Statistical Analysis of the Chikungunya Fever

Purvasha Chakravarti

Department of Statistics
Carnegie Mellon University

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 unpredented 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^oC and sometimes reaching 104^oC. 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 Runion 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 CHIKV 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 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 August 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.

To understand and predict the spread of the Chikungunya disease we model the infected case counts using SIR compartment models for the different countries. We also consider the travel between countries and incorporate infected people traveling from one to the another.

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. 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, 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`). Currently 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 are usually errors in the reported cumulative infected cases either due to misdiagnosed cases or miscounting. These errors are usually corrected in subsequent weeks. As a result of these corrections, sometimes the cumulative count reported decreases. For example on plotting the difference in the cumulative counts of consecutive weeks for Colombia and French Guiana, we notice in Figure 1, that the number of infected cases is negative at week 45 and week 30 for Colombia and French Guiana respectively. Since we do not know if the error was made the previous week or the current week, we just assume zero new cases in that week instead of negative count.

3 Methods

We have used two different types of models to model the number of infected cases of Chikungunya, namely a multi-country SIR model and a multi-country ARIMA model. The multi-

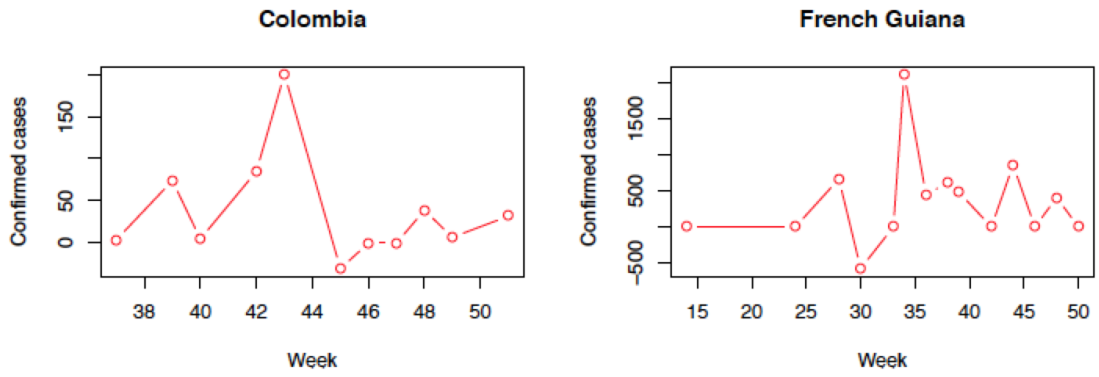

Figure 1: Confirmed new cases per epidemic week in Colombia and French Guiana. The count at week 45 for Colombia and at week 30 for French Guiana are negative due to error.

country SIR and ARIMA models have been discussed in subsections 3.1 and 3.3 respectively. We used the multi-country SIR compartment model to understand the spread of the disease. The model helps us understand whether Chikungunya will cause an epidemic. The multi-country ARIMA model was considered for a different reason, namely, in order to estimate the number of infected cases in the future. In this section, we also consider an extension of the simple SIR compartment model, given in subsection 3.2, that includes the mosquito population. As we do not have necessary data on the mosquito population we do not use this model. Despite this it is discussed in this section to exlain how multi-country SIR models can be extended to incorporate mosquito population.

In order to model the dynamics of the disease in this section, we account for infected people travelling from one country to another. In the SIR compartment 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 optimum model is found by minimizing the sum of squared errors in estimating new infected cases per week in all the countries. The data for the movement comes from flight itineraries. Currently we just assume the number of people traveling every week is a constant due to unavailability of data.

In the ARIMA modeling, the travel between the countries is incorporated by considering multi-variate ARIMA models. The number of infected cases of Chikungunya in each country is considered to be a variable. Hence considering a multi-variate model explains the influence that a country has on another country in speading the disease.

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. 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, $S + I + R = N$.

