-1})**	**95% CR**
39	31	0.24	[0.23, 0.26]

Table 4–7: Mean estimate of the initial exponential growth rate at week 31 by using a Bayesian approach.

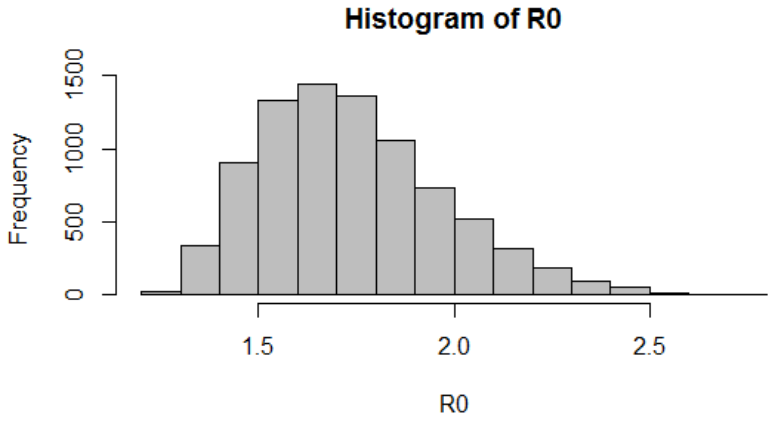


Figure 4–12: Probability distribution of the basic reproductive number (R_0), using Chowell et al.'s method.

4.4 Discussion

The purpose of this chapter was on the estimation of the basic reproductive number (R_0) taking into consideration the initial exponential growth rate of the 2015-2016 Zika epidemic in Puerto Rico, where two different approaches were taken: Non-Bayesian and Bayesian. Figure 4–13 shows a summary of the estimates for the exponential growth rate (ρ) according to the methods discussed in this chapter, while Figure 4–14 shows a summary of the estimates for basic reproductive number, based on the initial exponential growth rate of the 2015-2016 ZIKV epidemic in Puerto Rico.

Method	**End of the Exponential Phase**	**$\widehat{\rho}$ (weeks^{-1}) 95%CI or 95%CR**	
		Non-Bayesian	**Bayesian**
Visual	Week 32	0.26 [0.23, 0.29]	0.24 [0.23, 0.26]
	Week 35	0.25 [0.23, 0.28]	0.24 [0.23, 0.25]
	Week 39	0.24 [0.22, 0.26]	0.23 [0.22, 0.24]
Favier	Between week 33 and 34	0.29 [0.27, 0.35]	
Chowell	Week 31		0.24 [0.23, 0.26]

Figure 4–13: Summary of the estimates for the exponential growth rate.

Method	End of the Exponential Phase	$\widehat{R_0}$ 95%CI or 95%CR	
		Non-Bayesian	Bayesian
Visual	Week 32	1.80 [1.40, 2.38]	1.72 [1.36, 2.23]
	Week 35	1.76 [1.38, 2.31]	1.70 [1.35, 2.19]
	Week 39	1.72 [1.36, 2.23]	1.67 [1.34, 2.14]
Favier	Between week 33 and 34	1.91 [1.45, 2.60]	
Chowell	Week 31		1.73 [1.37, 2.24]

Figure 4–14: Summary of the estimates for basic reproductive number (R_0).

At this point, we noticed that the estimates from the Visual Method provides similar estimates for each endpoint of the exponential phase. However, the method defined by Favier et al. is the method that estimated the highest value for the initial exponential growth rate and R_0. In addition, the estimates of ρ, from Chowell et al.'s method, seems to be similar to the estimates from Visual Method, especially if the exponential phase ends in week 35 or 39.

The estimates obtained in this chapter, for the basic reproductive number, ranged from 1.67 to 1.91, where the maximum estimated value is provided by the Favier et al.'s method, while the minimum is provided by the Visual method when a Bayesian approach was applied to the exponential phase ending in the epidemiological week 39.

Since the Zika virus is an emergent disease in Puerto Rico, our estimates of ρ, as we know, can be the first estimate of the basic reproductive number of the ZIKV epidemic in Puerto Rico. Because of few studies about the ZIKV during the time of this work, the Bayesian approach seems to be the best scheme to follow because it enables the researcher to update their belief about the parameters in the mathematical model, by using the available data provided by the health agencies.

CHAPTER 5
ESTIMATION OF R_0 THROUGH THE INVERSE PROBLEM ESTIMATION

In this chapter, the scenario of the under-reporting problem and estimate of the appropriate reporting rate, r, for the 2015-2016 ZIKV epidemic in Puerto Rico, will be discuss before the estimation of R_0 through the the inverse problem estimation.

The Zika virus is mostly asymptomatic, and according to the World Health Organization [77], the symptoms are typically mild even upon clinical presentation, with a very similar diagnosis to that of Dengue and Chikungunya virus. The ZIKV is not only transmitted by mosquitoes but also from human to human, for example: through sexual transmission, blood transfusion and mother-to-fetus. In Chapter 1, we briefly discussed the concern about the connection between the Zika virus with the increased risk of the Guillan-Barré syndrome and neonate Microcephaly [26], and in addition, in most of the cases, pregnant women may not even know that they have the virus. On the other hand, since the symptoms of the ZIKV are mild, some people prefer not to seek medical assistance.

According to the discussion from previous paragraph, about asymptomatic cases and infected people that are not seeking medical assistance, we can conclude that our Zika weekly incidence data may just be a fraction of the true number of infected people in Puerto Rico. Fauci and Morens [26] confirmed that there exist a high rate of asymptomatic infection. An estimated 80% of persons infected with Zika virus are asymptomatic [24, 48, 53]. In the work of Kucharski et al. [40], the number of infected individuals who reported their symptoms is estimated to be in the range

of 7% and 17% of the total number infected by the virus. This situation about non counted cases is called the *under-reporting problem*.

We considered previous results about reporting rates, in particular the works by Kucharski et al. [40] and Shutt et al. [67], to use a Bayesian approach, in order to estimate the reporting rate, r. As we discussed in Subsection 4.1.1, our weekly incidence data follows a Negative Binomial distribution. Since the reporting rate (r) of the incidence data represent a fraction of the true incidence data (that is reported by health organizations), we can rewrite Equation 4.5 for the probability distribution as follow:

$$P\left(y_j \mid \phi, \rho\right) = \frac{\Gamma(y_j + \phi)}{y_j! \cdot \Gamma(\phi)} \left(\frac{\phi}{\phi + r \cdot y_0 e^{\rho t}}\right)^{\phi} \left(\frac{r \cdot y_0 e^{\rho t}}{\phi + r \cdot y_0 e^{\rho t}}\right)^{y_j}, \tag{5.1}$$

where y_j is the observed data (Zika weekly incidence data at time iȷ), ρ the initial exponential growth rate, ϕ the over dispersion parameter, r represent the reporting rate, and y_0 is the number of cases reported at time $t = 0$.

After running the mathematical model in System 2.9 and the resulted simulated incidence is plotted against the Zika weekly incidence data from Puerto Rico, we obtained the graph presented in Figure 5–1. It can be observe that the Zika weekly incidence data (black dots) is a small fraction of the predicted curve by the model (red line). The model in System 2.9 was performed by considering the total population of Puerto Rico (3.7 million [19]) as the initial condition (S_0) of the susceptible population (S_h). Since at the beginning of an infection, no one is already recovered, the initial number of recovered human is set to zero at time $t = 0$ (or $R_h(0) = 0$). On the other hand, $I_h(0)$ will be the initial number of infected people. We assumed $I_h(0) = 1$, which is, according to the data, the first human case detected in November 2015. For the compartment for mosquitoes population, we assumed a total of 8.3 million susceptible mosquitoes (which is approximately 2.24 mosquitoes

per human [67]) and 2.24 infected mosquitoes, at time $t = 0$. Therefore, to run the mathematical model, we used the vector of initial conditions that are given by:

$$[S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), E_v(0), I_v(0)] = [3699999, 0, 1, 0, 8287997.76, 0, 2.24], \tag{5.2}$$

where $S_h(0) + E_h(0) + I_h(0) + R_h(0) = 3700000$ and $S_v(0) + E_v(0) + I_v(0) = 8288000$.

On the other hand, for the parameters in the model, we used the mean value of the probability distributions and results provided by Kucharski et al. in [40]. Table 5–1 shows the mean value of the probability distributions for the transmission rates, incubation periods, infectious period, and mosquito lifespan, in which Kucharski et al. assigned a Gamma distribution according to their literature review. First, the mean value of the probability distribution defined by Kucharski et al. will be use to run the mathematical model and to estimate the parameters. Later in this chapter, a discussion about the probability distributions of each parameter will be presented. For the human-to-human transmission rate (κ), we assumed a value according to the work presented by Towers et al. [72], in which a transmission rate between humans correspond to a low value in comparison with β_v and β_h.

