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, Figure 1. Since we do not know ~~if the~~ when the error was made t~~he previous week or the current week~~, we just assume zero new cases in that e ~~week instead of~~ week with negative count.

3 Methods

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

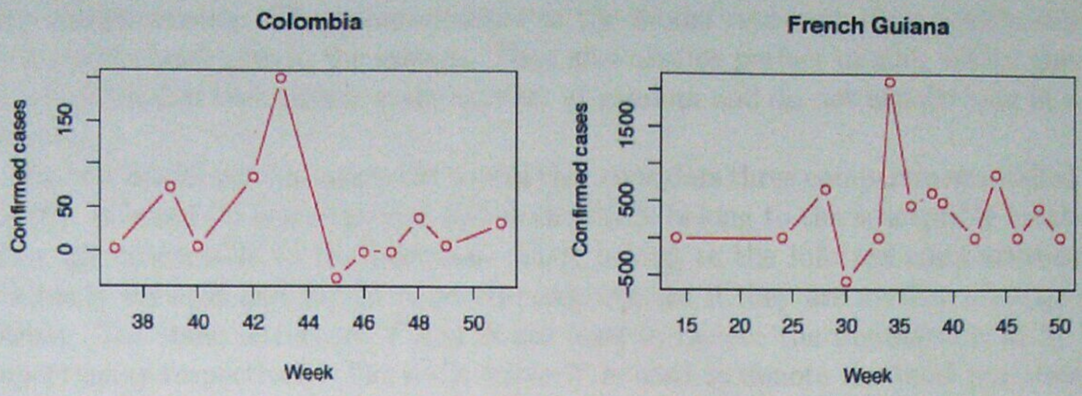

Figure 1: Confirmed new cases per epidemic week in Colombia and French Guiana. The count at week 45 for Colombia and at week 30 for French Guiana are negative due to error.

country SIR and ARIMA models ~~have been~~ will be discussed in subsections 3.1 and 3.3 respectively. We used the multi-country SIR compartment model to understand the spread of the disease. 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 ~~for a different reason, namely,~~ in order to estimate (predict ?) the future number of infected cases ~~in the future.~~ In this section, we also consider a theoretical extension of the simple SIR compartment model, ~~given in~~ subsection 3.2, that includes the mosquito population. Nevertheless ~~As~~ we do not have necessary data on the mosquito population but we do not use this model investigate the theor properti. ~~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~~ is modeled by accounting for infected people travelling from one country to another. ~~In the SIR compartment modeling,~~ for SIR model impl. 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~~ of estimating new infected cases per week in all the countries. The data for the movement comes from flight itineraries, but not all of them are available therefore ~~Currently~~ we ~~just~~ assume the number of people traveling every week ~~is a~~ constant ~~due to unavailability of data.~~

Idk

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~~ the 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

IN THE ARIMA MODELING WE INCLUDED A VARIABLE FOR COUNTING THE #... COUNTRY AND WE INCORPORATED THE INFORMATION ABOUT TRAVELING BETWEEN COUNTRIES USING A MULTIVAR. STRUCT. HENCE...

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, $S + I + R = N$.

According to this assumption ~~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,

$$(1) \quad \frac{dS}{dt} \propto -\frac{SI}{N}. \quad (\text{numb})$$

Letting be Lebrum use of pronouns

~~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~~ β becomes the proportionality constant in ~~the above relation~~. equation ( )

~~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,

$$\frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I. \quad (2)$$

eq (3)

(1, 2, 3)
Combining the eq above

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

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

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 ~~also~~ assume that the 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~~ Lastly, we 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:

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

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 have also considered 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~~ have been :

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

[illegible] or pdf

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~~ apply 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:

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

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,

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

Eq. (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 of the disease between the countries.
across

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 6th, 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 8th to 15th, 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