Now 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,

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

Now if β is defined as the contact rate, which takes into account the probability of getting the disease in a contact between a susceptible and an infectious subject then it becomes the proportionality constant in the above relation.

Now 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. γ is considered as the recovery rate, indicating the average proportion of infected people who recover every instant. Hence γ 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,

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

The people who recover just move to the Removed (R) compartment. Hence SIR models are usually defined by the following differential equations.

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta}{N}SI \\ \frac{dI}{dt} &= \frac{\beta}{N}SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{1}$$

where,
β is the contact rate,
γ 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,
N is the total population.

The basic reproduction number, $R_0 = \frac{\beta}{\gamma}$ is 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.

We model every country with a different compartment model and include travel between the infected compartments of different countries. Due to the unavailability of weekly travel data between the countries, we assume that the number of people who cross borders between a pair of countries is constant per week. We also assume that the populations of the countries remain 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 movement is homogeneous, that is, the ratio of people belonging to the different compartments among the people who cross borders is same as the ratio of people belonging to the compartments in the country. It is also assumed 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.

Therefore the cross-border SIR compartment model for countries i = 1,2,..,m is characterized by the following differential equations:

$$\begin{aligned}
\frac{dS_i}{dt} &= -\frac{\beta_i}{N_i} S_i I_i \\
\frac{dI_i}{dt} &= \frac{\beta_i}{N_i} S_i I_i - \gamma_i I_i - \sum_{j=1, j\neq i}^{m} r_{ij} \frac{I_i}{N_i} + \sum_{j=1, j\neq i}^{m} r_{ji} \frac{I_j}{N_j} \\
\frac{dR_i}{dt} &= \gamma_i I_i
\end{aligned} \tag{2}$$

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

CHIKV is transmitted by mosquitos but the cross-border SIR compartment model does not really take into account the mosquito population. To incorporate the mosquito population we could consider a compartment model which included mosquitos.

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 populations and between the mosquito and human population. It also assume constant population of the humans and mosquitos. The model is given by:

$$\begin{aligned}\frac{dI_H}{dt} &= abI_M\frac{N_H - I_H}{N_H} - \gamma I_H \\ \frac{dI_M}{dt} &= ac(N_M - I_M)\frac{I_H}{N_H} - \delta I_M\end{aligned} \tag{3}$$

where,
a is the mosquito biting rate,
b is the mosquito to human transmission probability, per bite
c human to mosquito transmission probability, per bite
γ human recovery rate: inverse of average duration of infection in humans,
δ 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 could consider a Ross-Macdonald model for each country and then incorporate the travel between the infected compartments of the countries. Then the differential equations for the system would be

$$\begin{aligned}\frac{dI_{Hi}}{dt} &= ab_iI_{Mi}\frac{N_{Hi} - I_{Hi}}{N_{Hi}} - \gamma_i I_{Hi} - \sum_{j=1,j\neq i}^{m} r_{ij}\frac{I_{Hi}}{N_{Hi}} + \sum_{j=1,j\neq i}^{m} r_{ji}\frac{I_{Hj}}{N_{Hj}} \\ \frac{dI_{Mi}}{dt} &= ac_i(N_{Mi} - I_{Mi})\frac{I_{Hi}}{N_{Hi}} - \delta_i I_{Mi}\end{aligned} \tag{4}$$

where the $a, b_i, c_i, \gamma_i, \delta_i, N_{Hi}, I_{Hi}, N_{Mi}, I_{Mi}$ are as defined in (3) for country $i = 1, 2, ..., m$. r_{ij} is as defined in (2) for the cross-border SIR compartment model.

Due to the lack of data on mosquito population, we do not use this approach for the results discussed in the next section.

3.3 Autoregressive Integrated Moving Average (ARIMA) Model

While the previously discussed compartment models used Differential equations, a different approach for modeling disease counts is an ARIMA model which uses data at previous time points to estimate the present. ARIMA models are used to fit time series data either to better understand the data or to predict future points in the series (forecasting). They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to reduce the non-stationarity (Box and Jenkins, 1990).