Parameter	**Baseline (weeks^{-1})** [40, 72]	**Range (weeks^{-1})** [40, 72]
α_v	0.67	[0.58 - 3.50]
α_h	1.19	[0.47 - 1.75]
γ	1.40	[1.00 - 2.33]
δ	0.90	[0.35 - 1.17]
β_v	1.45	[0.70 - 14.00]
β_h	4.85	[0.70 - 16.80]
κ	0.39	[0.007 - 0.70]

Table 5–1: Chosen values for the vector of parameters, to run the mathematical model in System 2.9.

The vector of parameters we used, in weeks^{-1}, to run the model, is given by:

$$[\alpha_v, \alpha_h, \gamma, \delta, \beta_v, \beta_h, \kappa] = [0.67, 1.19, 1.40, 0.90, 1.45, 4.85, 0.39] \tag{5.3}$$

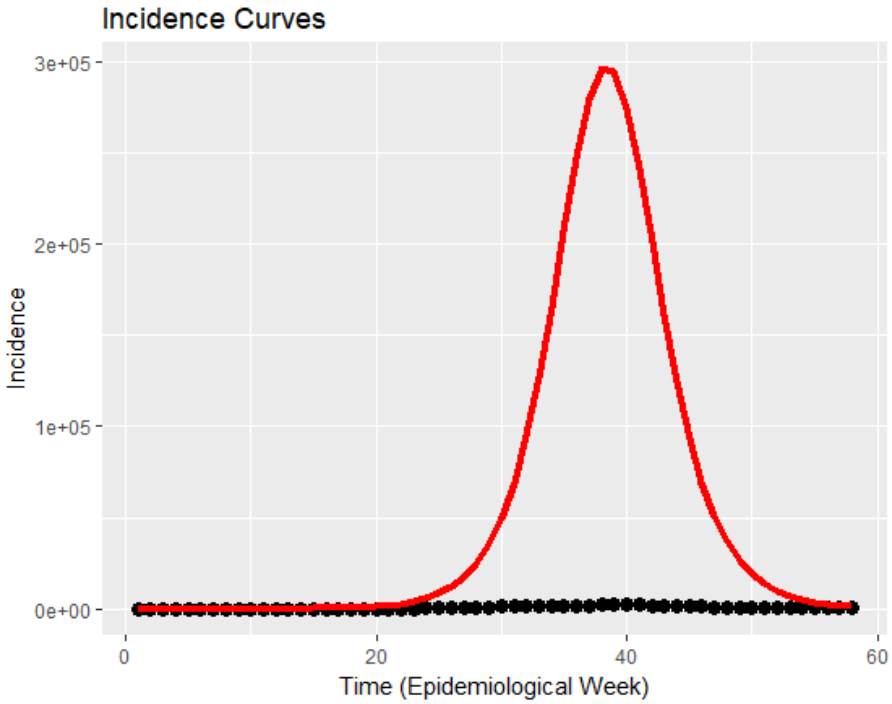


Figure 5–1: The red line represent the time series for the weekly incidence as calculated by the mathematical model in System 2.9 and the black dots are the Zika weekly incidence reported by Puerto Rico Department of Health (our data set).

The intention with this chapter is to calculate the reporting rate in order to obtain a better fit of the mathematical model incidence curve to the available data. The analysis on this chapter is divided in three sections. First, a Markov Chain Monte Carlo (MCMC) simulation will be perform to generate a posterior distribution of the reporting rate, knowing that the Zika weekly incidence data follows a Negative Binomial distribution, as described in Subsection 4.1.1. In the second section, we will estimate the transmission rates β_h and κ, from the mathematical model (System 2.9), that provide the best fit of the simulated data to the observed data (ZIKV weekly incidence in Puerto Rico), by performing an optimization process.

Finally, a sampling process will be use to generate a probability distribution for the basic reproductive number, by using the expression for R_0 given by

$$\begin{aligned} R_0 &= R0_{vector} + R0_{sex} \\ &= \beta_v \beta_h \frac{\alpha_v}{\delta\gamma(\alpha_v + \delta)} + \frac{\kappa}{\gamma}, \end{aligned} \tag{5.4}$$

where the parameters are presented in Table 5–1.

5.1 Estimation of the Reporting Rate

A Markov Chain Monte Carlo (MCMC) simulation to obtain the posterior distribution of the reporting rate, r, according to Kucharski et al. [40], we can assume a Unif(0,1) distribution as a prior for r. Since the reporting rate can be interpreted as a ratio between the observed data (Zika weekly incidence data) and the predicted incidence data (from System 2.9), then the estimate of r will range between 0 and 1. From Figure 5–1, we can observe that the weekly incidence values in our data set are lower than the predicted values. For the MCMC simulation, we run 8 replicates of 25000 iterations, each one with a burn-in period of 5,000 iterations by using "rjags" package [56] from the statistical software R [70]. The results of this process are presented in Figure 5–2, which is an output from the statistical software R [70] and the posterior distribution for the reporting rate is in Figure 5–3.

From the output in Figure 5–2, we can conclude that the weekly incidence data for the ZIKV epidemic in Puerto Rico represent approximately 2% [1.7%, 2.3%] of the true incidence data set from the simulation of the mathematical model. Now, by taking 0.02 times the predicted data and plotting it against the observed data (our data set), we obtained the graph presented in Figure 5–4. The estimated value for the reporting rate can be extracted from the posterior distribution, presented in Figure 5–3, by taking the mean of the histogram. The mean correspond to the value $r = 0.02$.

```
Iterations = 10001:35000
Thinning interval = 1
Number of chains = 8
Sample size per chain = 25000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

          Mean             SD       Naive SE Time-series SE
     2.009e-02      1.500e-03      3.355e-06      4.404e-06

2. Quantiles for each variable:

   2.5%     25%     50%     75%   97.5%
0.01738 0.01905 0.02000 0.02103 0.02328
```

Figure 5–2: Output from the statistical software R showing the mean estimate of the reporting rate for the Zika weekly incidence data from Puerto Rico.

A simple way to approximate the reporting rate (or confirmed ratio as defined by Towers et al. [71]) is by taking the sum of the observed data (Zika weekly incidence data in Puerto Rico) and divide it by the sum of the predicted incidence data. We know, from Section 4.1, that in Puerto Rico we had a total of 37,500 infected people (with ZIKV) at the end of 2016, while the predicted incidence curve indicates a total of 3,313,279 infected people. Then, by computing the division between 37,500 and 3,313,279, we obtained 0.011 as a point estimate for the reporting rate or confirmed ratio. This latest computation help us to support the procedure performed by the MCMC simulation, even when the curves shows a gap between them. It is important to remark that multiply the estimated reporting rate by the predicted incidence data is not an optimization process and it does not achieve any fitted curve.

It is important to mention that the current section is yet under study , and therefore, needs a more deeper analysis. In consequent, the fitting procedure will be performed using the partial results of the reporting rate.

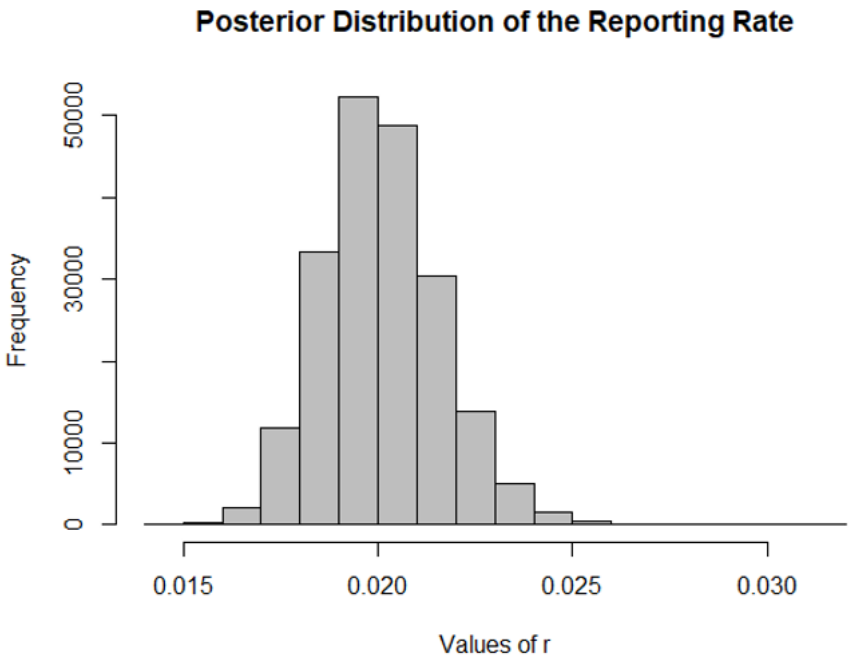


Figure 5–3: Posterior distribution of the reporting rate.

