# Statistical Analysis of the Chikungunya Fever

**Purvasha Chakravarti**

Carnegie Mellon University

Department of Statistics

Pittsburgh, PA 15213

## ABSTRACT

The first locally-acquired Chikungunya virus (CHIKV) infections in the Americas were reported in the Caribbean in December, 2013. Due to the recent emergence of the disease in the Americas, the current extent of spread and risk is uncertain. Understanding the spread of the disease is paramount for taking preventive measures, but challenging due to limitations of current data on the outbreak and on CHIKV transmission. We propose two models for modeling the number of infected cases of Chikungunya: a multi-country SIR model and a multi-country ARIMA model. The different models explain different aspects of the disease spread. While the former explains the rate of spread of the disease, the latter forecasts the number of new infected cases of Chikungunya.

Advanced Data Analysis Report

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at

Carnegie Mellon University

2015-2016

# Statistical Analysis of the Chikungunya Fever

Presented by

**Purvasha Chakravarti**[1,2]

Statistics Advisor — William F. Eddy[1,2]

External Advisor — Virginia Dato[3]

[1] Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213
[2] Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213
[3] Department of Biomedical Informatics, University of Pittsburgh, PA 15206

# Contents

# 1 Introduction

Chikungunya Fever is an emerging viral disease in the Americas, caused by the alphavirus, Chikungunya virus (CHIKV) and transmitted by mosquitos. The most common symptoms of Chikungunya are fever and joint pain; the joint pains are sometimes known to last for years. The fever may be accompanied with headache, muscle pain, joint swelling, or rash. Chikungunya has occurred in outbreaks of unprecedented magnitude in Asia, Africa, Europe and the Americas since 2004. The disease has affected approximately two million people, with some areas having attack rates as high as 68% [Roth et al., 2014].

The Chikungunya virus is transmitted to humans by the bite of infectious mosquitos, predominantly mosquitos of the Aedes genus; *Aedes aegypti* and *Aedes albopictus* [Lahariya and Pradhan, 2006]. The first indication of Chikungunya can be identified by the sudden onset of fever two to four days after exposure. The fever typically lasts for two to seven days and is usually accompanied by joint pains which typically last for weeks or months and sometimes for years. Sometimes there are other symptoms like muscle pain, head ache, nausea, fatigue and rash [Sourisseau et al., 2007]. Chikungunya has a mortality rate of little less than 1 in 1000. Usually the elderly, infants or those having underlying chronic medical problems having higher risk of complications [Mavalankar et al., 2008].

Chikungunya virus has an incubation period ranging from one to twelve days, and is most typically three to seven days [Thiberville et al., 2013]. That is, it takes typically three to seven days after the exposure of the disease, for an individual to show symptoms. The disease occurs in two stages. The first stage usually begins with a very high fever, usually above 102°C and sometimes reaching 104°C. The fever lasts

from a week to ten days, during which viremia occurs. However, other symptoms like headache and extreme exhaustion last for another five to seven days [Chhabra et al., 2008].

The second stage of the disease lasts for approximately ten days during which symptoms improve and the virus disappears from the blood. This is followed by strong joint pains and stiffness in muscles, which last for weeks but may last for years. Joint pain is reported by 87% to 98% of the patients and often results in near immobility of the affected joints. During the La Reunion outbreak (in Reunion Island in the Indian Ocean) in 2006, more than 60% of the people reported painful joints three years after the original Chikungunya infection [Schilte et al., 2013]. Similarly after a local epidemic of chikungunya in Italy, 66% of the people reported muscle pains or joint pains one year after acute infection [Moro et al., 2012].

The word 'Chikungunya' is believed to have been derived from 'Kungunyala', meaning "that which bends up" in Makonde language, which refers to the contorted posture of people affected with the severe joint pain associated with this disease [CDC, 2006]. Chikungunya was discovered by Marion Robinson and W.H.R. Lumsden in 1955 after an outbreak in 1952 on the Makonde Plateau, the mainland part of modern-day Tanzania. They found that in Africa, the virus largely cycles between other non-human primates, like monkeys, birds, cattle, and rodents, and mosquitos between human outbreaks [Powers and Logue, 2007]. Due to the high concentration of virus in the blood of those infected (or in the acute stage of infecton), the virus can circulate to and fro between humans and mosquitos very easily. Hence outbreaks are usually related to heavy rainfall which implies increase in mosquito population [Burt et al., 2012].

Since its discovery, periodic outbreaks have been documented in Africa, South Asia, and Southeast Asia. After some years of inactivity, in 2005 Chikungunya caused large outbreaks in Africa and Asia. For example in 2006, in India it re-appeared after 32 years of absence in an outbreak that reported 1.25 million suspected cases [Lahariya and Pradhan, 2006]. Before that, the largest Chikungunya epidemic that had been documented was in 2005 in an outbreak on the Runion Island in the Indian Ocean. It was estimated that 266,000 people were affected on the island which had a population of approximately 770,000 people [Roth et al., 2014].

The outbreak which started in 2005 was very severe and its severity is attributed to a change in the genetic sequence of the virus which allows it to multiply more easily in mosquito cells. The mutation also allows the virus to be carried by the Asian tiger mosquito, *Aedes albopictus*, in addition to its main vector or carrier *Aedes aegypti*. This could increase the risk of outbreaks since *Aedes aegypti* grows strictly in tropical climate whereas *Aedes albopictus* is a more invasive species which has spread through Europe, the Americas, the Carribean, Africa and the Middle East [Schuffenecker et al., 2006] [Tsetsarkin et al., 2007].

While Chikungunya transmission had never been documented in the Americas before 2013, the potential for outbreaks had long been recognized because of the prevalence of the vectors (carriers) and their efficiency at transmitting dengue viruses [CDC, 2014]. In December 2013, Pan American Health Organization (PAHO) and World Health Organization (WHO) reported the first cases of locally acquired Chikungunya infections in Americas, reported from St. Martin [Leparc-Goffart et al., 2015]. As of November 2015, 1.5 million cases have been reported in Americas since its start in December 2013, which has amplified the concern and awareness about this disease.

Due to the recent emergence of the disease in the Americas, the current extent of spread and risk is uncertain. It is important for us to understand the spread of Chikungunya for effective intervention, but it is a difficult task as cases might be unrecognized or confused with other diseases such as dengue. Some cases might not even get reported. Analyzing travel patterns is also important to understand the spread of transmissions. But it is very difficult to capture travel patterns in real-time and sometimes the patterns change due to the outbreak itself. Further, epidemics are themselves stochastic in nature [Johansson et al., 2014].

In 2013, Pan American Health Organisation (hereby called PAHO) in collaboration with the U.S. Center for Disease Control and Prevention (CDC) published new guidelines on Chikungunya. PAHO recommends that countries must maintain the capacity to detect and confirm Chikungunya cases, manage patients and implement social communication strategies to reduce the presence of mosquitos [PAHO, 2013]. PAHO then published the cumulative number of Chikungunya cases for all the countries in the Americas.

In this paper we try to understand and predict the spread of the Chikungunya disease in the Americas. We propose two different models to predict the infected case counts. The first model, named multi-country SIR compartment model, uses a SIR compartment model for every country and incorporates information about travel between the countries. The second model, named multi-country ARIMA model uses a multi-variate ARIMA model to forecast the number of infected cases of Chikunguya. The models have been explained in detail in Section 3. One of the major porblems encountered in modeling the spread of the disease is the scarcity of data. This has been discussed in Section 2 in more detail.

# 2 Data on Chikungunya Transmission in the Americas

Countries affected by Chikungunya in the Americas are required by PAHO to maintain a record of the progress of the disease since December 2013. The countries maintain a record of the number of suspected, confirmed and imported cases of Chikungunya in their country. The suspected and confirmed cases are counts for autochthonous (locally acquired) transmissions of the disease. Autochthounus cases are those cases which are native rather than descended from migrants or colonists and hence their presence in a country signifies the presence of the virus in the mosquito population of the country. In addition to collecting the raw counts of the people currently infected, PAHO computes the incidence rate of the disease in every country, that is, it reports the number of confirmed autochthonous transmissions per hundred thousand population.

PAHO maintains the weekly record of the cumulative counts for all the countries in Americas on their website (www.paho.com). As of November 2015, fifty-one countries in the Americas have been affected by Chikungunya and so the data consists of the cases reported weekly in each of these countries since December, 2013. There have been a total of $61,282$ confirmed autochthonus cases in the Americas in a total of $97$ epidemic weeks counting until November $6^{\text{th}}$, 2015.