Non-seasonal 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. ARIMA models form an important part of the Box-Jenkins approach to time-series modelling.

Given a time series of data X_t where t is an integer index and the X_t are real numbers, then an ARIMA(p, d ,q) model is given by:

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t, \tag{5}$$

where L is the lag operator, the α_i are the parameters of the autoregressive part of the model, the θ_i are the parameters of the moving average part and the ε_t are error terms. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

The above can be further be generalized as follows.

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right) (1 - L)^d X_t = \delta + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \tag{6}$$

This defines an ARIMA(p,d,q) process with drift $\delta/(1 - \sum_{i=1}^{p} \alpha_i)$. ARIMA(p,d,q) are very useful for forecasting a time series. We use multivariate ARIMA models to explain the spread between the countries.

4 Results

4.1 Exploratory Data Analysis

The chikungunya epidemic started in the Americas in December, 2013. There have been a total of $61,282$ confirmed autochthonus cases in the Americas in a total of 97 epidemic weeks counting uptill November 6^{th}, 2015. As mentioned earlier, the case counts are updated cumulatively and sometimes due to manual errors, the counts are updated in the consecutive weeks. As a result of the updates, sometimes the cumulative counts decease in consecutive weeks instead of being non-decreasing. For example, we notice a sudden drop in the total cumulative confirmed cases in the Americas from Epidemic week 71 to 72, that is from May 8^{th} to 15^{th}, 2015. The counts drop 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. As the reason of the abrupt change is unknown, we assume that the cumulative counts were computed newly from Epidemic week 72. So we adjust for the change and add $31,223$ to all the cumulative counts henceforth.

On taking a difference of the cumulative counts to get the new confirmed autochthonus cases per week, it is seen that due to the adjustment, the new count of $8,790$ at week 72 is way higher than in any other week, see Figure 2. This implies that our assumption that a new set of cumulative counts were started at week 72 is false. To avoid complications and not lose too much information, we just assume that there were no new confirmed cases between week 71 and 72. The number of new confirmed cases from week 73 onwards are

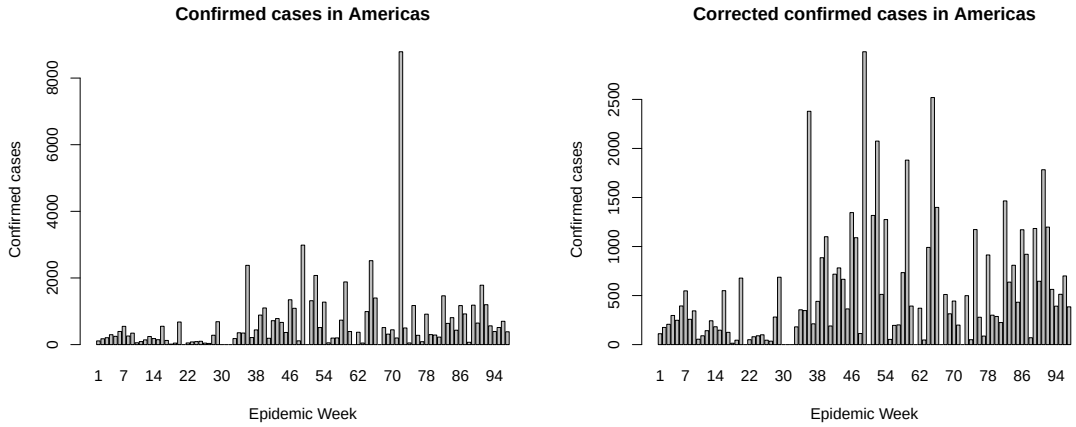

Figure 2: Total infectious subjects per week in Americas. We observe an anamoly at epidemic week 72 in the first figure due to an adjustment in the process of updating the cumulative number of confirmed autochthonus cases. The second one is the corrected version of the first and the corrected cumulative counts are the cumulative sum of the counts given in this figure.