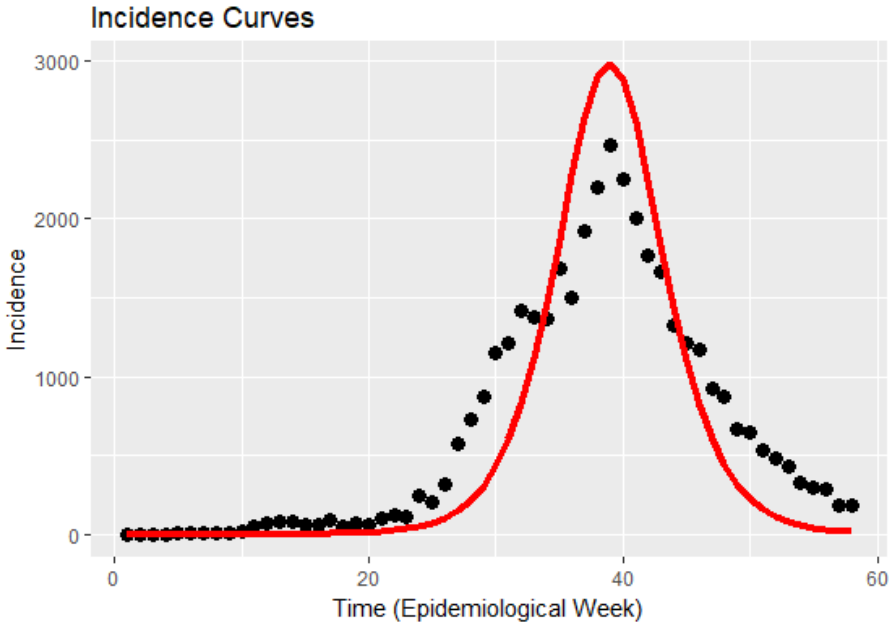


Figure 5–4: Incidence curves after we adjust using the estimate of the reporting rate (2%). Red line is the adjusted simulated data and the black dots are the observed data from Puerto Rico.

5.2 Fitting the parameters of an SEIR/SEI model to the ZIKV data

In this section, we cover the estimation of the transmission rates β_h and κ, by using the observed data and a predicted data that will be produced from the mathematical model (System 2.9). Since our data set, as presented in Section 4.1, correspond to weekly incidence, the predicted data also needs to be the same type of data. In order to obtain the predicted weekly incidence data, from the mathematical model in System 2.9, $f(t_j;\vec{\theta})$ (note that these predictions depend on the model parameters and initial conditions), we needed to add a new ordinary differential equation that help us to keep the count of the weekly cumulative incidence for the ZIKV epidemic in Puerto Rico, and eventually to compute $f(t_j;\vec{\theta})$.

The new ordinary differential equation is given by

$$\frac{dC}{dt} = \alpha_h E_h(t), \tag{5.5}$$

which it was added only as part of the simulation in the statistical software R [70], but is not part of the mathametical model as defined by Brauer et al. [7]. Therefore, the predicted values can be computed by the difference between the cumulative incidence in week j and the cumulative incidence at week $j-1$,

$$f(t_j;\vec{\theta}) = C(j) - C(j-1). \tag{5.6}$$

In most of the cases, the Ordinary Least Squares (OLS) scheme is employed assuming that the variances associated with the epidemic observations were longitudinally constant and not dependent on the values of the observations [38]. Also, according to Khan et al. [38], it is known that the assumption of constant variance may not be realistic, because the epidemic data could be influenced by a particular event, such an under-reporting process as described in [38].

In our case, we have the under-reporting problem that it is affecting the count of infected people with the Zika virus, leading us to choose the Generalized Least Squares (GLS) method as the optimization process, to estimate κ and β_h, and reject the implementation of the OLS method. Later in this section, a residual analysis from both methods (OLS and GLS) will be presented, to confirm our choice. For information about OLS and GLS methods, see Chapter 3.

According to Banks et al. [4] and Cintron et al. [16], the implementation of the GLS method requires the definition of a weight that will be included as part of the minimization process. In order to estimate the parameters through this method, we need to minimize the square of the point-by-point distances between the predicted values and the observed data.

As defined by Cintron et al. in [16] (see Chapter 3), the GLS estimator for a set of observations $\vec{Y} = (\vec{Y}_1, \vec{Y}_2, \ldots, \vec{Y}_N)$ is the solution of the normal equations

$$\sum_{j=1}^{N} W_j \cdot \left[\vec{Y}_j - f(t_j; \vec{\theta})\right]^T \nabla_{\vec{\theta}} f(t_j; \vec{\theta}) = 0, \tag{5.7}$$

where the W_j defined for our optimization process are a set of non-negative weights (assuming $\omega = 1/2$ in Equation 3.29), are given by:

$$W_j = \frac{1}{f(t_j; \vec{\theta})}. \tag{5.8}$$

As part of the optimization process, we need to define the vector of fixed parameters, which is given by

$$[\alpha_v, \alpha_h, \gamma, \delta, \beta_v] = [0.67, 1.19, 1.40, 0.90, 1.45], \tag{5.9}$$

the vector of parameters to optimize, given by

$$\theta = [\beta_h, \kappa]^T, \tag{5.10}$$

and the vector of initial conditions, given by

$$[S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), E_v(0), I_v(0)] = [3699999, 0, 1, 0, 8287997.76, 0, 2.24]. \tag{5.11}$$

Since the observation process is affected by the under-reporting problem, the GLS method will be considered as the first attempt to estimate the parameters in θ, we performed the GLS method, by using the same vector of initial conditions as presented in Equation 5.11, but we noticed that our estimates did not converge to a reasonable values according to the literature, and shows some biological disagreements, for example, negative values for the parameters or estimates out of range. This results comes when we tried to optimize the vector of parameters $\theta = (\beta_h, \kappa)^T$. After several subsequent tries in the optimization, we finally noticed that the problem was in the initial number of susceptible human. This problem was pointed by Shutt et al. [67], when they said that simulations performed by using the entire country population as the number of initially susceptible humans mischaracterized the disease dynamics, leading to over estimates in the final size of an epidemic. We compared our graphs to those presented in their work, and they obtained similar graphs to our plot in Figure 5–1, where the observed values are lower than the predicted values by the model. We tried to truncate the observed data, to change the baseline for the parameters to optimize, shifted the initial conditions, and even considering those scenarios, always obtained estimates that are not reasonable according to the epidemiological interpretation.

The next attempt was to reduce the initial susceptible population by taking the product of our estimate of the reporting rate ($\hat{r} = 0.02$) with the total population of Puerto Rico (3.7 million), obtaining 74,000 people. After running the optimization process, we solved the problem about negative estimates and we fix the problem of estimates out of range, but the fitted curve was not explaining the dynamics adequately.

Following the idea presented by Shutt et al. in [67], they suggest to estimate the number of people that is needed to be in $S_h(0)$ such that the convergence occurs without any biological (or mathematical) disagreements. They assumed that Dengue and Zika occur in the same areas, sharing a common vector and with similar asymptomatic rates. This assumption lead them to the calculation of a reasonable population value rather than fitting the initial susceptible population to the data as a parameter. They calculated the at-risk population size per country for a Zika outbreak based on historical data for dengue, but we addressed this issue in a different way. In the work presented by Cintron et al. [16], they performed a statistical analysis (OLS and GLS method) that includes the initial susceptible population ($S_h(0)$) as another parameter to optimize, but here in this work, a similar approach to those presented by Cintron et al. [16] was employed, to estimate the transmission rates β_h and κ. The most intrigue questions about this optimization process was: Why we did not get reasonable estimates by using the total population of Puerto Rico as the initial susceptible population? What is the biological explanation for this? Or there is a mathematical problem involved here? and why we cannot achieve a good fitted curve?

The solution of this analysis (the optimization process) took us some time to solve. Since the optimization achieve convergence with a low value for $S_h(0)$, but we do not know how low this value needs to be, we prefer to define a new parameter, ψ, which will allow us to measure the proportion of people that needed to be included in the model to obtain the best fitted curve. But, this new parameter need to be linked with the mathematical model in a biological sense.

In a work presented by Shresta et al. [66] in November 2017, they suggest that the reason that some people infected with Zika do not come down with the disease is due to prior exposure to Dengue. According to Lalita et al. [59], the ZIKV shares a high degree of sequence and structural homology compared with other flaviviruses,

including Dengue virus (DENV), resulting in immunological cross-reactivity. This might be the answer to explain why the susceptible population cannot be equal to the Puerto Rico total population. According to the Center of Disease Control (CDC), Puerto Rico has faced various epidemic dengue activities since 1963 [11], and in accordance with Shresta et al. [66], the prior exposure to Dengue would reduce the number of people that are vulnerable (or not immune) to the Zika virus, allowing us to select a lower value for the initial condition $S_h(0)$.

In addition to the work presented by Shresta et al. [66], we also reviewed other articles that widely discuss about the cross-reactivity between Dengue and Zika, and some studies about the development of a vaccine that will target both viruses, ZIKV and DENV. See references [17, 32, 52, 58, 59, 63, 66] for further information.

After considering this new information, we made a few changes. Now, for the optimization process, the vector of parameters is given by:

$$\theta = [\beta_h, \kappa, \psi]^T, \tag{5.12}$$

and the vector of initial conditions will be similar to those in Equation 5.11, but now we have that the susceptible compartments will be given by

$$\begin{aligned} S_h(0) &= 3700000 \cdot \psi \\ S_v(0) &= 2.24 \cdot S_h(0) \end{aligned} \tag{5.13}$$

where ψ is parameter to optimize, and it will be in charge of measuring the proportion of people that are not immune, which is, the people that did not had a previous exposure to either of both viruses, ZIKV and DENV.