One of the main challenges in extracting the data is that the weekly record of the cumulative counts published by PAHO, is in the pdf format and most automated text scraping tools do not suffice to extract the data. Some of the entries are also in Spanish instead of English which makes it really difficult to regularize the data for modeling purposes. We used different individualized scripts for every week to extract

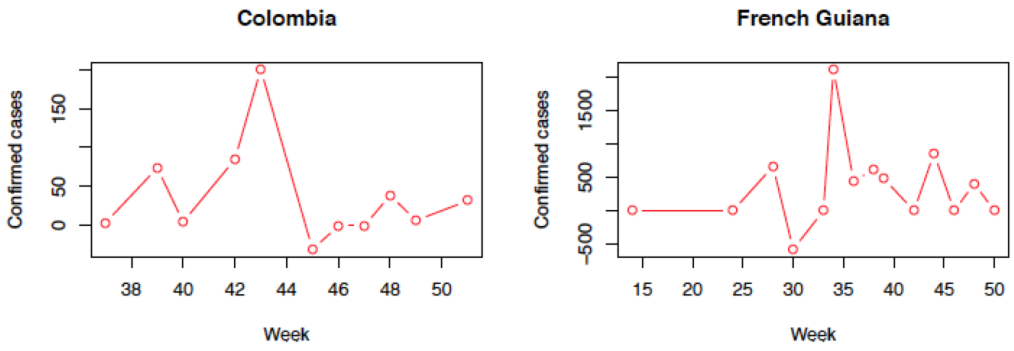


FIGURE 2.1: Confirmed new cases per epidemic week in Colombia and French Guiana. The count at epidemic week 45 for Colombia and at epidemic week 30 for French Guiana are negative due to error.

the data for that week.

Another major challenge is that data is not available for every week. There are 29 countries with more than 10 weeks worth of data and just 6 countries with more than 30 weeks worth of data. So we don't have sufficient data to model the spread of the disease for every country. The countries that have more than 30 weeks of data are Mexico, El Salvador, French Guiana, Puerto Rico, Colombia and Paraguay.

As mentioned earlier, the case counts are updated cumulatively, that is, every week, the number of new cases for that week are added to the previous week's case counts. Sometimes the new cases in a week are miscounted either due to a wrong diagnosis or due to manual error in entering the counts. These errors are usually corrected in subsequent weeks. As a result of the updates, sometimes the cumulative counts decrease in consecutive weeks instead of being non-decreasing. For example, plotting the difference in the cumulative counts of consecutive weeks for Colombia and French Guiana, we notice that the number of infected cases is negative at epidemic week 45 and epidemic week 30 for Colombia and French Guiana respectively, see

Figure 2.1. Since we do not know when the error was made, we just assume zero new cases in the weeks with negative count.

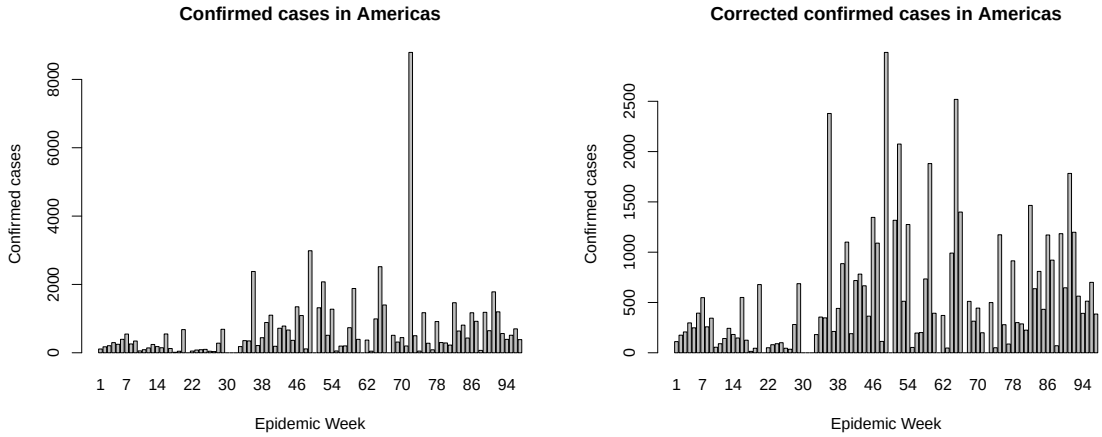


FIGURE 2.2: Total infectious subjects per week in Americas. We observe an anamoly at epidemic week 72 in the plot of Confirmed cases in Americas due to an adjustment in the process of updating the cumulative number of confirmed autochthonus cases. The corrected confirmed cases in Americas is corrected by just assuming that there were no new confirmed cases in epidemic week 72.

We also noticed a sudden drop in the total cumulative confirmed cases in the Americas from epidemic week 71 to 72, that is from May $8^{\text{th}}$ to $15^{\text{th}}$, 2015. The counts dropped from $31,223$ to $8,790$ in a week. The most likely reason for such an abrupt change is a change in the process of updating the cumulative counts. But as the reason of the abrupt change is unknown, we try to explore the different reasons that could have caused this change.

First, we take a difference of the cumulative counts to get the new confirmed autochthonus cases per week. This would result in a huge negative count at epidemic week 72. So instead, we decide to treat the cumulative count at epidemic week 72 as the number of new cases that week, see Figure 2.2. We notice that the new count of $8,790$ at epidemic week 72 is much higher than in any other week. Therefore it is not

likely to be the number of new autochthonus cases in epidemic week 72.

To avoid further complications and not lose too much information, we just assume that there were no new confirmed cases between epidemic weeks 71 and 72. The number of new confirmed cases from epidemic week 73 onwards are assumed to be correct and then the new cumulative counts are taken into consideration. From the corrected new confirmed cases per week in Figure 2.2, we see that assuming no new confirmed cases in epidemic week 72 does not really change the local distribution.

We use the thus corrected new confirmed cases per week to model the spread of the disease in the Americas. Due to lack of data, we don't model the spread of the disease for every country. Instead, we model the spread of the disease for the Americas on the whole and for a few specific countries in the results section.

# 3 Methods

We have used two models for modeling the number of infected cases of Chikungunya: a multi-country SIR model and a multi-country ARIMA model. We used the multi-country SIR compartment model to understand the spread of the disease, see subsection 3.1. In fact the model helps us to understand whether Chikungunya will cause an epidemic. On the other hand, the multi-country ARIMA model was considered in order to predict the future number of infected cases, see subsection 3.3. In this section, we also consider a theoretical extension of the simple SIR compartment model that includes the mosquito population, in subsection subsection 3.2. As we do not have necessary data on the mosquito population, we investigate the theory behind the model but do not use this model in the Results section. Despite this, it is

discussed in this section to explain how multi-country SIR models can be extended to incorporate mosquito population.

The dynamics of the disease in the multi-country models, are modeled by accounting for infected people travelling from one country to another. For multi-country SIR modeling, we use a different SIR compartment model for each country and include a variable for people travelling between the infected compartments of the different countries. The data for the movement of people between the different countries could be taken from flight itineraries but as not all the countries have the data readily available, we assumed constant number of people traveling between the different countries, every week.

In the multi-country ARIMA model we include a variable for counting the number of infected cases of Chikungunya in each country and we incorporate the information regarding travel between the countries using a multi-variate ARIMA model. Hence, the multi-variate model explains the influence of a country in spreading the disease in another country.

## 3.1 Multi-country SIR Compartment Model

Compartment models are one of the most commonly used methods for modeling epidemics. The method is founded upon differential equations and was introduced by Kermack and McKendrick in the early 1900s [Kermack and McKendrick, 1927]. These models serve as a base mathematical framework for understanding the complex dynamics of diseases. The model assumes the population to be a homogeneous mixture of people who are divided between compartments. The compartments in the model represent their health status with respect to the pathogen in the system. They also

assume perfect mixing within the population which implies that people make contact at random and do not usually mix in a smaller subgroup.

The SIR model is a compartment model that considers three compartments called Susceptible (S), Infected (I) and Removed (R). Individuals belong to the susceptible compartment if they are susceptible to the infection. They belong to the infected compartment if they are already infected and to the removed compartment if they are neither infected nor susceptible to the infection. The italic letters, $S$, I and $R$ are used to denote the populations in S, I and R compartments respectively. The italic letter $N$ is used to denote the total population, that is, $S + I + R = N$.