Parameters	β	γ	R_0
Americas	1	0.9695	1.0314

Table 1: The estimates of parameters β, the contact rate, γ, the recovery rate and Basic Reproduction Number $R_0 = \frac{\beta}{\gamma}$ for the Americas.

assumed to be correct and then the new cumulative counts are taken under consideration. The corrected new confirmed cases per week can be seen in Figure 2.

4.2 SIR compartment model for the total counts in the Americas

We model the spread of chikungunya in the whole western hemishphere, i.e., in the Americas using an SIR compartment model. We select the optimal values of β and γ for the Americas by minimizing the log-sum of error squares using Nelder-Mead optimization algorithm. We start the algorithm at $I = 111, R = 0$ and $S = N - I$, where N is the population of Americas, which is currently $991,134$ thousand. Since the cumulative counts of the confirmed cases give the sum of the I and R compartments till the given week, the error is computed as the difference between the cumulative counts and the sum of estimated I and R from the model. We try different starting values for the parameters and select the optimum value with the minimum objective function. The optimum values can be seen in Table 1. The R_0 value is 1.0314 which is greater than 1, which implies that chikungunya will cause an epidemic in the Americas. Ebola's basic reproduction number was found to be 1.51 for Guinea, 2.53 for Sierra Leone and 1.59 for Liberia (CL, 2014).

The drawback of a compartment model is that though it can predict if a disease will be an epidemic or an endemic, it fits an exponential curve to the number of people in the infected compartment and so the prediction of the number of new cases is not very accurate, see Figure 3.

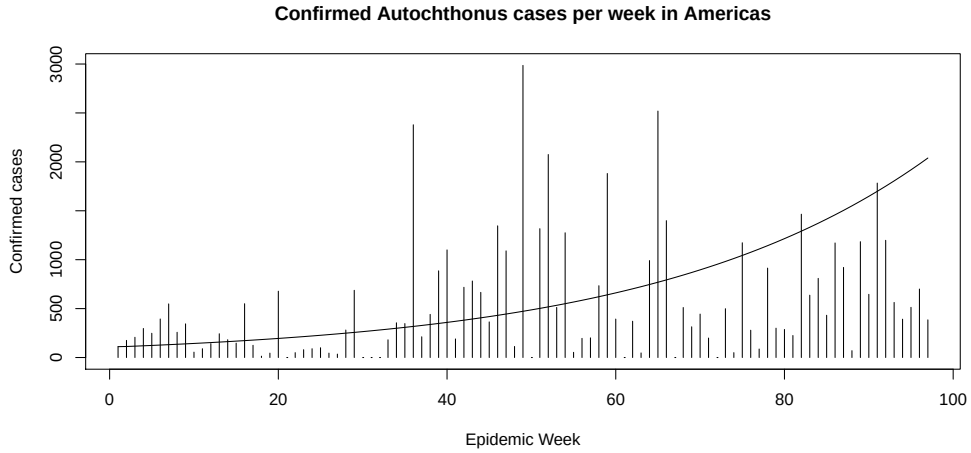

Figure 3: Predicted number of confirmed cases per week in Americas.

4.3 Multi-country SIR model for St. Martin and St. Barthelemy

The CHIKV 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 cross-border SIR compartment model given in (2), where number of countries $m = 2$. 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 approximately 30 people travel from one to the other per day. This number is obtained by looking at flight itineraries 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 case of the whole of Americas. 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.

Modeling the CHIKV transmissions in St. Martin and St. Barts using both the SIR model (1) and the cross-border SIR model, we get the values of β and γ for the two countries as given in Table 2. The estimates of the parameters do not vary too much between the SIR compartment model and the cross-border SIR compartment model. Interestingly, 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. This could be because 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

Compartment model	**St. Martin**			**St. Barthelemy**		
Parameters	β	γ	R_0	β	γ	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 2: The estimates of parameters β, the contact rate, γ, the recovery rate and Basic Reproduction Number $R_0 = \frac{\beta}{\gamma}$ for St. Martin and St. Barthelemy.

