lion 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 Methods and Materials

According to the PAHO guidelines published in 2[1] countries affected by Chikungunya in the Americas, 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 indigenous 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. PAHO maintains the weekly record of the cumulative counts for all the countries in Americas on their website (`www.paho.com`).

PAHO also compute the incidence rate of the disease in every country. The incidence rate is the number of confirmed autochthonous transmissions per hundred thousand population. 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.

As PAHO reports the cumulative infected cases per week for each country and not the number of new cases per week, sometimes due to error 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

Summary of Comments on Chikungunya_V3_Comments.pdf

Page: 3

Author: loaner Subject: Sticky Note Date: 1/29/2016 9:32:47 AM
Not a good line for topic position. Perhaps putting this at the end as a second clause might be more clear.

Author: loaner Subject: Highlight Date: 1/29/2016 9:00:47 AM
I would include this part as a topic sentence for next paragraph. It doesn't add much to the existing line of argument.

Author: loaner Subject: Highlight Date: 1/29/2016 9:04:36 AM
what kind of error? Also since this paragraph expans more on how the error in reporting affects data, you could use this in the topic position of this paragraph.

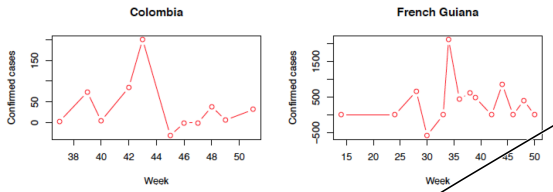

Figure 1: Confirmed new cases per epidemic week in Colombia and French Guiana. The count at week 45 for Colombia and at week 30 for French Guiana are negative due to error.

just assume zero new cases in that week instead of negative count.

To model the dynamics of the disease we use a different SIR compartment model for each country and allow for travel 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.

2.1 Multi-Country SIR Compartment Model

Compartment models which are founded upon differential equations are one of the most commonly used methods for modeling epidemics. The method was introduced by Kermack and McKendrick in the early 1900s (Kermack and McKendrick, 1927). These models serve as a basic mathematical framework for understanding the complex dynamics of diseases. They consider the population to be homogeneous mixture of people who are divided between compartments which 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 mix mostly in a smaller subgroup. The SIR model just considers three compartments 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. SIR models are usually defined by the following differential equations:

Page: 4

Author: loaner Subject: Highlight Date: 1/29/2016 9:33:55 AM
transmission?

Author: loaner Subject: Highlight Date: 1/29/2016 9:34:09 AM
big separation between subject and verb here.

Author: loaner Subject: Sticky Note Date: 1/29/2016 9:12:29 AM
they = Kermack and McKendrick?

Author: loaner Subject: Sticky Note Date: 1/29/2016 9:11:39 AM
split into two sentences? in the second part after which I think you talk about the compartments, right?

Author: loaner Subject: Highlight Date: 1/29/2016 9:35:25 AM
The three compartments that the SIR model considers are ....? (is this what you meant to say here? It wasn't quite clear!)

Also this might be a good place to start a new paragraph, now that you are switching from a general compartment model to a specific SIR moidel and talk about how the individuals are grouped under this model.

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

where,
β is the contact rate, which takes into account the probability of getting the disease in a contact between a susceptible and an infectious subject,
γ is the recovery rate, which is inverse of the average duration of the infection,
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. We just use a single number for the number of people crossing borders per week between a pair of companies due to the unavailability of weekly travel data between the countries. We 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. Hence we incorporate the travel between infected compartments of the different countries. 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. We also assume that 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:

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

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.

Page: 5

Author: loaner Subject: Sticky Note Date: 1/29/2016 9:21:53 AM
seems like you are talking about a solution to a problem you talk in the end of the sentence. Why not present the problem first and solution in the following sentence?

Author: loaner Subject: Highlight Date: 1/29/2016 9:36:07 AM
repititive.

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.

2.2 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.

2.3 Multi-Country Ross-Macdonald Model

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.

This page contains no comments

2.4 Autoregressive Integrated Moving Average (ARIMA) Model

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 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.

3 Results

The chikungunya epidemic started in the Americas in December, 2013. There have been 61,282 confirmed autochthonus cases in the Americas in a total of 97 epidemic weeks counting uptill November 6^{th}, 2015. As mentioned earlier, due to the process in which the counts are updated, we notice a sudden drop in the 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. This is most likely because of a change in the process of updating the cumulative counts. So assuming that the cumulative counts were computed newly from Epidemic week 72, we adjust for the change and add 31,223 to all the 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

Page: 7

Author: loaner Subject: Sticky Note Date: 1/29/2016 9:46:14 AM

It is not clear to me from this section, what approach you actually used and what you did not? Also why are you talking about the models here when it could just as well be in literature review? Is it a motivation for something you are going to do later?

It is important that you clearly outline the methods you will be using and others that exist but you will not use, and why you made the decision of using certain approach. I think the all the materials are there in this section, but they are hidden and it is hard for a reader to uncover them.