Since individuals in the R compartemnt are not susceptible to infection, only people in the susceptible (S) compartment can get infected in the population. Also they get infected only when they come in contact with an infected person (with some probability). Hence the rate at which people get infected is proportional to the rate of contacts between susceptible and infected people, that is, it is proportioanl to $SI/N$. Once the suspectible people are infected they leave S compartment. Hence the rate at which susceptible people get infected is also equal to the rate at which the S compartment's population decreases. Therefore,

$$\frac{dS}{dt} \propto -\frac{SI}{N}. \tag{3.1}$$

Let $\beta$ be the contact rate, which takes into account the probability of getting the disease in a contact between a susceptible and an infectious subject. Then $\beta$ becomes the proportionality constant in equation 3.1.

Considering the infected (I) compartment, we notice that the population increases as the infected people from S compartment move to the I compartment. But some

of the infected people also recover from the disease and hence are removed to the R compartment. Let $\gamma$ be the recovery rate, indicating the average proportion of infected people who recover every instant. Hence $\gamma$ can also be seen as the inverse of the average recovery time. Then the change in the population of I compartment can be given by,

$$\frac{dI}{dt} = \frac{\beta}{N} SI - \gamma I. \tag{3.2}$$

As the number of people in the removed (R) compartment can be found by $R = N - S - I$, the change in population of R compartment can be given by,

$$\frac{dR}{dt} = -\frac{dS}{dt} - \frac{dI}{dt} = \gamma I. \tag{3.3}$$

Combining the equations above (3.1, 3.2, 3.3), SIR models can be defined by the following set of differential equations. [Kermack and McKendrick, 1927]

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta}{N} SI \\ \frac{dI}{dt} &= \frac{\beta}{N} SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{3.4}$$

where,

$\beta$ is the contact rate,

$\gamma$ is the recovery rate,

$S$ is the number of susceptible people,

$I$ is the number of infected people,

$R$ is the number of removed people,

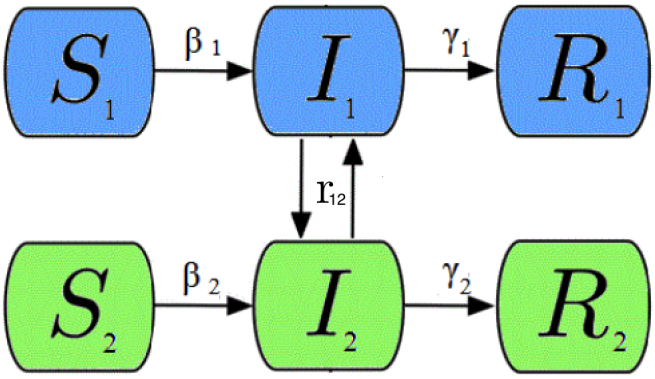


FIGURE 3.1: Two-country SIR compartment model

$N$ is the total population.

For a fixed SIR compartment model, the basic reproduction number, $R_0 = \frac{\beta}{\gamma}$ can be defined as the expected number of new infections from a single infection in a population where all people are susceptible. Therefore, having a value of $R_0 > 1$ indicates an epidemic where the infection peaks and eventually dies down and a value of $R_0 < 1$ indicates that the infection will die out without an epidemic.

To propose a multi-country SIR compartment model, we model every country with a different compartment model and include travel between the infected compartments of the different countries. Let us consider $m$ countries with variables $\beta_i, \gamma_i, S_i, I_i, R_i$ and $N_i$ as defined before, for every country $i = 1, 2, \ldots, m$ and let $r_{ij}$ denote the number of people traveling between any two countries $i$ and $j$.

We assume that the population of the countries ($N_i$) remains constant over time. Hence movement between the susceptible and removed compartments of different countries is inconsequential to the dynamics of the disease. We also assume that the movement is homogeneous, that is, the number of people belonging to a particular compartment among the people who cross borders, is proportional to the number of

people belonging to the compartment in the country. Lastly, we assume that there is no migration between countries and so the number of people traveling from country $i$ to $j$ is the same as the number of people moving from $j$ to $i$, that is, $r_{ij} = r_{ji}$. Figure 3.1 gives a pictographic representation of a two-country SIR compartment model.

Therefore the multi-country SIR compartment model for countries $i = 1, 2, \ldots, m$ is characterized by the following differential equations:

$$\begin{aligned}
\frac{dS_i}{dt} &= -\frac{\beta_i}{N_i} S_i I_i \\
\frac{dI_i}{dt} &= \frac{\beta_i}{N_i} S_i I_i - \gamma_i I_i - \sum_{j=1, j\neq i}^{m} r_{ij} \frac{I_i}{N_i} + \sum_{j=1, j\neq i}^{m} r_{ji} \frac{I_j}{N_j} \\
\frac{dR_i}{dt} &= \gamma_i I_i
\end{aligned} \tag{3.5}$$

where $\beta_i, \gamma_i, S_i, I_i, R_i$ and $N_i$ are defined as before for country $i = 1, 2, \ldots, m$ and $r_{ij} = r_{ji}$ denotes the number of people traveling between any two countries $i$ and $j$.

While the multi-country model considers the change in the dynamics of the disease as infected people travel from one country to another, they don't really consider the prevalence of the disease carrying vector, the mosquitos in the case of Chikungunya. Ideally, to understand the dynamics of the spread of any mosquito-borne infectious disease, for example Chikungunya, we should consider a compartment model that includes the mosquito population. Therefore, we propose a multi-country compartment model that includes the mosquito population in the next subsection.

## 3.2 Multi-country Ross-Macdonald Model for Mosquito-borne Infectious Diseases

Ronald Ross and George Macdonald developed a mathematical model of mosquito-borne transmissions commonly known as Ross-Macdonald Model [Smith et al., 2012]. The model considers homogeneous human and mosquito population and perfect mixing within the two populations and between the mosquito and human population. It also assumes constant population of the humans and mosquitos. The model is given by:

$$\begin{aligned} \frac{dI_H}{dt} &= abI_M \frac{N_H - I_H}{N_H} - \gamma I_H \\ \frac{dI_M}{dt} &= ac(N_M - I_M)\frac{I_H}{N_H} - \delta I_M \end{aligned} \tag{3.6}$$

where,

$a$ 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

$\gamma$ human recovery rate: inverse of average duration of infection in humans,

$\delta$ mosquito death rate: inverse of average duration of mosquito infection.

$I_H$ number of infected humans,

$N_H$ total number of humans in population,

$I_M$ number of infected mosquitos,

$N_M$ total number of mosquitos in population.

We propose a multi-country model, similar to the multi-country SIR compartment

model, that considers a Ross-Macdonald model for each country and then incorporates the travel between the infected compartments of the countries. Then the set of differential equations for the system is given by,

$$\begin{aligned}
\frac{dI_{Hi}}{dt} &= ab_i I_{Mi}\frac{N_{Hi}-I_{Hi}}{N_{Hi}} - \gamma_i I_{Hi} - \sum_{j=1,j\neq i}^{m} r_{ij}\frac{I_{Hi}}{N_{Hi}} + \sum_{j=1,j\neq i}^{m} r_{ji}\frac{I_{Hj}}{N_{Hj}} \\
\frac{dI_{Mi}}{dt} &= ac_i(N_{Mi}-I_{Mi})\frac{I_{Hi}}{N_{Hi}} - \delta_i I_{Mi}
\end{aligned} \tag{3.7}$$

where the variables $a, b_i, c_i, \gamma_i, \delta_i, N_{Hi}, I_{Hi}, N_{Mi}, I_{Mi}$ are as defined in equation (3.6) for every country $i = 1, 2, \ldots, m$. Similar to the multi-country SIR compartment model in equation (3.5), we define $r_{ij}$ as the number of people traveling between any two countries $i$ and $j$. Notice that, the mosquito biting rate $a$ is kept a constant and is assumed to not depend on which country the mosquito is from. Therefore, it need not be indexed by the country.

Due to the lack of data on mosquito population, we do not apply this approach in this paper. Vector Map provides data on number of mosquitos caught in particular areas, which in the future, can be used to get the mosquito population in a particular area. The backdraw of compartment models is that the assumptions made might not be accurate and hence might affect the accuracy of the predictions. In the Results section, we see that the SIR compartment models are not very accurate and hence we need a better model to forecast the spread of the disease. We therefore propose a multi-variate time series model, the proposed multi-country ARIMA model, to forecast the spread of Chikungunya in the Americas.