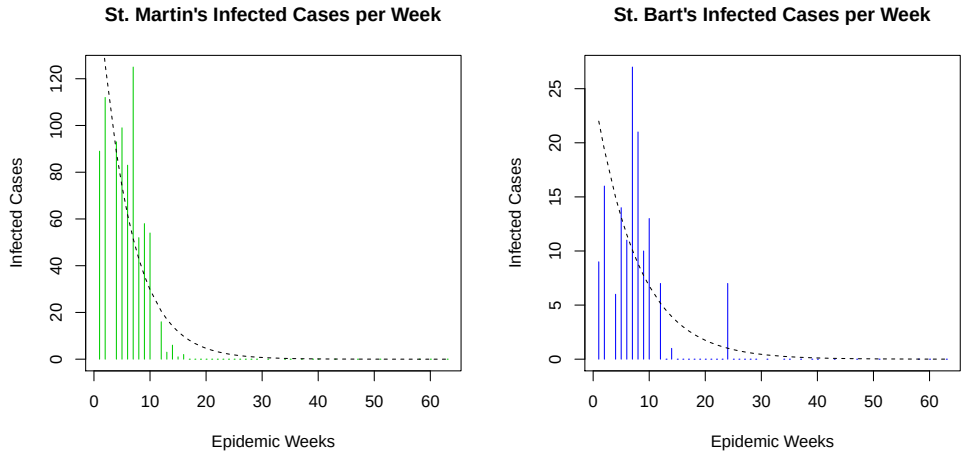

Figure 4: Infectious subjects per week in St. Martin and St. Barthelemy. The peak of the infectious period occurs at sixth week. In case of St. Barts the peak does not seem to have been captured.

infection indeed dies down in ten weeks in both countries. The number of infections peak in the sixth week but we can see that the CHIKV infection did not cause an epidemic in these two countries, see Figure 4. It is also noticeable that the peak is not captured very well by the model and hence, it would not serve as a very good forecasting model.

4.4 ARIMA model for predicting counts in Americas

In order to create a good fore-casting model, we consider ARIMA models. For fore-casting the total cumulative number of confirmed cases in whole of the Americas using an ARIMA(p,d,q) model as given in (6) (Section 2.4), 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 5).

To forecast the cumulative counts in the different countries, we could either fit an

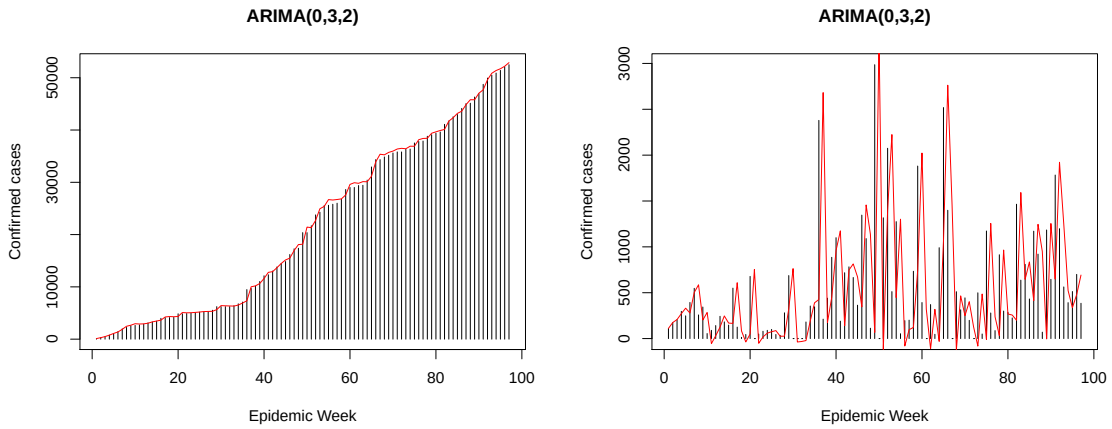

Figure 5: Predicted number of total cumulative confirmed cases and the total confirmed cases per week in Americas. The black dotted line with the predicted value. We notice that its much better than the fit of the SIR compartment model.