The optimization process was implemented by using a direct search method, in our case, we decided to choose the Nelder-Mead simplex algorithm, as recommended by Cintron et al. [16]. We used the function *optim*() in R and the argument "method" was defined to be the Nelder-Mead algorithm.

The baseline and estimates for each parameters in θ are presented in Table 5–2. The implementation of the GLS method in R, through the *optim*() function, gives the option to obtain the information of the Hessian matrix. Since we are performing a minimization process, then the covariance matrix of the estimates will be approximate (asymptotically) to the inverse of the Hessian matrix [1]. The standard errors are the square roots of the diagonal elements of the covariance matrix. To compute the confidence intervals for the estimates, $\hat{\theta}_{GLS}$, the normal confidence intervals (Wald confidence intervals) will be use, that are based on the asymptotic normality of the parameter estimators [1].

The 95% confidence interval for $\hat{\theta}_{GLS}$ is given by

$$\hat{\theta}_{GLS} \pm Z_{0.975} \cdot \hat{SE}(\hat{\theta}_{GLS})$$

After the iterations were completed, we found that the GLS estimates for the transmission rates, in weeks^{-1}, are $\hat{\beta}_h = 2.85$ [2.65, 3.04] 95% CI, $\hat{\kappa} = 0.63$ [0.53, 0.73] 95% CI and the proportion of non-immune population was estimated to be $\hat{\psi} = 0.0140$ [0.0138, 0.0142] 95% CI.

Parameter	**Baseline [67, 72]**	$\hat{\theta}_{GLS}$	**Standard Error**
β_h	4.85	2.85	0.0984
κ	0.39	0.63	0.0479
ψ	1	0.0140	0.0001

Table 5–2: GLS estimates of the transmission rates (weeks^{-1}) and the proportion of non-immune population, from the mathematical model with vector and sexual transmission.

According to Table 5–2, $\hat{\psi} = 0.0140$, therefore, by using Equation 5.13, we estimated that

$$\hat{S}_h(0) = 3700000 \cdot 0.0140 = 51800$$

and

$$\hat{S}_v(0) = 2.24 \cdot 51800 = 116032.$$

Then, the vector of initial conditions is given by

$$[\hat{S}_h(0), E_h(0), I_h(0), R_h(0), \hat{S}_v(0), E_v(0), I_v(0)] = [51799, 0, 1, 0, 116029.8, 0, 2.24]. \quad (5.14)$$

The mathematical model (System 2.9) will be run again, by using the GLS estimates, $\hat{\theta}_{GLS}$, provided in Table 5–2, with the rest of parameters in Equation 5.9. Since the third parameter in $\hat{\theta}_{GLS}$ correspond to the proportion of non-immune population, these value will cause an effect in the vector of initial conditions, allowing to estimate a value for $S_h(0)$.

Finally, we have all of the necessary tools to produce the best fitted curve to the observed data. In Figure 5–5, you will observe the predicted weekly incidence curve that best fit the observed data, through the Generalized Least Squares (GLS) method as an optimization process, for the unknown parameters in the mathematical model (β_h, κ).

In addition, a residuals analysis was performed to ensure that the selection of the GLS method is appropriate. Figure 5–6 shows the residuals obtained from the OLS method while Figure 5–7 presents the residuals through the GLS method. According to Agresti [1], if the residuals plot shows a random pattern, it suggests that the model fits the data well. On the other hand, if a non-random pattern is evident in the residuals, then the model fits the data poorly. In this case, the residuals from the OLS method displays a non-random structure, indicating that OLS method is not the correct model to explain the data. From this analysis, we support the choice we made about GLS method as the optimization procedure.

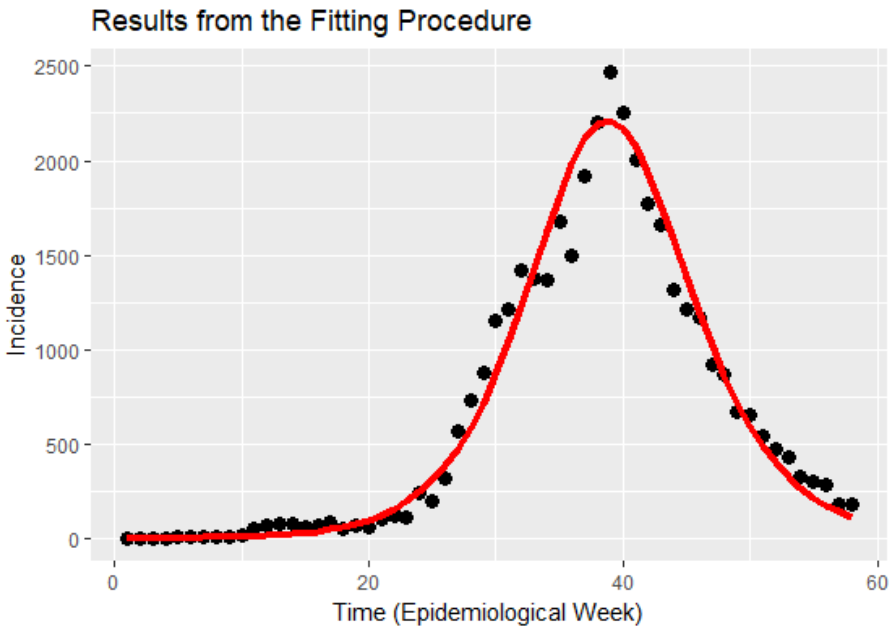


Figure 5–5: The red line indicate the best fitted curve, while the black dots represents the observed data (weekly incidence data) for 2015-2016 Zika epidemic in Puerto Rico.

The next step is to use the GLS estimates, $\hat{\theta}_{GLS}$, provided in Table 5–2, with the rest of parameters in Equation 5.9, to compute the basic reproductive number (R_0) for the 2015-2016 ZIKV epidemic in Puerto Rico, by using Equation 2.10. We found that

$$\begin{aligned} R_0 &= R0_{vector} + R0_{sex} \\ &= 1.40 + 0.45 \\ &= 1.85, \end{aligned} \tag{5.15}$$

which is similar to those estimates of R_0 from previous chapter. According 5.15, we can support that the Zika virus will not produce an epidemic if the only transmission is from human to human, and this is because $R0_{sex} < 1$. On the other hand, if the only transmission is through the vector, an epidemic will occur because we found that $R0_{vector} > 1$. In general, we can say that in average, one infected individual can produce an approximate of two more infected people.

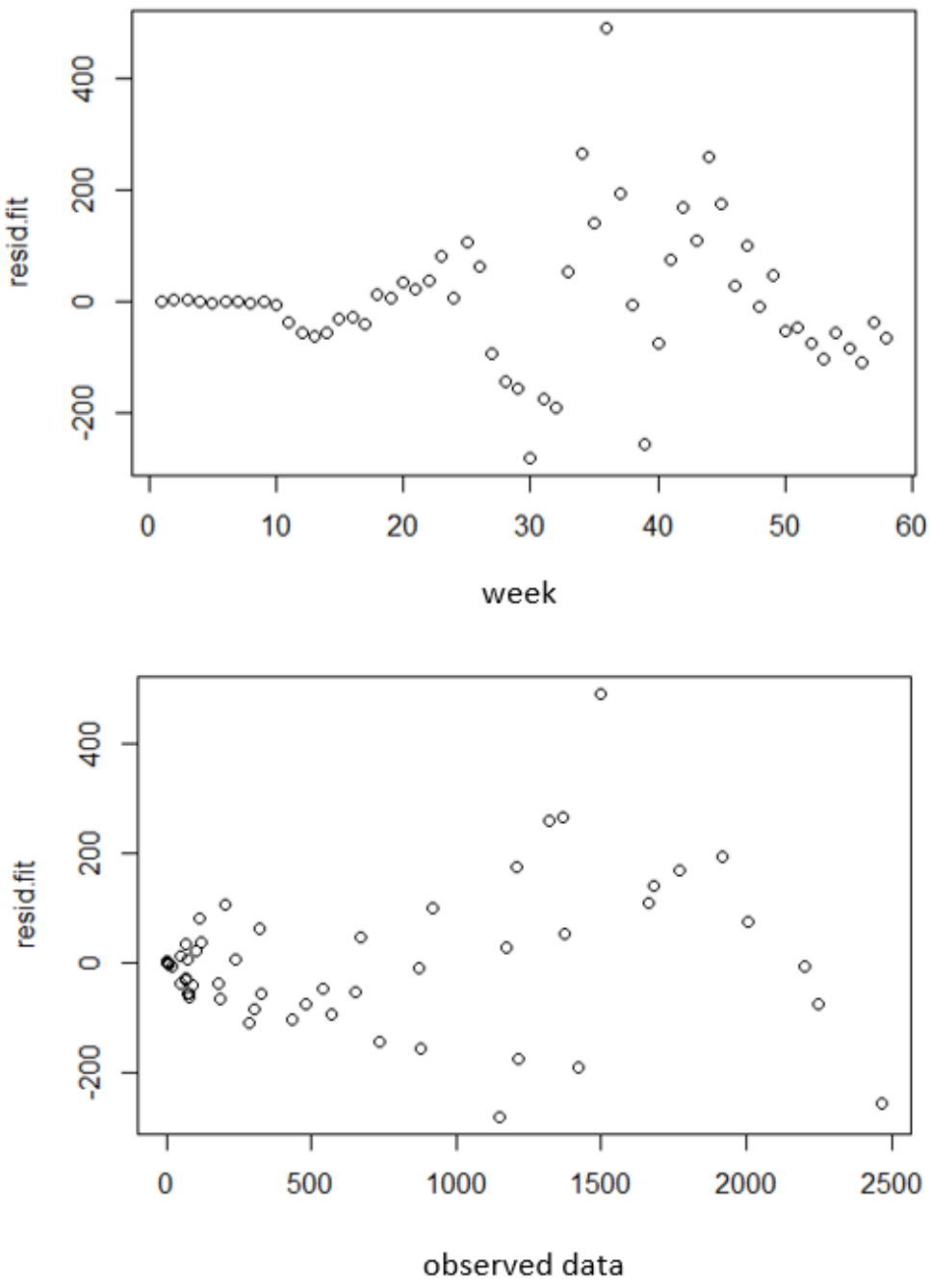


Figure 5–6: Residual plots for the Zika incidence data through the OLS method.