## 3.3 Multi-country Autoregressive Integrated Moving Average (ARIMA) Model

While the previously discussed compartment models use a set of differential equations, a different approach for modeling disease counts, is an ARIMA model which uses data at previous time points to estimate the present. The advantage of ARIMA models over compartment models is that, ARIMA models can be used to fit time series data both to better understand the data and to predict future points in the series (forecast). Compared to other time series models, ARIMA models are generally used in cases where there is evidence of non-stationarity in the data but where an initial differencing step can be applied to reduce the non-stationarity [Box and Jenkins, 1990].

ARIMA models are generally denoted by ARIMA$(p, d, q)$ where parameters $p$, $d$, and $q$ are non-negative integers, $p$ is the order of the Autoregressive model, $d$ is the degree of differencing, and $q$ is the order of the Moving-Average model. Then given time series data $X_t \in \mathbb{R}$, where $t$ is an integer, an ARIMA$(p, d, q)$ model is given by the following equation:

$$\left(1 - \sum_{i=1}^{p} \alpha_i B^i\right) (1 - B)^d X_t = \delta + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t, \tag{3.8}$$

where, $B$ is backward shift operator, i.e., $(1 - B)$ is the difference operator. Hence, $BX_t = X_{t-1}$ and $B^k X_t = X_{t-k}$ for any $k$ and $t$. $\alpha_i$'s are the parameters of the autoregressive part of the model, $\theta_i$'s are the parameters of the moving average part and the $\varepsilon_t$'s are error terms that are assumed to be independent and identically distributed variables sampled from $N(0, \sigma_t^2)$. $\delta/(1 - \sum_{i=1}^{p} \alpha_i)$ is the drift of the ARIMA$(p, d, q)$ process.

While the ARIMA model can be used to model the number of infections in one area, we would like to incorporate the effect of a country in the spread of the disease in another country. So similar to the multi-country SIR compartment model, we propose an ARIMA model that could also explore the relationship of the infections between the different countries. We propose using a multivariate ARIMA model to explain the spread of the disease across the countries where each variate is the number of infections in a different country. Then given time series data $\mathbf{X}_t \in \mathbb{R}^m$, where $\mathbf{X}_{it}$ is the infections count in country $i$, a multi-country ARIMA$(p, d, q)$ model is given by the following equation:

$$\left(\mathbf{I} - \sum_{i=1}^{p} \alpha_i B^i\right)(1 - B)^d \mathbf{X}_t = \left(\mathbf{I} + \sum_{i=1}^{q} \Theta_i B^i\right) \varepsilon_t, \tag{3.9}$$

where, $B$ is backward shift operator as before, such that, $B\mathbf{X}_t = \mathbf{X}_{t-1}$ and $B^k\mathbf{X}_t = \mathbf{X}_{t-k}$ for any $k$ and $t$. $\alpha_i$'s are matrices which give the parameters of the autoregressive part of the model, $\theta_i$'s are also matrices that give the parameters of the moving average part and the $\varepsilon_t$'s are error vectors that are assumed to be independent and identically distributed variables sampled from multivariate normal distribution, $N_m(\mathbf{0}, \sigma^2\mathbf{I})$.

# 4 Results

In this section, we model the spread of Chikungunya using the multi-country SIR compartment model and the multi-country ARIMA model. Section 4.1 models the spread of Chikungunya in the Americas using a single SIR compartment model. Travel is incorporated into the model in Section 4.2 where we fit a multi-country model to the number of autochthonus cases in St. Martin and St. Barthelemy, which were the first

two islands affected by Chikungunya in the Americas. We consider only two countries St. Martin and St. Barthelemy, due to the lack of data to create more complicated models that consider more countries.

Due to the failure of the compartment models in forecasting the spread of the disease we use ARIMA model to forecast the total number of confirmed autochthonus cases in the Americas in Section 4.3. We then use the multi-country ARIMA model to incorporate travel into the ARIMA model and use it to model the spread of the disease for French Guiana, Puerto Rico and Colombia. Once again, we only model for these three countries as we have insufficient data to include other countries in the model.

## 4.1 SIR Compartment Model for the Total Counts in the Americas

We first model the spread of chikungunya in the Americas using an SIR compartment model. That is, we estimate $\beta$ and $\gamma$ for the Americas in the set of differential equations 3.4 given in Section 3.1. The set of differential equations were as follows:

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta}{N}SI \\ \frac{dI}{dt} &= \frac{\beta}{N}SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{4.1}$$

The advantage of using an SIR model is that it gives an estimate for the basic reproduction number, $R_0$. As any disease having an $R_0$ of value greater than one is expected to become an epidemic, this is a number epidemiologists are generally interested in.

The set of differential equations 4.1, do not have any closed form solution as discussed in Section 3.1. Therefore to model the data using SIR compartment model, we need a non-linear optimization technique. We use Nelder-Mead optimization algorithm to select the optimal values of $\beta$ and $\gamma$ for the Americas. We consider the log-sum of error squares as our objective function in order to minimize the prediction errors.

As there were $111$ total number of autochthonus cases in the Americas initially, we start the algorithm at $I_0 = 111, R_0 = 0$ and $S_0 = N - I$, where $N$ is the population of Americas. The population of Americas is currently around $991.1$ million. Then for each value of the parameters $\beta$ and $\gamma$, we can get estimates of $S_t, I_t$ and $R_t$ at every week $t$.

Since the cumulative counts of the confirmed cases give the sum of the I and R compartments for the given week, the error is computed as the difference between the observed cumulative counts and the sum of estimated I and R from the model. That is, if $Y_{t,\beta,\gamma}$ is the total number of cumulative autochthonus cases in the Americas in week $t$ and $\hat{I}_{t,\beta,\gamma}$ and $\hat{R}_{t,\beta,\gamma}$ are the predicted number of infected and removed cases as predicted by the differential equations, then the error at week $t$ is given by,

$$e_{t,\beta,\gamma} = Y_{t,\beta,\gamma} - \hat{I}_{t,\beta,\gamma} - \hat{R}_{t,\beta,\gamma}. \tag{4.2}$$

Therefore for any value of $\beta$ and $\gamma$, we can find the corresponding objective function, that is the log-sum of error squares as

$$l(\beta, \gamma) = \log \sum_t \left(Y_{t,\beta,\gamma} - \hat{I}_{t,\beta,\gamma} - \hat{R}_{t,\beta,\gamma}\right)^2. \tag{4.3}$$

We run the Nelder-Mead optimization algorithm to find the $\beta$ and $\gamma$ that minimizes the objective function given above. As the optimization problem is not convex, the algorithm usually gives a local optima. Hence we try different starting values for the parameters $\beta$ and $\gamma$ and run the Nelder-Mead optimization algorithm for each of them. We then select the optimum parameters with the minimum objective function.

The optimum $\beta$ and $\gamma$ for the Americas can be seen in Table 4.1. The basic reproduction number, $R_0$, that is computed as a ratio of $\beta$ over $\gamma$ is 1.0314 for the Americas. Though the value is greater than one, it is not that far from one. We compare the value with the basic reproduction number of Ebola. Ebola's basic reproduction number was found to be 1.51 for Guinea, 2.53 for Sierra Leone and 1.59 for Liberia by Christian L. Althaus [Althaus, 2014]. We see that the $R_0$ for Ebola is higher than that of Chikungunya's. Hence it does not clearly answer whether Chikungunya will become and epidemic or not.

| **Parameters** | $\beta$ | $\gamma$ | $R_0$ |
|---|---|---|---|
| Americas | 1 | 0.9695 | 1.0314 |

TABLE 4.1: The estimates of parameters $\beta$, the contact rate, $\gamma$, the recovery rate and Basic Reproduction Number $R_0 = \frac{\beta}{\gamma}$ for the Americas.

The major drawback of a compartment model is that it makes very rigorous assumptions about the homogenity of the compartments and perfect-mixing between compartments as mentioned in Section 3.1. Therefore the prediction of the number of new cases is usually not very accurate. The predicted number of total autochthonus cases in the Americas can be seen in Figure 4.1. We notice that the SIR compartment model does not predict the number of new cases of Chikungunya very well but it can

be used to understand if Chikungunya will cause an epidemic or not.

