Scientific Computing with Python

Group Project Report

Midterm Checkpoint - 11/13/2020

The SIR Model of Disease Spread: Simulations and Possible Extensions

Cole Irwin

Daniel Gold

Shruthi Ramesh

1 Introduction

In this report, we run several simulations of the SIR Model of disease spread. We use two variations of the model: the agent-based (discrete) model and the ODE Model.

For the Discrete Model, each person belongs to one of three categories - susceptible (prone to get the disease), infected (currently has the disease), and recovered (is either immune to the disease after being infected or has died). We start with a population with about 0.1 percent infected people. In this simple model, we assume that everyone interacts with everyone else. Every infected person could spread the disease to the susceptible population with probability b at a given time period. Similarly, every infected person recovers from the disease with probability k at every time period. It is to be noted that the three states are mutually exclusive and collectively exhaustive, and the population is fixed at a constant N.

The ODE system at hand is defined as below:

$$\dot{x} = -bxy \tag{1}$$

$$\dot{y} = bxy - ky \tag{2}$$

$$\dot{z} = ky \tag{3}$$

where x(t) is the proportion of the population that is susceptible (S) to the disease at time t, y(t) is the proportion of the population that is infected (I) by the disease at time t, and z(t) is the proportion of the population that is removed (R) at time t. The parameters b, k > 0 are the infection and removal rates, respectively. The first parameterizes the rate at which individuals go from susceptible to infected, and the second parameterizes the rate that individuals go from infected to removed. There is also related parameter, R_0, which is defined by b and k. $R_0 = b/k$ is typically what is referred to as the 'basic reproduction number' Weiss (2013). This is because it is estimating the number of people an infected person will infect (based on b) before they recover (based on k).

We find that our equilibrium solutions form only when y = 0. Specifically, there are two instances; y goes to 0 because there is nobody left to infect ($x^* = 0$, $z^* = 1$), or the virus is removed from the

population before everyone is infected ($x^* > 0$, $0 < z^* < 1$). Note that $\dot{y} = 0$ when y is zero and also when $x = \frac{k}{b}$ ($x = \frac{1}{R_0}$). With regards to the latter, we can also see that when $x > \frac{k}{b} = \frac{1}{R_0}$, $\dot{y}$ is positive, noting an increase in infections. Conversely, when $x < \frac{k}{b} = \frac{1}{R_0}$, $\dot{y}$ is negative, noting an decrease in infections. This helps us analytically construct the maximum of infections at the point where x = $\frac{k}{b} = \frac{1}{R_0}$, as well as the infection curve at large.

2 Structure of sir Package

We have two files in the sir package. The first file 'discretemodel.py' contains the class and relevant functions to create agents and run simulations for the discrete SIR model. The second file 'ODE_Function.py' contains functions pertaining to the ODE model. Our scripts for running various simulations using these classes and functions can be found in the 'scripts' directory, which has also been divided into two parts (one for each model).

3 Preliminary Results

3.1 Discrete Model

In this section, we introduce the discrete agent-based model. This model is parameterized by b, k, T and N, where b corresponds to the likelihood of disease transmission and k is the likelihood of recovery for a given time period. T and N correspond to the time horizon and number of agents, respectively.

In our current configuration, each agent interacts with all other agents per time period. Thus, b can be thought of as being comprised of two elements: the probability, p, of transmission, and the proportion of the population they are interacting with $0 \leq n \leq 1$. Thus, we have $b = p \cdot n$.

We expect that, with higher values of b, the entire population might become infected. This is expressed in the model with $b = 0.0001$ and $k = 0.1$ seen in Figure 1, wherein the entire population becomes infected and recovers within 600 iterations of the model. However, this is highly unlikely in reality.

Another corner example is expressed in Figure 2, which shows the model parameterized by $b =$

0.0001 and $k = 0.1$. In this model, the person who is infected recovers before others in the population, so nobody becomes sick.