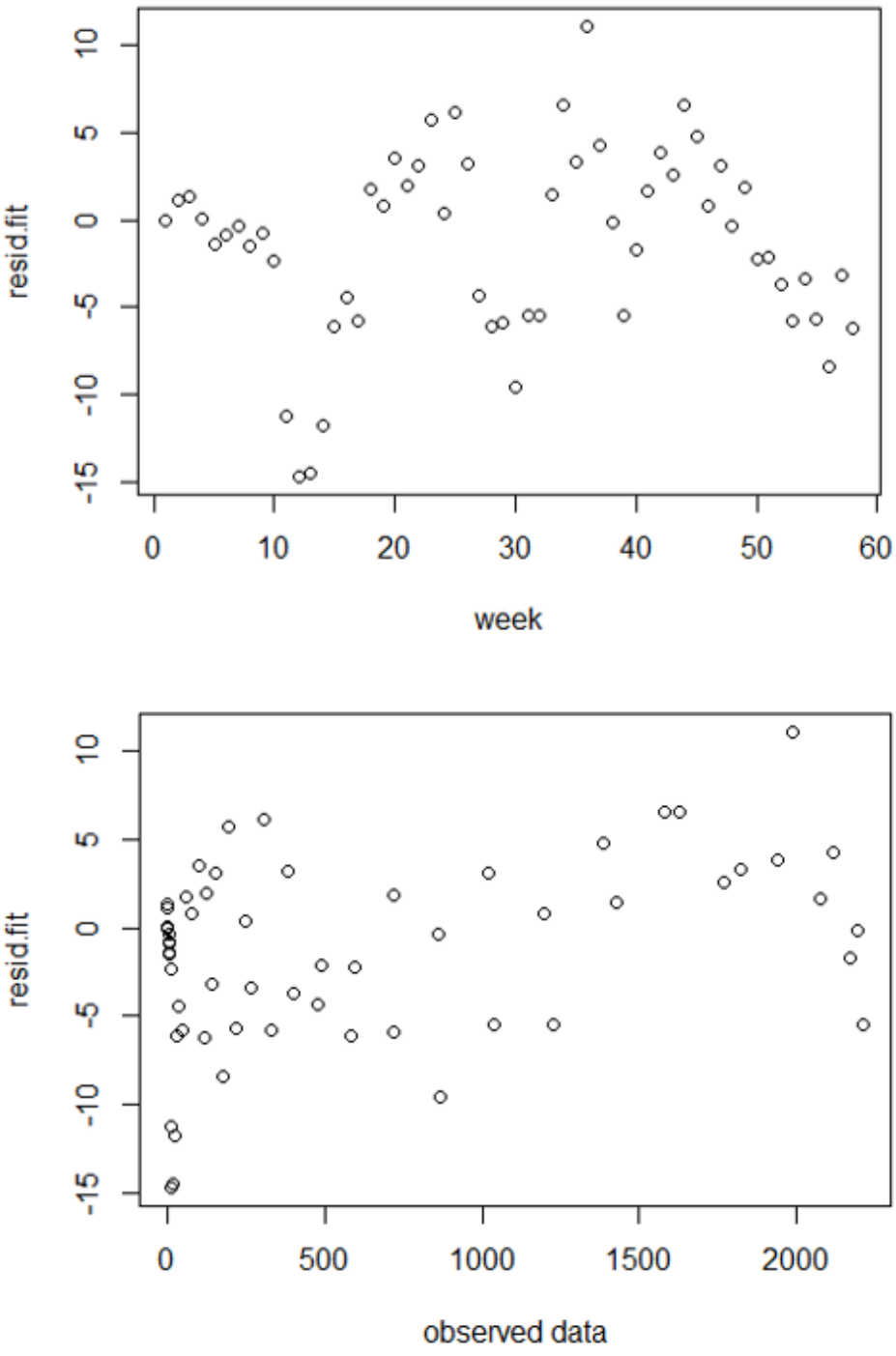


Figure 5–7: Residual plots for the Zika incidence data through the GLS method.

5.3 Estimation of R_0

The last analysis of this work is a sampling process, to generate a probability distribution of the basic reproductive number (R_0), by using distributions of the parameters in the mathematical model, which is a similar procedure as presented in Chapter 5. From the work of Kucharski et al. [40], Shutt et al. [67], and Towers et al. [72], three scenarios for the probability distributions of the parameters can be extracted. First, a Uniform distribution is used taking into consideration the range of values of each parameter. In the second scenario, a Gamma distribution is generated, also by considering the range of values of the parameters. Finally, a Log-Normal distribution is selected, where the mean is the logarithm of the baseline value presented in Table 5–1. In the three scenarios, the transmission rates have a Normal distribution, where GLS estimates of β_h and κ are the means of the distributions and we used their respective standard errors for the standard deviations. For the probability distribution of β_v, the baseline vale is chosen, from Table 5–1, as the mean and a standard deviation of 0.1, defined accordingly to the other resulted standard error from the GLS method. A summary for the three scenarios are presented in Figure 5–8.

The sampling process need to obey the constraint that the serial interval derived from the sampled parameters had to be within the observed serial interval of 10 to 23 days, or 10/7 to 23/7 weeks, in our case [46]. After the 10,000 iterations, the probability distribution of $R0_{model}$ for each scenario are presented in Figure 5–9, where the mean, median, and credible interval are computed.

If the descriptive statistics included in Figure 5–9 is observed, we can conclude that the scenario A and B throw similar results, but are quite far from the estimated values from previous chapter. From this two scenarios, the resulted estimates for the basic reproductive number are $\bar{R}_0 = 2.67$ [1.09, 6.28] and $\bar{R}_0 = 2.67$ [1.02, 5.86], respectively. The last scenario results on a estimate of $\bar{R}_0 = 1.85$ [1.51, 2.23],

which is similar with the results of previous chapter. It is clear that these scheme of plugging and playing with the parameters distribution needs to be considered in a more deeper analysis, but it will be leave as a future work.