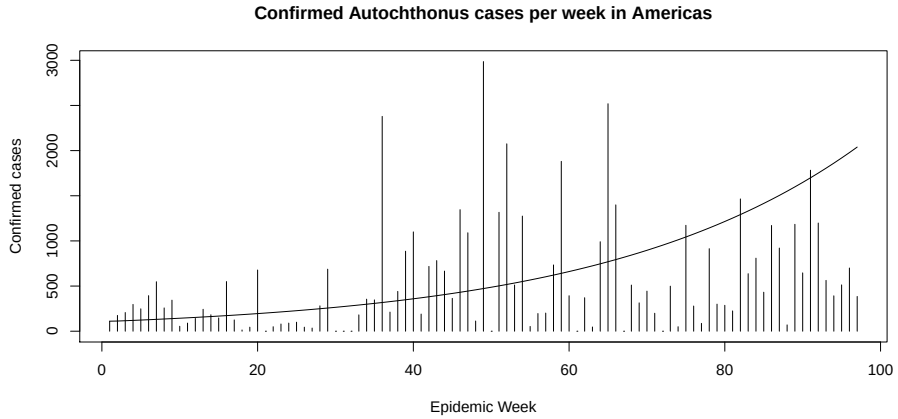


FIGURE 4.1: Predicted number of confirmed cases per week in Americas.

## 4.2 Multi-country SIR Model for St. Martin and St. Barthelemy

The Chikungunya transmission first started in St. Martin in the Caribbean islands and spread to St. Barthelemy, also known as St. Barts. We modeled the transmission between them using a multi-country SIR compartment model given by the differential equations in the set of differential equations 3.5 mentioned in Section 3.1. The differential equations were

$$\begin{aligned}
\frac{dS_i}{dt} &= -\frac{\beta_i}{N_i} S_i I_i \\
\frac{dI_i}{dt} &= \frac{\beta_i}{N_i} S_i I_i - \gamma_i I_i - \sum_{j=1, j\neq i}^{m} r_{ij} \frac{I_i}{N_i} + \sum_{j=1, j\neq i}^{m} r_{ji} \frac{I_j}{N_j} \\
\frac{dR_i}{dt} &= \gamma_i I_i
\end{aligned} \tag{4.4}$$

Here the number of countries, $m = 2$ and hence $i, j = 1, 2$. The number of people traveling from St. Martin to St. Baths and vice-versa is assumed to be, $r_{12} = r_{21} = 210$, that is, we assume approximately 30 people travel from one to the other per day. This number was obtained by looking at flight itineraries for flights between the two islands and capacity of each flight.

Due to the availability of only the cumulative number of confirmed cases, we have the sum of the number of people in infected and removed compartments. So the optimum parameters of the model can be found by using Nelder-Mead optimization algorithm, similar to what we did in the previous section. We minimize the log-sum of squared errors where the errors are computed by taking the difference between cumulative confirmed cases and sum of estimated $I$ and $R$ for the two countries seperately.

| **Compartment model** | **St. Martin** | | | **St. Barthelemy** | | |
|---|---|---|---|---|---|---|
| Parameters | $\beta$ | $\gamma$ | $R_0$ | $\beta$ | $\gamma$ | $R_0$ |
| SIR | 0.7979 | 0.9669 | 0.8252 | 0.7456 | 0.8706 | 0.8564 |
| Multi-Country SIR | 0.8046 | 0.9738 | 0.8262 | 0.7029 | 0.8261 | 0.8509 |

TABLE 4.2: The estimates of parameters $\beta$, the contact rate, $\gamma$, the recovery rate and Basic Reproduction Number $R_0 = \frac{\beta}{\gamma}$ for St. Martin and St. Barthelemy.

Modeling the Chikungunya transmissions in St. Martin and St. Barts using both the SIR model and the multi-country SIR model, we get the values of $\beta$ and $\gamma$ for the two countries as given in Table 4.2. The estimates of the parameters do not vary too much between the SIR compartment model and the multi-country SIR compartment model. Interestingly, even though the $R_0$ value for whole of Americas was seen to be greater than $1$, in this case for both countries it is less than $1$. A reason for this could be that the disease died down pretty quickly in St. Martin and St. Barts

due to their small populations. The data confirms this as we see that the infection indeed dies down in ten weeks in both countries. The number of infections peak in the sixth epidemic week but we can see that the Chikungunya infection did not cause an epidemic in these two countries, see Figure 4.2.

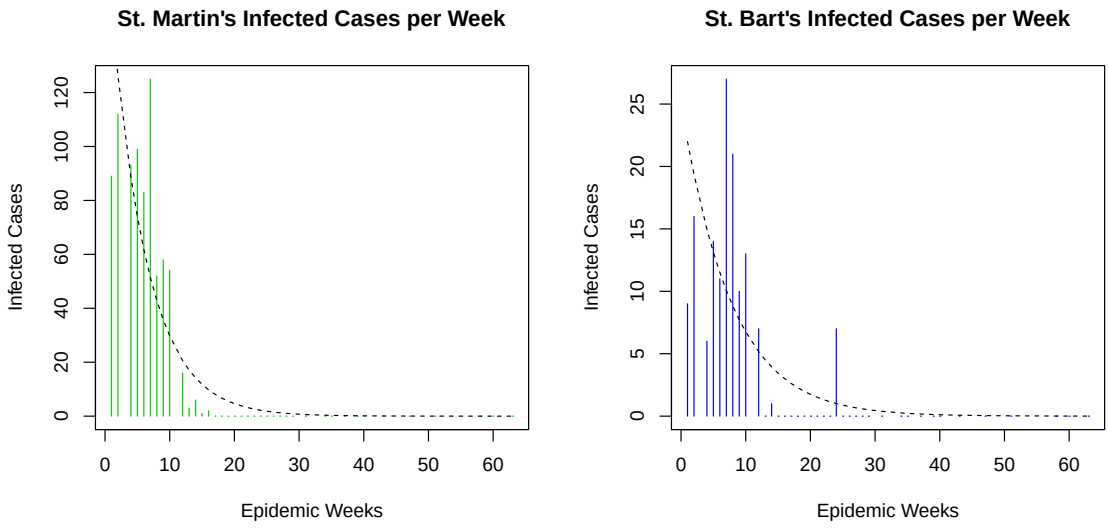


FIGURE 4.2: Infectious subjects per week in St. Martin and St. Barthelemy. The peak of the infectious period occurs at sixth epidemic week. In case of St. Barthelemy the peak does not seem to have been captured.

We notice that even in this model there is a similar backdraw like the previous model in Section 4.1. Due to the assumption of homogenity within the compartments and perfect mixing between the compartments, the compartment model tends to provide smoother estimates of the number of new confirmed cases every week. Hence, as we can notice in Figure 4.2, the peak at epidemic week 6 is not really captured well. Hence it does not really capture the full effect of the spread of Chikungunya and would not serve as a very good forecasting model.

## 4.3 Multi-country ARIMA Model

In order to create a good forecasting model, we consider ARIMA models. For forecasting the total cumulative number of confirmed cases in whole of the Americas using an ARIMA$(p, d, q)$ model as given in Equation 3.8 (Section 3.3), we choose the ideal parameters $p$, $d$ and $q$ by minimizing the Akaike Information Criterion (AIC). The model thus chosen is ARIMA$(4, 3, 8)$ with an AIC value of $1477.632$. But this model fits a total of $13$ parameters and seems to be overfitting the data and so we pick ARIMA$(0, 3, 2)$ whose AIC is $1480.766$ which is pretty close to the AIC value of the previous model. Comparing this prediction to the predicted values of the compartment model, we notice a huge improvement in the prediction (see in Figure 4.3).

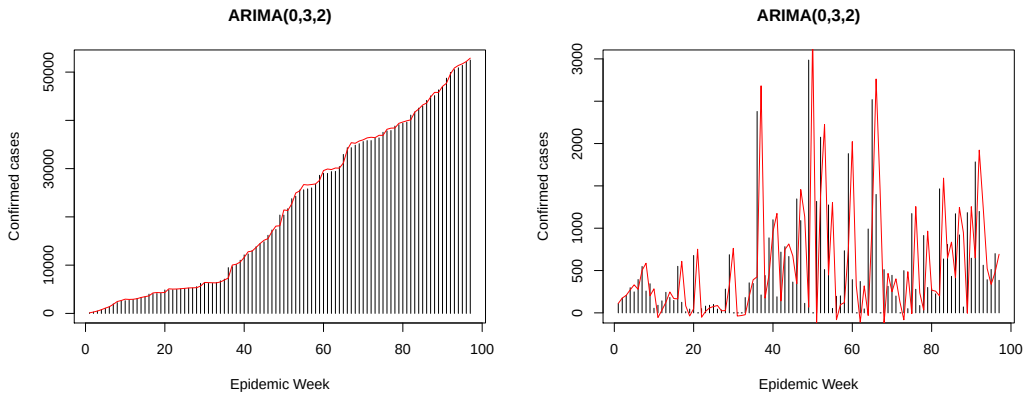


FIGURE 4.3: Predicted number of total cumulative confirmed cases and the total confirmed cases per week in Americas. The black dotted line with the predicted value. We notice that it is much better than the fit of the SIR compartment model.