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 (7) (Section 2.5) 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 earlier we just have six 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 6 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. The model hence chosen is a multi-variate ARIMA(0,1,1). Similarly, the univariate ARIMA model chosen for just Purto Rico is ARIMA(2,3,1). The models are fit on the data before epidemiological week 90 and we use it to forecast the number of cumulative cases for the next 8 weeks. The number of confirmed cases for the weeks can then be found by taking a difference. In Figure 6 where 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 we are estimating.

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. 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 7. The plot shows us that other than a huge negative residual on

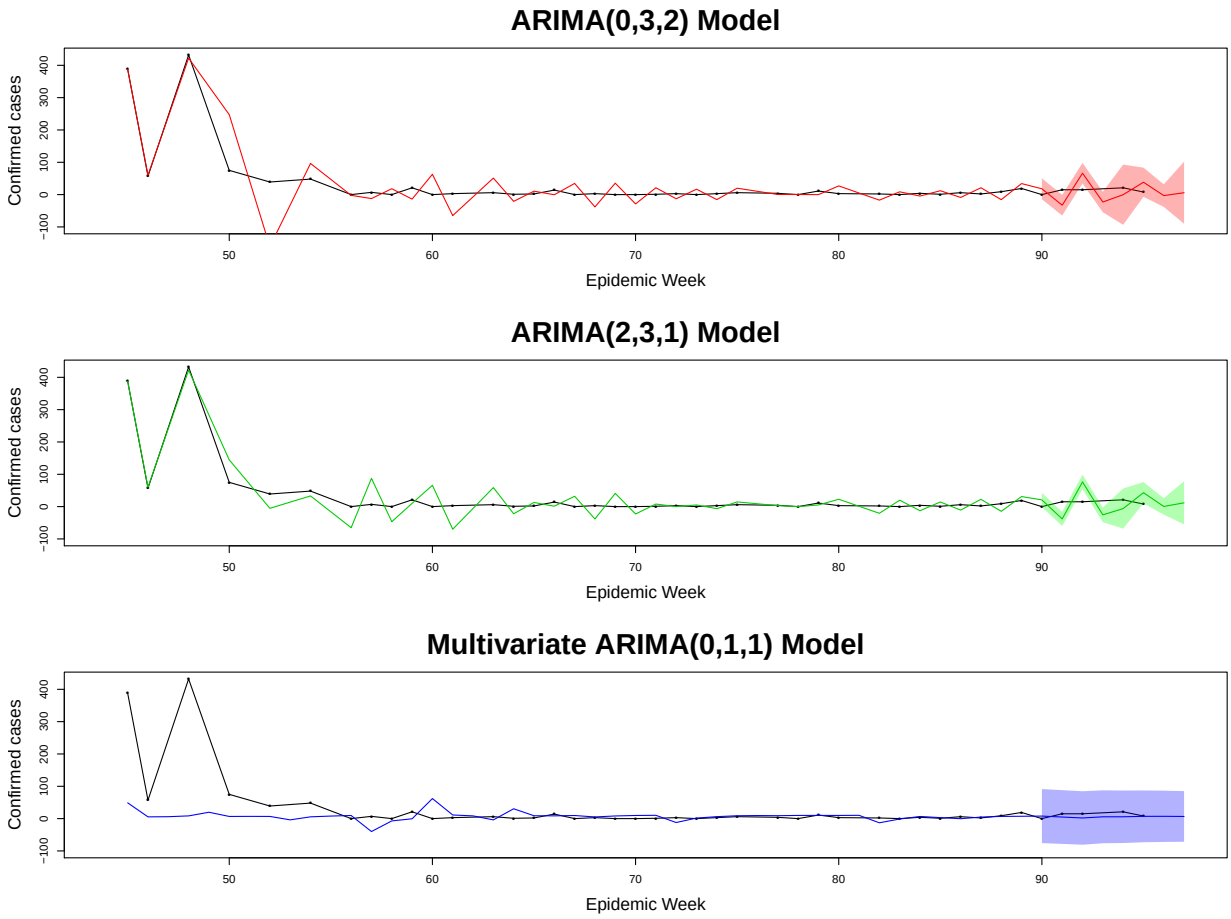

Figure 6: Comparing ARIMA(0,3,2), which was selected for the total cumulative counts in Americas, ARIMA(2,3,1), which was selected for Puerto Rico alone and multi-variate ARIMA(0,1,1). Observed confirmed counts are given by the black line. The predictions are made for epidemiological weeks 90 to 98, based on available previous data.