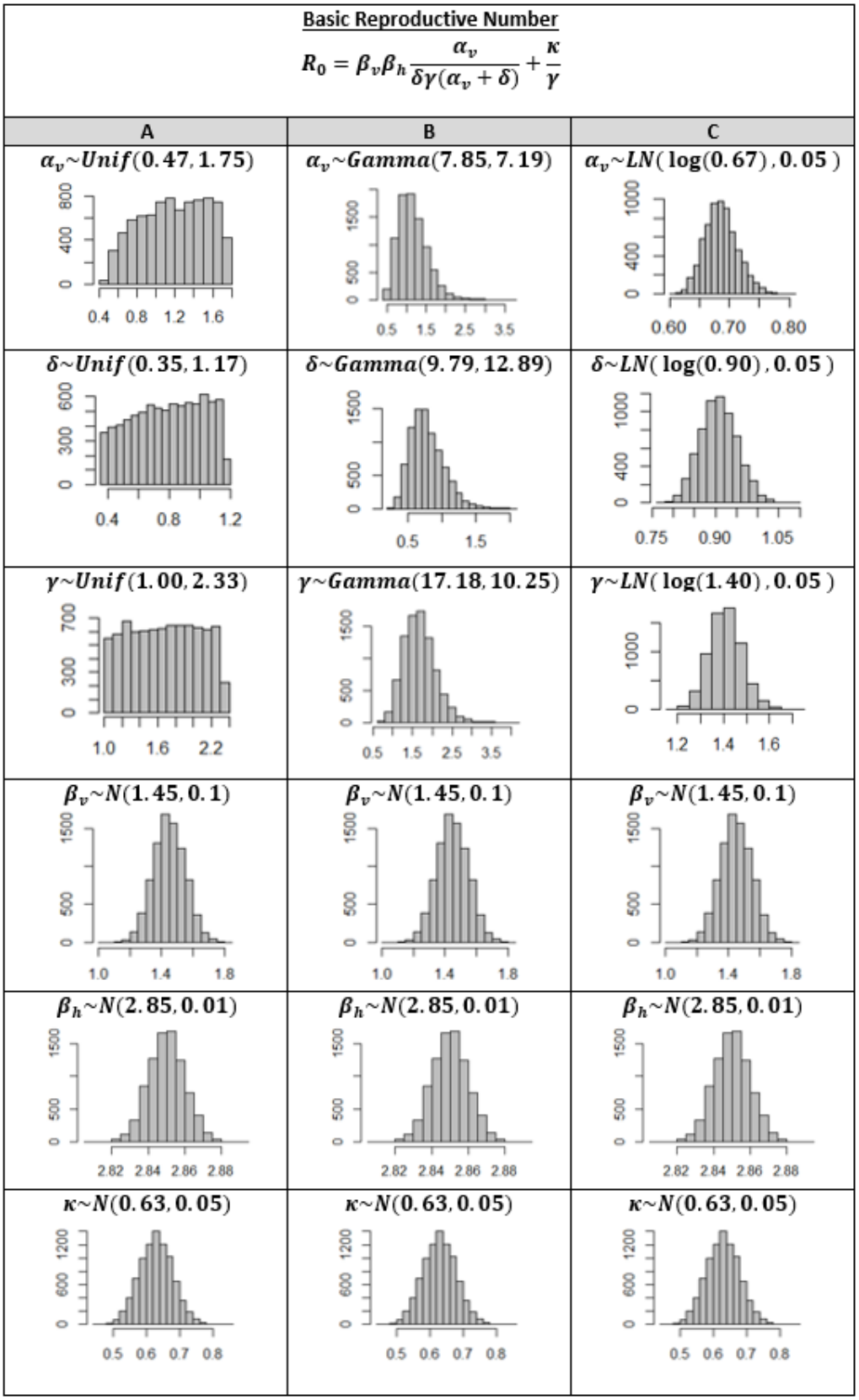


Figure 5–8: Probability distribution for the parameters in the mathematical model from System 2.9. Three different scenarios for the simulation were considered (A, B and C).

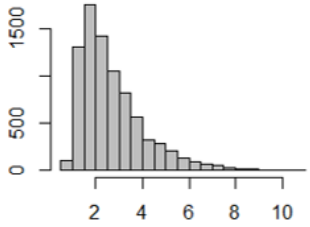
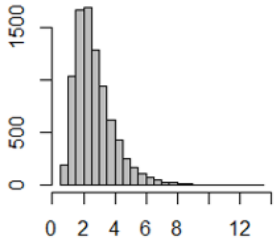
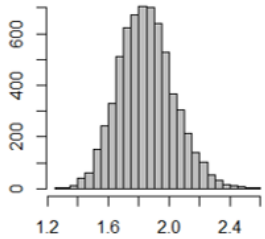
Results	A	B	C
Probability distribution for R_0			
Mean	**2.67**	**2.68**	**1.85**
Median	**2.31**	**2.40**	**1.84**
95% CR	**[1.09, 6.28]**	**[1.02, 5.86]**	**[1.51, 2.23]**

Figure 5–9: Probability distribution of the basic reproductive number ($R0_{model}$). The distributions comes from the sampling process, following the three parameter scenarios from Figure 5–8. In these scenarios (A, B and C), the transmission rates follows a Normal distribution, while the parameters α_v, δ and γ follows different distributions: (A) Uniform distribution, (B) Gamma distribution, and (C) Log-Normal distribution.

CHAPTER 6
CONCLUSION AND FUTURE WORK

Every time that an infectious disease is impacting a population, several analysis need to be performed in order to understand the behavior of the disease, so we can be able to control it and eventually, eradicate it from the population. One of those analysis is the estimation of the basic reproductive number (R_0), which is used to measure the transmission potential of a disease, properly defined as the average number of secondary infections produced by a typical case of an infection in a population that is totally susceptible [65].

The mission of this work was on the estimation of the basic reproduction number (R_0) by various methods including the initial exponential growth rate (ρ), considering the Zika outbreak occurred in Puerto Rico from November 2015 to December 2016. Since the Zika virus was an emergent disease in the island, our estimates, as we know, are the first estimates of this epidemiological parameters. In order to obtain the estimates of ρ and R_0, a mathematical model of ordinary differential equations was adopted, provided by Brauer et al. [7] and the expression for R_0 comes from the analysis of Towers et al. [72]. The principal contribution of this work is the estimate of the initial exponential growth rate and the basic reproductive number through a Bayesian approach, for the Zika outbreak in Puerto Rico, assuming that the weekly incidence data follows a Negative Binomial distribution.

From the different analysis in this work, a summary of the estimates are presented in Table 6–1, according to the methods covered. In addition, we have proved

that our data set represent only the 2% of the true incidence data, by analyzing the reporting rate, also through a Bayesian approach.

According to the statistical methods that we employed to estimate ρ and R_0, the method stablished by Favier et al. [27] is recommended to select the exponential phase of a epidemic, while the Bayesian approach leads to a more precised estimation of ρ. The estimates for the initial exponential growth rate ranged from 0.23 to 0.29 (see Chapter 4). For the 2015 Zika oubtreak in Barranquilla, Colombia, Towers et al. [72] estimated ρ to be 0.071 [0.055,0.089] days^{-1} (or 0.497 weeks^{-1}). In the 2013-14 Zika oubtreak ocurred in French Polynesia, Kucharski et al. [40] estimated this parameter to be around the values 0.075 and 0.20 weeks^{-1} (see [40] to know the estimates for each region of French Polynesia).

The estimates for the basic reproductive number ranged from 1.67 to 1.91, considering an expression for R_0 in terms of ρ, where the discussion of each procedure are presented in Chapter 4. The estimates of R_0 through a Inverse Problem ranged between 1.51 and 2.23, which is the scenario where a Log-Normal distribution was assigned to α_v, δ and γ, and a Normal distribution for β_h, β_v and κ. The Zika outbreak in Colombia was estimated to be 3.8 [2.4, 5.6] 95% CI, while the median estimates from the regions in French Polynesia ranged from 2.6 to 4.8.

For future works, it will be appropriate to re-estimate the values of the initial exponential growth rate (ρ) and the basic reproductive number (R_0), by using a real weekly incidence data, which are the data from Puerto Rico Department of Health, instead of using approximations as we used on this work. Since the scenario of a cross-reactivity (or cross-protection) between the Dengue and Zika virus is something that has been in study recently, a deeper analysis of a mathematical model that includes this information is highly recommended. Therefore, a new computation of the R_0 will be necessary, considering real data and a more complete dynamic model

about the Zika virus in Puerto Rico or other populations, where Zika and Dengue were circulating simultaneously.

Method	End of the Exponential Phase	$\widehat{\rho}$ (weeks^{-1}) 95%CI or 95%CR		$\widehat{R_0}$ 95%CI or 95%CR	
		Non-Bayesian	Bayesian	Monte Carlo Simulation	
Visual	Week 32	0.26 [0.23, 0.29]	0.24 [0.23, 0.26]	1.80 [1.40, 2.38]	1.72 [1.36, 2.23]
	Week 35	0.25 [0.23, 0.28]	0.24 [0.23, 0.25]	1.76 [1.38, 2.31]	1.70 [1.35, 2.19]
	Week 39	0.24 [0.22, 0.26]	0.23 [0.22, 0.24]	1.72 [1.36, 2.23]	1.67 [1.34, 2.14]
Favier	Between week 33 and 34	0.29 [0.27, 0.35]		1.91 [1.45, 2.60]	
Chowell	Week 31		0.24 [0.23, 0.26]		1.73 [1.37, 2.24]
Inverse Problem Estimation				(A) Mean: 2.67 Median: 2.31	[1.09, 6.28]
				(B) Mean: 2.68 Median: 2.40	[1.02, 5.86]
				(C) Mean: 1.85 Median: 1.84	[1.51, 2.23]

Figure 6–1: A summary of the resulted estimates from this work. The rows identified by Visual, Favier, and Chowell were the methods used for the estimation of R_0, with and without the initial exponential growth rate (ρ). In the Inverse Problem Estimation, the estimates of R_0 comes from the sampling process, following the three parameter scenarios from Figure 5–8. In these scenarios (A, B and C), the transmission rates follows a Normal distribution, while the parameters α_v, δ and γ follows different distributions: (A) Uniform distribution, (B) Gamma distribution, and (C) Log-Normal distribution.

REFERENCE LIST

[1] Agresti, A. *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons, Inc., Hoboken, New Jersey., 2015.

[2] Anderson, R. M., and May, R. M. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1991.

[3] Ayuso, I. R. R., Geerman, K., and Pesante, F. Puerto Rico Community Health Assesstment: Secondary Data Profile. *Puerto Rico Department of Health* (2012).

[4] Banks, H. T., Davidian, M., Samuels, J., and Sutton, K. An Inverse Problem Statistical Methodology Summary. *Mathematical and Statistical Estimation Approaches in Epidemiology* (2009). DOI 10.1007/978-90-481-2313-1 11.