To forecast the cumulative counts in the different countries, we could either fit an ARIMA$(0, 3, 2)$ to them (the model used for the total counts in Americas), or find the best ARIMA$(p, d, q)$ model for the country using minimum AIC, or fit a multivariate

ARIMA$(p, d, q)$ model, given in Equation 3.9 (Section 3.3) to all the countries and use that to forecast in the given country.

The multivariate ARIMA model fits a lot of parameters and so we need suffiecient data to predict them. Unfortunately, as discussed in Section 2, we just have 6 countries which have more than 30 weeks worth of data but even that does not suffice. Therefore we fit the multivariate ARIMA model for just the three countries with the maximum data, namely French Guiana with data for 45 weeks, Puerto Rico with data for 60 weeks and Colombia with data for 54 weeks. Figure 4.4 compares the three different models for Puerto Rico.

We fit the multivariate ARIMA model with the minimum AIC to French Guiana, Puerto Rico and Colombia. As a result we choose a multi-variate ARIMA$(0, 1, 1)$ to model the spread of Chikungunya in these three countries. Similarly, we choose ARIMA$(2, 3, 1)$ as the best univariate ARIMA model for just Purto Rico. We fit the models on the data before epidemic week 90 and use it to forecast the number of cumulative cases for the next 8 weeks. The number of confirmed cases for the weeks are then found by taking a difference of the cumulative counts. In Figure 4.4, we compare the forecasts for Puerto Rico. We notice that, the multi-variate model forecasts much better because it is smoother. Also its variance is higher because of the multiple parameters that need to be estimated in the model.

Comparitively the two univariate ARIMA models are much less smoother, because of which they predict the increase in the counts much better in the first couple of weeks. The two univariate ARIMA models, the one that was selected for the total number of counts in the Americas, ARIMA$(0, 3, 2)$, and the one that was selected as the best model for Puerto Rico, ARIMA$(2, 3, 1)$, seem to be performing similarly.

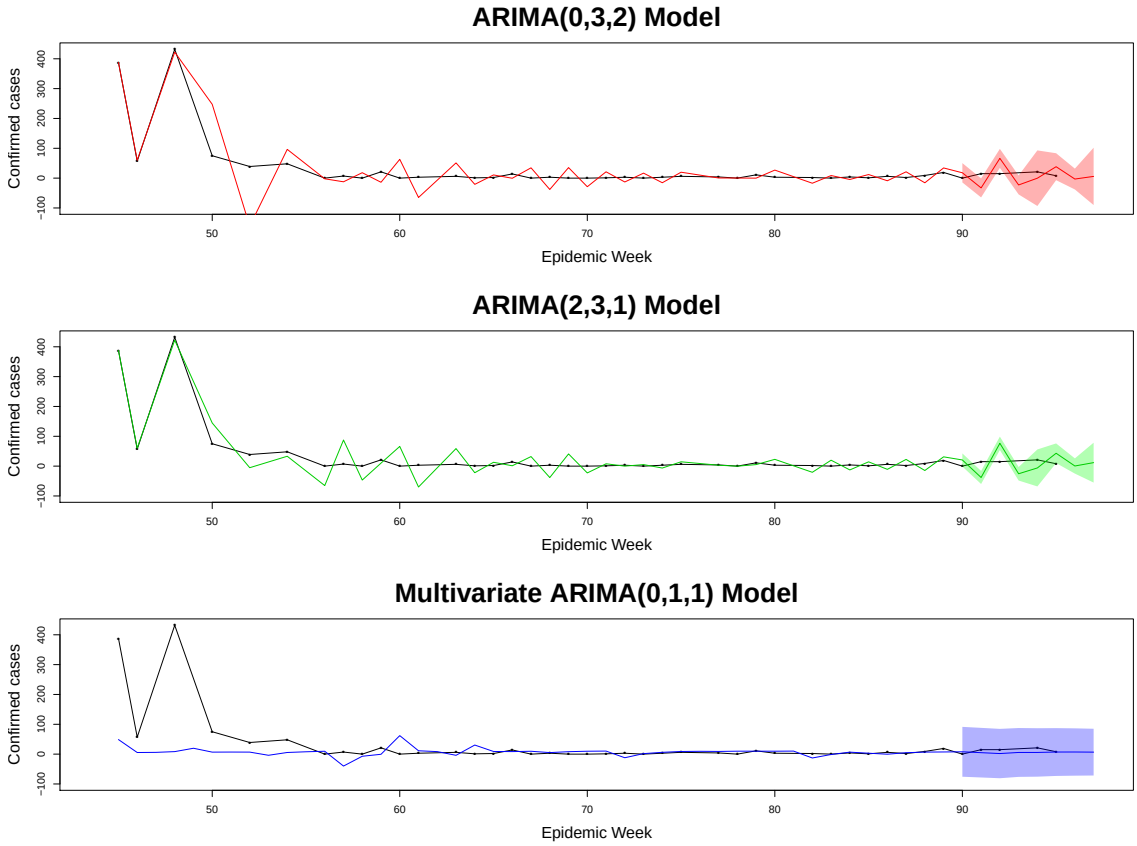


FIGURE 4.4: Comparing ARIMA$(0,3,2)$, which was selected for the total cumulative counts in Americas, ARIMA$(2,3,1)$, which was selected for Puerto Rico alone and multi-variate ARIMA$(0,1,1)$. Observed confirmed counts are given by the black line. The predictions are made for epidemic weeks 90 to 98, based on available previous data.

This could mean that instead of fitting a seperate model for every country, it could generally be a good idea to fit ARIMA$(0,3,2)$ to all the countries. We look at the residual plots for ARIMA$(0,3,2)$ and ARIMA$(2,3,1)$ in Figure 4.5. The plot shows us that other than a huge negative residual on epidemic week 50, the fit of ARIMA$(0,3,2)$ and ARIMA$(2,3,1)$ are similar. This validates our statement that ARIMA$(0,3,2)$ works pretty well for Puerto Rico. Similarly, it also performs well for French Guiana and Colombia.

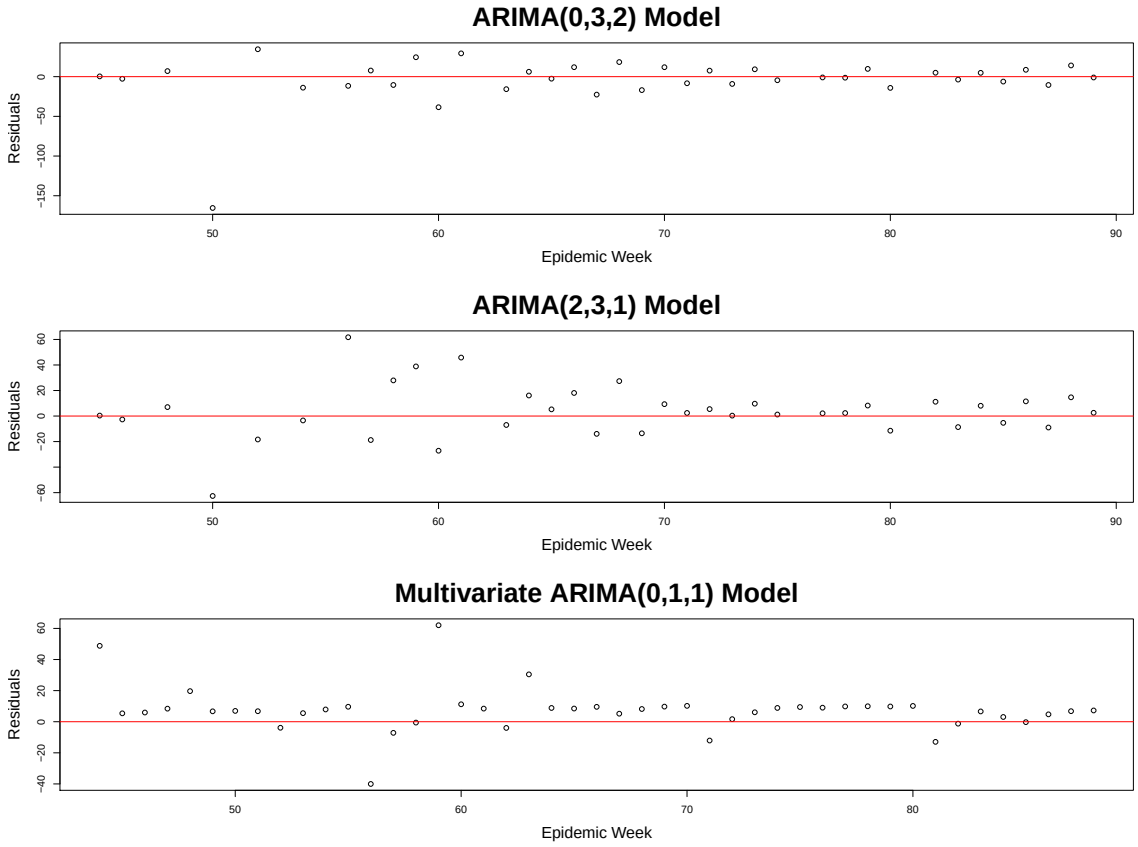


FIGURE 4.5: Comparing the residuals for Puerto Rico of ARIMA$(0,3,2)$, which was selected for the total cumulative counts in Americas, ARIMA$(2,3,1)$, which was selected for Puerto Rico alone and multi-variate ARIMA$(0,1,1)$. The predictions are made for epidemic weeks 90 to 98, based on available previous data.