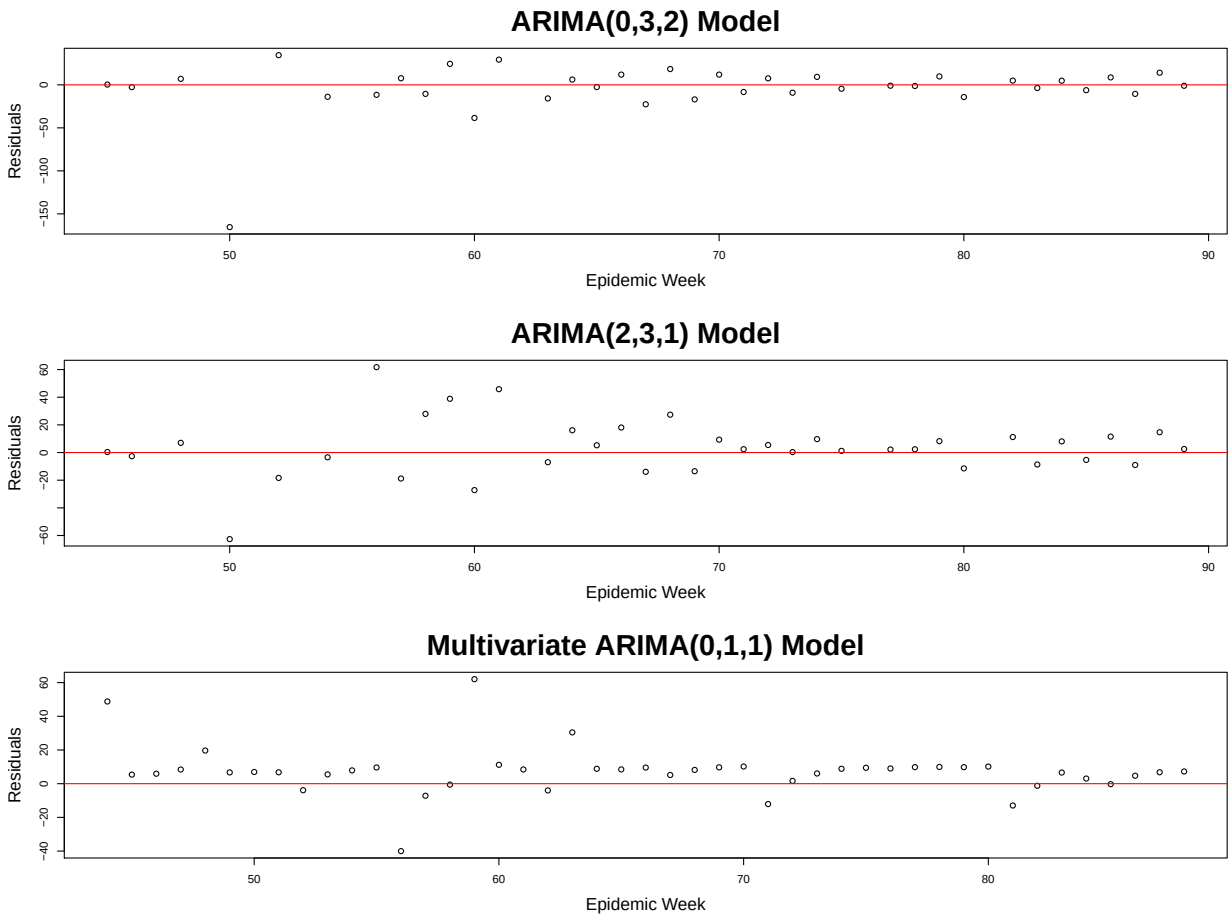

Figure 7: Comparing the residuals for Puerto Rico of ARIMA(0,3,2), which was selected for the total cumulative counts in Americas, ARIMA(2,3,1), which was selected for Puerto Rico alone and multi-variate ARIMA(0,1,1). The predictions are made for epidemiological weeks 90 to 98, based on available previous data.

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.

5 Discussion

The CHIKV 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 cross-border SIR compartment model to model the CHIKV 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 CHIKV transmission did not cause an epidemic in the two countries and died down after a while.

We plan to use the cross-border SIR compartment model and the SIR compartment model for all the countries in America. The data of travel between the different countries is not very easily available which makes it challenging to fit a cross-border 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). The estimates of mosquito populations could be used to fit Ross-Macdonald model.

References

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

Burt, F. J., Rolph, M. S., Rulli, N. E., Mahalingam, S., and Heise, M. T. (2012). Chikungunya: a re-emerging virus. *The Lancet*, 379(9816):662–671.

CDC (2006). Chikungunya fever diagnosed among international travelers–united states, 2005-2006. *MMWR Morb Mortal Wkly Rep*, 55(38):1040–1042.

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

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

CL, A. (2014). Estimating the reproduction number of ebola virus (ebov) during the 2014 outbreak in west africa. *PLOS Currents Outbreaks*, 1(12).

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

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