[5] Bettencourt, L., and Ribeiro, R. Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Deseases. *PLOS One 3*, 5 (2008), e2185.

[6] Boyer, C., and Merzbach, U. *A History of Mathematics*. John Wiley & Sons, Inc., 2010. 3rd Edition.

[7] Brauer, F., Castillo-Chavez, C., Mubayi, A., and Towers, S. Some Models for Epidemics of Vector-Transmitted Diseases. *KeAi Advancing Research Evolving Science Infectious Disease Modelling 1* (2016).

[8] Burton, D. *The History of Mathematics*, 7th ed. The McGraw-Hill Company, Inc., 2011.

[9] Casella, G., and George, E. Explaining the Gibbs Sampler. *The American Statistician, by the American Statistical Association 46*, 3 (Aug. 1992), 167–174.

[10] CDC. CDC Concludes Zika Causes Microcephaly and Other Birth Defects. www.cdc.gov/media/releases/2016/s0413-zika-microcephaly.html.

[11] CDC. Dengue in Puerto Rico. https://www.cdc.gov/dengue/about/inpuerto.html.

[12] Chowell, G., Diaz-Dueas, P., Miller, J., Alcazar-Velazco, A., Hyman, J., Fenimore, P., and Castillo-Chavez, C. Estimation of the Reproduction Number of Dengue Fever from Spatial Epidemic Data. *Elsevier, Mathematical Biosciences. 208*, 2 (August 2007), 571–589.

[13] Chowell, G., Nishiura, H., and Bettencourt, L. M. A. Comparative Estimation of the Reproduction Number for Pandemic Influenza from Daily Case Notification Data. *Journal of The Royal Society: Interface* (2006), 155–166.

[14] Chowell, G., Torre, C., Vladimir, C., Mac, J., Castillo, C., Suarez, L., and Lopez, R. Spatial and Temporal Dynamics of Dengue Fever in Peru: 1994-2006. *Epidemiol. Infect. 136*, 12 (2008), 1667–77.

[15] Chowell, G., Towers, S., Viboud, C., Fuentes, R., Sotomayor, V., Simonsen, L., Miller, M., Lima, M., Villaroel, C., Chiu, M., Villaroel, J., and Olea, A. The Influence of Climatic Conditions on the Transmission Dynamics of the 2009 A/H1N1 Influenza Pandemic in Chile. *BioMedCentral Infectious Diseases* (2012).

[16] Cintron-Arias, A., Castillo-Chavez, C., Bettencourt, L., Lloyd, A., and Banks, H. The Estimation of the Effective Reproductive Number from Disease Outbreak Data. *Mathematical Biosciences and Engineering 6*, 2 (Apr. 2009), 261–282. doi:10.3934/mbe.2009.6.261.

[17] Collins, M. H., McGowan, E., Jadi, R., Young, E., Lopez, C. A., Baric, R. S., Lazear, H. M., and de Silva, A. M. Lack of Durable Cross-Neutralizing Antibodies against Zika Virus from Dengue

Virus Infection. *Emerging Infectious Diseases 23*, 5 (May 2017). DOI: http://dx.doi.org/10.3201/eid2305.161630.

[18] Congdon, P. *Bayesian Statistical Modelling*. Wiley Series in Probability and Statistics, 2006.

[19] CountryMeters. Country Meters of the World Population: Puerto Rico. http://countrymeters.info/es/Puerto_Rico.

[20] Cowan, G. *Statistical Data Analysis*. Oxford Science Publications, 1998.

[21] Degallier, N., Favier, C., Boulanger, J., Menkes, C., and Oliveira, C. Une nouvelle mthode destimation du taux de reproduction des maladies (R0) : application ltude des pidmies de dengue dans le district fdral (Brsil). *Environnement, Risques & Sant 4*, 2 (2005).

[22] Delignette-Muller, M., and Dutang. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software 64*, 4 (2015), 1–34.

[23] Dietz, K. The Estimation of the Basic Reproduction Number for Infectious Diseases. *Statistical Methods in Medical Research 2* (1993), 23–41.

[24] Duffy, M., Chen, T.-H., Hancock, T., Powers, A., Kool, J., Lanciotti, R., Pretrick, M., Marfel, M., Holzbauer, S., Dubray, C., Guillaumot, L., Griggs, A., Bel, M., Lambert, A., Laven, J., Kosoy, O., Panella, A., Biggerstaff, B., Fischer, M., and Hayes, E. Zika Virus Outbreak on Yap Island, Federated States of Micronesia. *The New England Journal of Medicine 360* (June 2009), 2536–2543.

[25] Dym, C. *Principles of Mathematical Modeling*, 2nd ed. Academic Press, 2004.

[26] Fauci, A., and Morens, D. Zika Virus in the Americas Yet Another Arbovirus Threat. *The New England Journal of Medicine 374* (2016), 601–604.

[27] Favier, C., Degallier, N., Rosa-Freitas, M., Boulanger, J., Lima, J. C., Luitgards-Moura, J., Menkes, C., Mondet, B., Oliveira, C.,

and E. Weimann, P. T. Early Determination of the Reproductive Number for Vector-Borne Diseases: The Case of Dengue in Brazil. *Tropical Medicine and International Health 2*, 4 (2006), 343–351.

[28] Fay, R. W. The Biology and Bionomics of Aedes Aegypti in the Laboratory. *Mosquito News 24*, 3 (1964), 300–308.

[29] Gamerman, D., and Lopes, H. F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. Chapman and Hall/CRC, 2006.

[30] Gao, D., Lou, Y., He, D., Porco, T., Kuang, Y., Chowell, G., and Ruan, S. Prevention and Control of Zika as a Mosquito-Borne and Sexually Transmitted Disease: A Mathematical Modeling Analysis. *Scientific Reports* (2016).

[31] Gelis, K., Rios, K., Santana, D., and Davila, S. Estimacin de Tasa de Transmisin para Dos Serotipos de Dengue con un Modelo Matemtico para las Epidemias de los Aos 2010 y 2015 en Puerto Rico. mathesis, Universidad de Puerto Rico, Recinto de Mayaguez, 2014.

[32] Grifoni, A., Pham, J., Sidney, J., ORourke, P. H., Paul, S., Peters, B., Martini, S. R., de Silva, A. D., Ricciardi, M. J., Magnani, D. M., Silveira, C. G. T., Maestri, A., Costa, P. R., de Oliveira-Pinto, L. M., de Azeredo, E. L., Damasco, P. V., Phillips, E., Mallal, S., de Silva, A. M., Collins, M., Durbin, A., Diehl, S. A., Cerpas, C., Balmaseda, A., Kuan, G., Coloma, J., Harris, E., James E. Crowe, J., Stone, M., Norris, P. J., Busch, M., Vivanco-Cid, H., Cox, J., Graham, B. S., Ledgerwood, J. E., Turtle, L., Solomon, T., Kallas, E. G., Watkins, D. I., Weiskop, D., and Settea, A. Prior Dengue Virus Exposure Shapes T Cell Immunity to Zika Virus in Humans. *Journal of Virology. Amercan Society for Microbiology 91*, 24 (Dec. 2017). https://doi.org/10.1128/JVI.01469-17.

[33] Halasa, Y. A., Shepard, D. S., and Zeng, W. Economic Cost of Dengue in Puerto Rico. *The American Society of Tropical Medicine and Hygiene* (2012).

[34] Heisey, D., Joly, D., and Messier, F. The Fitting of General Force of Infections Models to Wildlife Disease Prevalence Data. *Ecology* (2006), 2356–2365.

[35] Hilbe, J. M. *Negative Binomial Regression*, 2nd ed. Cambridge University Press, 2011.

[36] Holme, P., and Masuda, N. The Basic Reproduction Number as a Predictor for Epidemic outbreaks in Temporal Networks. *PLOS One 10*, 3 (2015).

[37] J.Faraway, J. *Linear Models with R.* Taylor & Francis e-Library, 2009.

[38] Khan, A., Hassan, M., and Imran, M. Estimating the Basic Reproduction Number for Single-Strain Dengue Fever Epidemics. *Infectious Diseases of Poverty 3:12* (2014). http://www.idpjournal.com/content/3/1/12.

[39] Khan, A., Naveed, M., e Ahmad, M. D., and Imran, M. Estimating the Basic Reproductive Ratio for the Ebola outbreak in Liberia and Sierra Leone. *Infectious Diseases of Poverty* (2015), 4–13.

[40] Kucharski, A., Funk, S., Eggo, R., Mallet, H.-P., Edmunds, W. J., and Nilles, E. Transmission Dynamics of Zika Virus in Island Populations: A Modelling Analysis of the 2013/14 French Polynesia Outbreak. *PLOS Neglected Tropical Diseases 10*, 5 (2016).

[41] Lipsitch, M., Cohen, T., Cooper, B., Robins, J., Ma, S., James, L., Gopalkrishna, G., Chew, S. K., Tan, C. C., Samore, M., Fisman, D., and Murray, M. Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science Magazine of American Association for the Advancement of Science 300* (2003).

[42] Lloyd, A. L. The Dependence of Viral Parameter Estimates on the Assumed Viral Life Cycle: Limitations of Studies of Viral Load Data. *The Royal Society*

(2000).