# 5 Discussion

The Chikungunya outbreak in the Americas started in December 2013 in St. Martin and soon spread to other countries of the Americas. Currently fifty-one countries in the Americas have been affected and understanding the spread of the disease is critical to alert people to the risk of disease and to implement control measures. We used multi-country SIR compartment model to model the Chikungunya transmission between St. Martin and St. Barthelemy, which were the first two islands in the

Caribbean to have been affected by the infection. We notice that the Chikungunya transmission did not cause an epidemic in the two countries and died down after a while.

In future, we plan to use the multi-country SIR compartment model for all the countries in America. But due to lack of data on the spread of Chikungunya and inaccessibility of data on travel between the different countries, it is challenging to fit a multi-country SIR compartment model.

As data on mosquito population is also not available it could be challenging task to fit the Ross-Macdonald model to the data. The current estimates of mosquitos in a location is primarily based on Centers for Disease Control and Prevention (CDC) light trap collections, which provide only point data. Logistic regression models have also been proposed to estimate mosquito abundance in areas not sampled by traps [Diuk-Wasser et al., 2006]. These estimates of mosquito populations could then be used to fit Ross-Macdonald model.

The multi-country ARIMA model seems to fix these drawbacks as it doesn't make any strong assumptions nor does it need data on travel or on mosquito population. But since we need to estimate a lot of parameters for even the simplest multi-country ARIMA model, we need sufficient data before we can continue to fit a multi-country model for all the countries in the Americas. As we have seen with Puerto Rico, French Guiana and Colombia, it might be a good idea to just use the ARIMA model that best fits the entire Americas to forecast in the individual countries too.

# 6 APPENDIX

## 6.1 Another Approach: Agent Based Model

Another approach in epidemiology to predict the spread of a disease, is to use an agent based model. In an agent based model, each individual is considered to be in one of the following states: susceptible (S), exposed (E), infected (I) or removed (R). Assume there are $N$ people in the population and let $S_t[i], E_t[i], I_t[i]$ and $R_t[i]$ denote the state to which individual $i$ belongs at time $t$, where $S_t[i] = 1$ if $i$ is susceptible, and zero otherwise. $E_t[i], I_t[i]$ and $R_t[i]$ can be defined simlarly. Then,

$$S_t[i], E_t[i], I_t[i], R_t[i] \in \{0, 1\}$$

$$S_t[i] + E_t[i] + I_t[i] + R_t[i] = 1 \ \forall \ i = 1, ..., N.$$

An individual moves from the susceptible to the exposed compartment when he gets infected, moves from exposed to infected if he shows symptoms or exposes symptoms and moves from infected to removed when the infection passes away or the subject is removed. So the movement of an individual from one state to another can be captured by Infection rate (IR), Exposure rate (ER) and Removal rate (RR).

$$IR_t[i] = \begin{cases} 1, & \text{if i moves from S to E at time t} \\ 0, & \text{o.w.} \end{cases}$$

$$ER_t[i] = \begin{cases} 1, & \text{if i moves from E to I at time t} \\ 0, & \text{o.w.} \end{cases}$$

$$RR_t[i] = \begin{cases} 1, & \text{if i moves from I to R at time t} \\ 0, & \text{o.w.} \end{cases}$$

The aggregate infection rate, exposure rate and removal rate are then given by,

$$IR_t = \sum_{i=1}^{n} IR_t[i], \quad ER_t = \sum_{i=1}^{n} ER_t[i], \quad RR_t = \sum_{i=1}^{n} RR_t[i].$$

Therefore we can write the progress of the model for each individual,
$j = 1, 2, ..., N$ as:

$$\begin{aligned} S_{t+1}[j] &= S_t[j] - IR_t[j] \\ E_{t+1}[j] &= E_t[j] + IR_t[j] - ER_t[j] \\ I_{t+1}[j] &= I_t[j] + ER_t[j] - RR_t[j] \\ R_{t+1}[j] &= R_t[j] + RR_t[j] \end{aligned}$$

In order to use an agent based model to predict the spread of Chikungunya in the Americas, we need to simulate the spread of the disease throughout the Americas which is computationally very heavy. We try to draw a comparision between the compartment models and agent based models to formulate a new model that can switch easily between them and use both of their advantages.

As we can see, agent based models are formulated in discrete time whereas compartmnet models are formulated in continuous time. In order to compare the two models we need to first formulate agent based model in continuous time.

### 6.1.1 Continuous Time Formulation of Agent Based Model

In continuous time the dynamics of a agent based model have to be invaariant to changes in the time step. Hence the states evolve as:

$$
\begin{aligned}
S_{t+1}[j] &= S_t[j] - dt IR_t[j] \\
E_{t+1}[j] &= E_t[j] + dt\{IR_t[j] - ER_t[j]\} \\
I_{t+1}[j] &= I_t[j] + dt\{ER_t[j] - RR_t[j]\} \\
R_{t+1}[j] &= R_t[j] + dt RR_t[j]
\end{aligned}
$$

where,

$$
IR_t[i] = \begin{cases} \frac{S_t[j]}{dt} & \text{w.p. P(j gets infected in dt)} \\ 0 & \text{o.w.} \end{cases}
$$

$$
ER_t[i] = \begin{cases} \frac{E_t[j]}{dt} & \text{w.p. P(j gets exposed in dt)} \\ 0 & \text{o.w.} \end{cases}
$$

$$
RR_t[i] = \begin{cases} \frac{I_t[j]}{dt} & \text{w.p. P(j gets removed in dt)} \\ 0 & \text{o.w.} \end{cases}
$$

### 6.1.2 Relationship between Agent Based Model and Compartment model

Compartment models assume that both the duration of the disease and the duration for emergence of symptoms follow exponential distribution for an individual. So the time any individual spends in the infected or exposed state follows exponential

distribution. Compartment models also assume perfect mixing, meaning every person meets every other person with the same probability.

Let us thus assume that the duration of disease for individual $j$ follows $\text{Exp}(\delta[j])$ and the duration for emergence of symptoms follows $\text{Exp}(\epsilon[j])$. Suppose $c_{IS}[j], c_{ES}[j]$ denotes the number of contacts that j has per unit time with infected people and exposed people respectively, when he is susceptible.

By assuming perfect mixing we get the probability that j gets infected by an infected person as,

$$\begin{aligned} P(\text{ j infected by I}) &= P(\text{ j meets infected})P(\text{ j infected} \mid \text{j meets infected}) \\ &= \frac{I}{N} i_{IS} \end{aligned}$$

where $i_{IS}$ is the probability that any person gets infected given he has had contact with an infected person. Similarly probability that j gets infected by an exposed person is,

$$\begin{aligned} P(\text{ j infected by E}) &= P(\text{ j meets exposed})P(\text{ j infected} \mid \text{j meets exposed}) \\ &= \frac{E}{N} i_{ES} \end{aligned}$$

where $i_{ES}$ is the probability that any person gets infected given he has had contact with an exposed person. So combining the two we get,

$$\begin{aligned} P(\text{ j infected in dt}) &= c_{IS}[j]dtP(\text{ j infected by I}) + c_{ES}[j]dtP(\text{ j infected by E}) \\ &= dt\left\{c_{IS}[j]i_{IS}\frac{I}{N} + c_{ES}[j]i_{ES}\frac{E}{N}\right\} \end{aligned}$$