Kermack, W. O. and McKendrick, A. G. (1927). 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.

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

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

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

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

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

Powers, A. M. and Logue, C. H. (2007). Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *Journal of General Virology*, 88(9):2363–2377.

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

Schilte, C., Staikovsky, F., Couderc, T., Madec, Y., Carpentier, F., Kassab, S., Albert, M. L., Lecuit, M., and Michault, A. (2013). Chikungunya virus-associated long-term arthralgia: A 36-month prospective longitudinal study. *PLoS Negl Trop Dis*, 7(3):e2137.

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

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

Sourisseau, M., Schilte, C., Casartelli, N., Trouillet, C., Guivel-Benhassine, F., Rudnicka, D., Sol-Foulon, N., Roux, K. L., Prevost, M.-C., Fsihi, H., Frenkiel, M.-P., Blanchet, F., Afonso, P. V., Ceccaldi, P.-E., Ozden, S., Gessain, A., Schuffenecker, I., Verhasselt, B., Zamborlini, A., Saïb, A., Rey, F. A., Arenzana-Seisdedos, F., Desprès, P., Michault, A., Albert, M. L., and Schwartz, O. (2007). Characterization of reemerging chikungunya virus. *PLoS Pathogens*, 3(6):e89.

Thiberville, S.-D., Moyen, N., Dupuis-Maguiraga, L., Nougairede, A., Gould, E. A., Roques, P., and de Lamballerie, X. (2013). Chikungunya fever: Epidemiology, clinical syndrome, pathogenesis and therapy. *Antiviral Research*, 99(3):345 – 370.

Tiao, G. C. and Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76(376):802–816.

Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E., and Higgs, S. (2007). A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog*, 3(12):e201.