[43] Lloyd-Smith, J. O. Maximum Likelihood Estimation of the Negative Binomial Dispersion Parameter for Highly Overdispersed Data, with Applications to Infectious Diseases. *PLoS ONE* (2007).

[44] Lozano-Fuentes, S., Hayden, M. H., Welsh-Rodriguez, C., Ochoa-Martinez, C., Tapia-Santos, B., Kobylinski, K. C., Uejio, C. K., Zielinski-Gutierrez, E., Monache, L. D., Monaghan, A. J., Steinhoff, D. F., and Eisen, L. The Dengue Virus Mosquito Vector Aedes aegypti at High Elevation in Mexico. *The American Society of Tropical Medicine and Hygiene 87*, 5 (2012), 902–909.

[45] Lumerical. Parameter Sweeps, Optimization and Monte Carlo Analysis, 2003 to 2018. https://kb.lumerical.com/en/index.html?sweeps_optimizations_yields.html.

[46] Majumder, M., Cohn, E., Fish, D., and Brownstein, J. Estimating a Feasible Serial Interval Range for Zika Fever. *Bulletin of the World Health Organization* (2016).

[47] McCracken, M. K., Gromowski, G. D., Friberg, H. L., Lin, X., Abbink, P., Barrera, R. D. L., Eckles, K. H., Garver, L. S., Boyd, M., Jetton, D., Barouch, D. H., Wise, M. C., Lewis, B. S., Currier, J. R., Modjarrad, K., Milazzo, M., Michelle Liu and, A. B. M., Putnak, J. R., Michael, N. L., Jarman, R. G., and Thomas, S. J. Impact of Prior Flavivirus Immunity on Zika Virus Infection in Rhesus Macaques. *PLOS Pathogens 13*, 8 (2017). https://doi.org/10.1371/journal.ppat.1006487.

[48] Melo, D., Rios, K., Santana, D., and Steinberg, L. A Mathematical Model for Zika Virus Under Vector and Sexual Transmission and the Effect of Variable Infectivity on the Asymptomatic and Symptomatic Infected Human. Master's thesis, Department of Mathematical Science. University of Puerto Rico

at Mayaguez, 2017.

[49] Nowak, M., Lloyd, A., Vazquez, G., Wiltrout, T., Wahl, L., Bischofberger, N., Williams, J., Kinter, A., Fauci, A., Hirsch, V., and Lifson, J. Viral Dynamics of Primary Viremia and Antiretroviral Therapy in Simian Immunodeficiency Virus Infection. *Journal of Virology 71*, 10 (1997), 7518–7525.

[50] Ntzoufras, I. *Bayesian Modeling Using WinBUGS*. Wiley Series in Computational Statistics, 2009.

[51] Ott, R. L., and Longnecker, M. *An Introduction to Statistical Methods and Data Analysis*. Brooks Cole Cengage Learning, 2010.

[52] Pantoja, P., Perez-Guzman, E. X., Rodriguez, I. V., White, L. J., Gonzalez, O., Serrano, C., Giavedoni, L., Hodara, V., Cruz, L., Arana, T., Martinez, M. I., Hassert, M. A., Brien, J. D., Pinto, A. K., de Silva, A., and Sariol, C. A. Zika Virus Pathogenesis in Rhesus Macaques is Unaffected by Pre-Existing Immunity to Dengue Virus. *Nature Communications 8* (June 2017). DOI: 10.1038/ncomms15674.

[53] Petersen, E., Staples, J., Meaney-Delman, D., Fischer, M., Ellington, S., Callaghan, W., and Jamieson, D. Interim Guidelines for Pregnant Women During a Zika Virus Outbreak - United States, 2016. *MMWR: Morbidity and Mortality Weekly Report 65*, 2 (2016), 30–33.

[54] Plummer, M. *JAGS user manual*, Aug. 2013.

[55] Plummer, M., and Murrell, P. R News: Bayesian Inference and Markov Chain Monte Carlo (MCMC) simulation. *The Newsletter of the R Project 6*, 1 (Mar. 2006).

[56] Plummer, M., Stukalov, A., and Denwood, M. *Bayesian Graphical Models using MCMC: Package "rjags"*, Feb. 2016.

[57] PRDH. Vigilancia del Zika. http://www.salud.gov.pr.

[58] Priyamvada, L., Hudson, W., Ahmed, R., , and Wrammert, J. Humoral Cross-Reactivity between Zika and Dengue Viruses: Implications for Protection and Pathology. *Emerging Microbes & Infections* (2017). doi:10.1038/emi.2017.42.

[59] Priyamvadaa, L., Quicke, K. M., Hudson, W. H., Onlamoon, N., Sewatanon, J., Edupuganti, S., Pattanapanyasat, K., Chokephaibulkit, K., Mulligan, M. J., Wilsoni, P. C., Ahmeda, R., Suthar, M. S., and Wrammert, J. Human Antibody Responses after Dengue Virus Infection are Highly Cross-Reactive to Zika Virus. *PNAS: Immunology and Inflamation* (2016).

[60] Rao, R., and Toutenburg, H. *Linear Models: Least Squares and Alternatives*, 2nd ed. Springer, 1999.

[61] Rohatgi, A. Web Plot Digitizer: Web Based Tool to Extract Data from Plots, Images, and Maps, 2016. http://arohatgi.info/WebPlotDigitizer/.

[62] Ross, S. M. *Introduction to Probability Models*, 10th ed. Elsevier, 2010.

[63] Roth, C., Delgado, F. G., Simon-Lorire, E., and Sakuntabhai, A. Immune Responses to Dengue and Zika VirusesGuidance for T Cell Vaccine Development. *International Journal of Environmental Research and Public Health* (Feb. 2018). doi:10.3390/ijerph15020385.

[64] Rothman, K. *Epidemiology - An Introduction*. Oxford University Press, 2002.

[65] Rothman, K., Greenland, S., and Lash, T. *Modern Epidemiology*, 3rd ed. Lippincott Williams & Wilkins, 2008.

[66] Shresta, S., Wen, J., Ngono, A. E., Regla-Nava, J. A., Kim, K., Gorman, M. J., and Diamond, M. S. Dengue Virus-Reactive CD8+ T Cells Mediate Cross-Protection Against Subsequent Zika Virus Challenge. *Nature Communications 8*, 1459 (2017). DOI: 10.1038/s41467-017-01669-z.

[67] Shutt, D., Manore, C., Pankavich, S., Porter, A., and Valle, S. D.

Estimating the Reproductive Number, Total Outbreak Size, and Reporting Rates for Zika epidemic in South and Central America. *Cornell University Library. ArXiv* (2017).

[68] Soetaert, K., Petzoldt, T., and Setzer, W. Package deSolve: Solving Initial Value Differential Equations in R. *Journal of Statistical Software* (2010).

[69] Svensson, A. A Note on Generation Times in Epidemic Models. *Elsevier, Mathematical Biosciences 208*, 1 (July 2007), 300–311.

[70] Team, R. C. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria.* (2016).

[71] Towers, S. Fitting the Parameters of an SIR Model to Influenza Data Using Least Squares and the Monte Carlo Parameter Sweep Method, January 2013. http://sherrytowers.com/2013/01/29/neiu-lecture-vi-fitting-the-parameters-of-an-sir-model-to-influenza-data/.

[72] Towers, S., Brauer, F., Castillo-Chavez, C., Falconar, A., Mubayi, A., and Romero-Vivas, C. Estimate of the Reproduction Number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and Estimation of the Relative Role of Sexual Transmission. *Elseiver 17* (2016), 50–55.

[73] Vale, C. Diagram Collection: World Map Of Zika Cases. 2017. http://gitea.org/diagrams/world-map-of-zika-cases.html.

[74] van den Driessche, P., and Watmough, J. Reproduction Numbers and Sub-threshold Endemic Equilibria for Compartmental Models of Disease Transmission. *Mathematical Biosciences 180* (2002), 29–48.

[75] van Ravenzwaaij, D., Cassey, P., and Brown, S. D. A Simple Introduction to Markov Chain MonteCarlo Sampling. *Springer* (2016). DOI 10.3758/s13423-016-1015-8.

[76] Villela, D., Bastos, L., de Carvalho, L., Cruz, O., Gomes, M., Durovni, B., Lemos, M., Saraceni, V., Coelho, F., and Codeco,

C. Zika in Rio de Janeiro: Assessment of Basic Reproduction Number and Comparison with Dengue Outbreaks. *Epidemiology and Infection. Cambridge University Press 8* (2017).

[77] WHO. Zika virus. www.who.int/mediacentre/factsheets/zika/en/.

[78] WHO. Report of the Scientific Working Group Meeting on Dengue. *Scientific working group WHO* (2007).

[79] WHO. Health News: How the Zika Virus Spread Around the World, 2016. www.independent.co.uk/life-style/health-and-families/health-news/zika-how-virus-spread-around-world-a6839101.html.

[80] WHO. Fact sheet about Dengue and Severe Dengue, 2018. http://www.who.int/mediacentre/factsheets/fs117/en/.