Let $T_E[j]$ and $T_R[j]$ be the time taken for exposure of symptoms and removal of the infection for individual $j$ respectively. Then assuming exponential time durations, $T_E[j] \sim \text{Exp}(\epsilon[j])$ and $T_R[j] \sim \text{Exp}(\delta[j])$. Therefore assuming $dt$ is very small, the probability of exposing symptoms and that of being removed are as follows.

$$P(\text{ j exposed in dt } |E_t[j] = 1) = P(T_E[j] \leq dt) = 1 - \exp^{-\frac{dt}{\epsilon[j]}} \approx \frac{dt}{\epsilon[j]}.$$
$$P(\text{ j removed in dt } |I_t[j] = 1) = P(T_R[j] \leq dt) = 1 - \exp^{-\frac{dt}{\delta[j]}} \approx \frac{dt}{\delta[j]}.$$

Therefore computing the expectation we get that

$$E\left[\frac{dS[j]}{dt}\right] = -E\left[IR[j]\right] = -S[j]\left\{c_{IS}[j]i_{IS}\frac{I}{N} + c_{ES}[j]i_{ES}\frac{E}{N}\right\} \tag{6.1}$$
$$E\left[\frac{dE[j]}{dt}\right] = E\left[IR[j] - ER[j]\right] = S[j]\left\{c_{IS}[j]i_{IS}\frac{I}{N} + c_{ES}[j]i_{ES}\frac{E}{N}\right\} - \frac{E[j]}{\epsilon[j]} \tag{6.2}$$
$$E\left[\frac{dI[j]}{dt}\right] = E\left[ER[j] - RR[j]\right] = \frac{E[j]}{\epsilon[j]} - \frac{I[j]}{\delta[j]} \tag{6.3}$$
$$E\left[\frac{dR[j]}{dt}\right] = E\left[RR[j]\right] = \frac{I[j]}{\delta[j]}. \tag{6.4}$$

This looks very simlar to the differential equations of compartment models but it talks about the change in state of one individual and not of the population itself. We need a few more assumptions to derive the differential equations of the compartment models. Firstly, we need homogenity within the compartments (states) and secondly we need the flow between compartments to be equal to the expected value of the sum of the underlying probabilistic rates for each individual (i.e. flow between states for each individual).

# References

C. Althaus. Estimating the reproduction number of ebola virus (ebov) during the 2014 outbreak in west africa. *PLOS Currents Outbreaks*, 1(12), Sep 2014. doi: 10.1371/currents.outbreaks.

G. E. P. Box and G. Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, Incorporated, 1990. ISBN 0816211043.

F. J. Burt, M. S. Rolph, N. E. Rulli, S. Mahalingam, and M. T. Heise. Chikungunya: a re-emerging virus. *The Lancet*, 379(9816):662–671, 2015/11/24 2012. doi: 10.1016/S0140-6736(11)60281-X. URL http://dx.doi.org/10.1016/S0140-6736(11)60281-X.

CDC. Chikungunya fever diagnosed among international travelers–united states, 2005-2006. *MMWR Morb Mortal Wkly Rep*, 55(38):1040–1042, Sep 2006. ISSN 1545-861X (Electronic); 0149-2195 (Linking).

CDC. Preparedness and response for chikungunya virus introduction in the americas, 2014.

M. Chhabra, V. Mittal, D. Bhattacharya, U. Rana, and S. Lal. Chikungunya fever: A re-emerging viral infection. *Indian Journal of Medical Microbiology*, 26(1):5–12, 2008. ISSN 0255-0857. doi: 10.4103/0255-0857.38850.

M. A. Diuk-Wasser, H. E. Brown, T. G. Andreadis, and D. Fish. Modeling the spatial distribution of mosquito vectors for west nile virus in connecticut, usa. *Vector-Borne & Zoonotic Diseases*, 6(3):283–295, 2006.

M. A. Johansson, A. M. Powers, N. Pesik, N. J. Cohen, and J. E. Staples. Nowcasting the spread of chikungunya virus in the americas. *PLoS ONE*, 9(8):e104915, 08 2014.

W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 115, pages 700–721. The Royal Society, 1927.

C. Lahariya and S. Pradhan. Emergence of chikungunya virus in indian subcontinent after 32 years: a review. *Journal of vector borne diseases*, 43(4):151, 2006.

I. Leparc-Goffart, A. Nougairede, S. Cassadou, C. Prat, and X. de Lamballerie. *Chikungunya in the Americas*, volume 383. Elsevier, 2015/10/10 2015.

D. Mavalankar, P. Shastri, T. Bandyopadhyay, J. Parmar, and K. V. Ramani. Increased mortality rate associated with chikungunya epidemic, ahmedabad, india. *Emerging Infectious Diseases*, 14(3):412–415, 03 2008.

M. L. Moro, E. Grilli, A. Corvetta, G. Silvi, R. Angelini, F. Mascella, F. Miserocchi, P. Sambo, A. C. Finarelli, V. Sambri, C. Gagliotti, E. Massimiliani, A. Mattivi, A. M. Pierro, and P. Macini. Long-term chikungunya infection clinical manifestations after an outbreak in italy: A prognostic cohort study. *Journal of Infection*, 65(2):165–172, 2015/11/24 2012. doi: 10.1016/j.jinf.2012.04.005. URL http://dx.doi.org/10.1016/j.jinf.2012.04.005.

PAHO. Preparedness and response plan for chikungunya virus introduction in the caribbean sub-region. 2013.

A. M. Powers and C. H. Logue. Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *Journal of General Virology*, 88(9):2363–2377,

2007. URL http://jgv.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.82858-0.

A. Roth, D. Hoy, P. F. Horwood, B. Ropa, T. Hancock, L. Guillaumot, K. Rickart, P. Frison, B. Pavlin, and Y. Souares. Preparedness for threat of chikungunya in the pacific. *Emerging infectious diseases*, 20(8), 2014.

C. Schilte, F. Staikovsky, T. Couderc, Y. Madec, F. Carpentier, S. Kassab, M. L. Albert, M. Lecuit, and A. Michault. Chikungunya virus-associated long-term arthralgia: A 36-month prospective longitudinal study. *PLoS Negl Trop Dis*, 7(3):e2137, 03 2013. doi: 10.1371/journal.pntd.0002137. URL http://dx.doi.org/10.1371%2Fjournal.pntd.0002137.

I. Schuffenecker, I. Iteman, A. Michault, S. Murri, L. Frangeul, M.-C. Vaney, R. Lavenir, N. Pardigon, J.-M. Reynes, F. Pettinelli, L. Biscornet, L. Diancourt, S. Michel, S. Duquerroy, G. Guigon, M.-P. Frenkiel, A.-C. Brhin, N. Cubito, P. Desprs, F. Kunst, F. A. Rey, H. Zeller, and S. Brisse. Genome microevolution of chikungunya viruses causing the indian ocean outbreak. *PLoS Med*, 3(7):e263, 05 2006. doi: 10.1371/journal.pmed.0030263. URL http://dx.doi.org/10.1371%2Fjournal.pmed.0030263.

D. L. Smith, K. E. Battle, S. I. Hay, C. M. Barker, and T. W. Scott. Ross, macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS pathog*, 8(4):e1002588, 2012.

M. Sourisseau, C. Schilte, N. Casartelli, C. Trouillet, F. Guivel-Benhassine, D. Rudnicka, N. Sol-Foulon, K. L. Roux, M.-C. Prevost, H. Fsihi, M.-P. Frenkiel, F. Blanchet, P. V. Afonso, P.-E. Ceccaldi, S. Ozden, A. Gessain, I. Schuffenecker,

B. Verhasselt, A. Zamborlini, A. Saïb, F. A. Rey, F. Arenzana-Seisdedos, P. Desprès, A. Michault, M. L. Albert, and O. Schwartz. Characterization of reemerging chikungunya virus. *PLoS Pathogens*, 3(6):e89, 06 2007.

S.-D. Thiberville, N. Moyen, L. Dupuis-Maguiraga, A. Nougairede, E. A. Gould, P. Roques, and X. de Lamballerie. Chikungunya fever: Epidemiology, clinical syndrome, pathogenesis and therapy. *Antiviral Research*, 99(3):345 – 370, 2013. ISSN 0166-3542. doi: http://dx.doi.org/10.1016/j.antiviral.2013.06.009. URL http://www.sciencedirect.com/science/article/pii/S0166354213001666.

K. A. Tsetsarkin, D. L. Vanlandingham, C. E. McGee, and S. Higgs. A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog*, 3(12):e201, 12 2007. doi: 10.1371/journal.ppat.0030201. URL http://dx.plos.org/10.1371%2Fjournal.ppat.0030201.