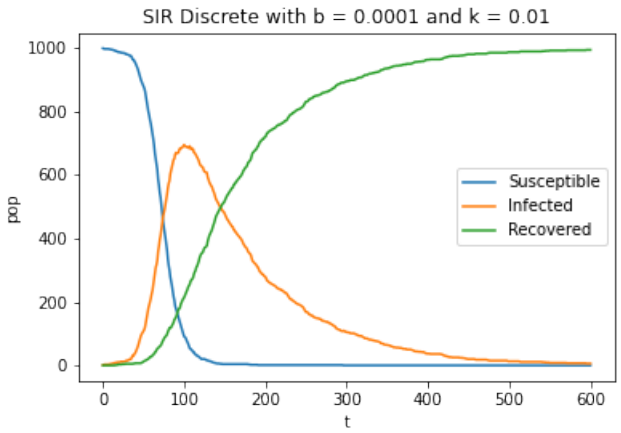

Figure 1

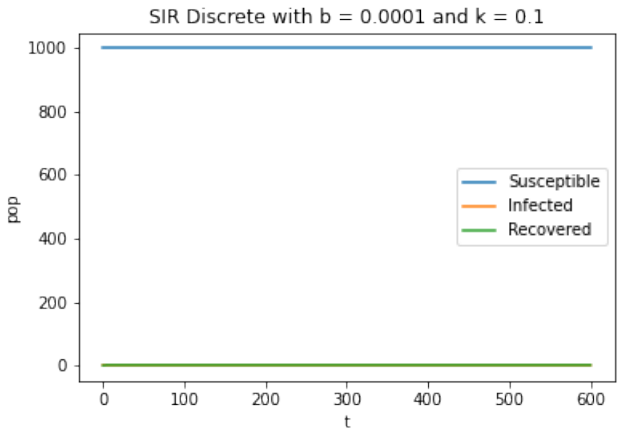

Figure 2

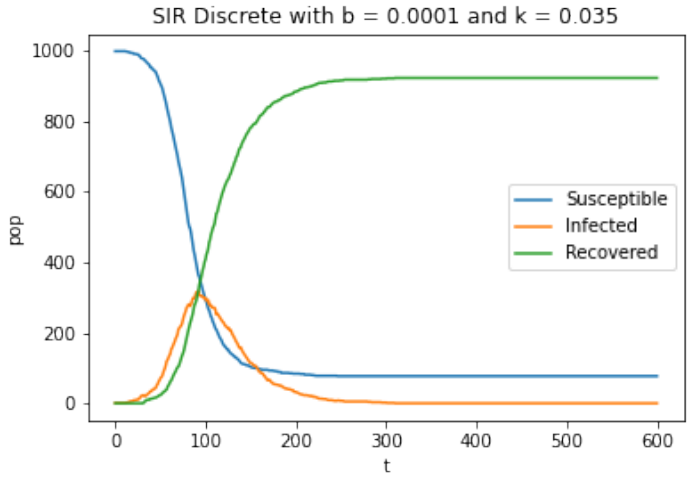

Figure 3

Of greater interest are models wherein only some segment of the population is infected and recover, as this seems to be more 'realistic.' This can be thought of as a wave of infection that ultimately leads to herd immunity in the population. A model paramaterized by $b = 0.0001$ and $k = 0.035$ describes this more realistic scenario, pictured in Figure 3.

Finally, we consider the phase diagram in 4, which can be used as an aide to choose interesting values of b and k to consider. As k increases, we see that the infected population decreases for a given b value decreases and vice versa, as expected.

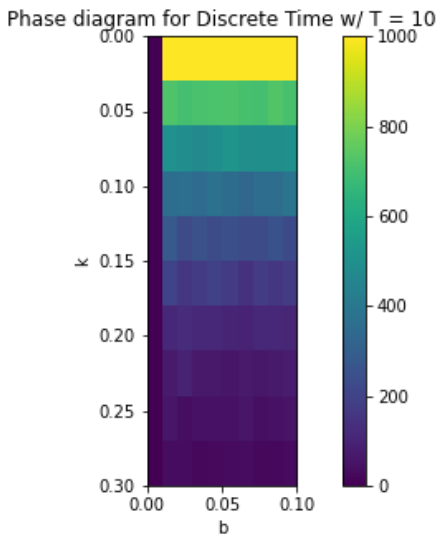

Figure 4

3.2 ODE Model

We set out to run numerical simulations on the system described in (1), (2), and (3) above. Our simulations were ran with the initial condition where .01 percent of the population has the infection at time 0. This equates to below:

$$SIR_0 = [x_0, y_0, z_0] = [0.99999, 0.00001, 0]$$

Note that the condition we are using here to describe the types of behavior of our system is with regards to the proportion of the population that is removed after 600 days. This is because there is a parametric range (namely, when $k > b$) such that the infection curve is null (infections go away immediately), as well as a parametric range (namely, when $3k < b$ such that everyone gets infected and the removed population tends towards 1.

See below for our phase diagram found by looping over our parametric range of b, k:

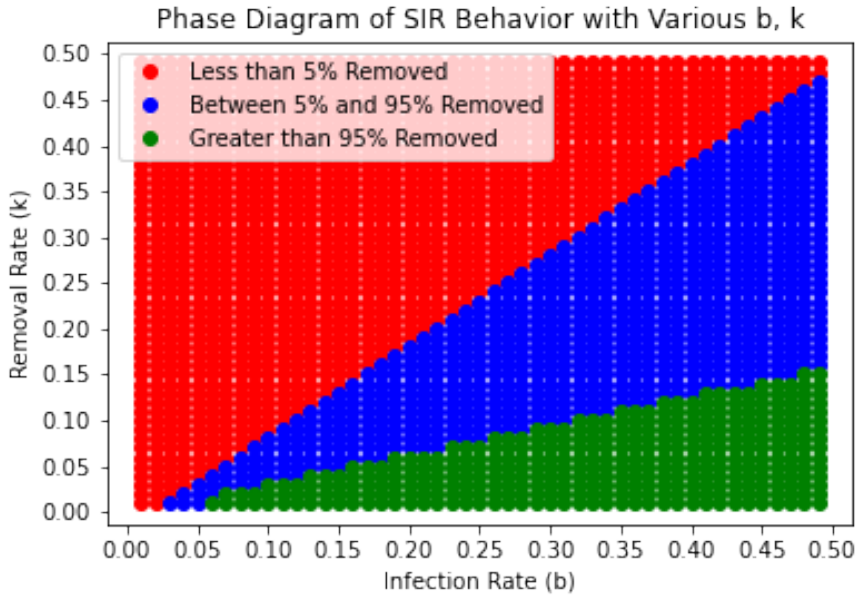

Figure 5: We can see a line along b=k, above which disease does not spread (red). We can also see another line along b = 3k, below which the infection spreads rapidly and infects the whole population (green)

A final remark regarding this exploration of b and k in our ODE model is as follows. A high infection rate b yields a faster introduction of the virus into the greater population. A high removal rate k is the only thing that can stop the virus from infecting the entire population. Parametric combinations b, k that yield the same R_0 yield similar maximum infected proportions as well as long term removed populations, however the magnitude of b, k will adjust how fast the infection sweeps through the population. In general, a lower R_0 will yield a smaller proportion of the population will have to be removed and that the maximum infected proportion will be smaller. The opposite applies for the converse - a greater R_0 yields a higher amount of removed population and larger maximum infected population. See below for a few results highlighting what our 3 SIR solution curves look like over time (days) with various parameter combinations set:

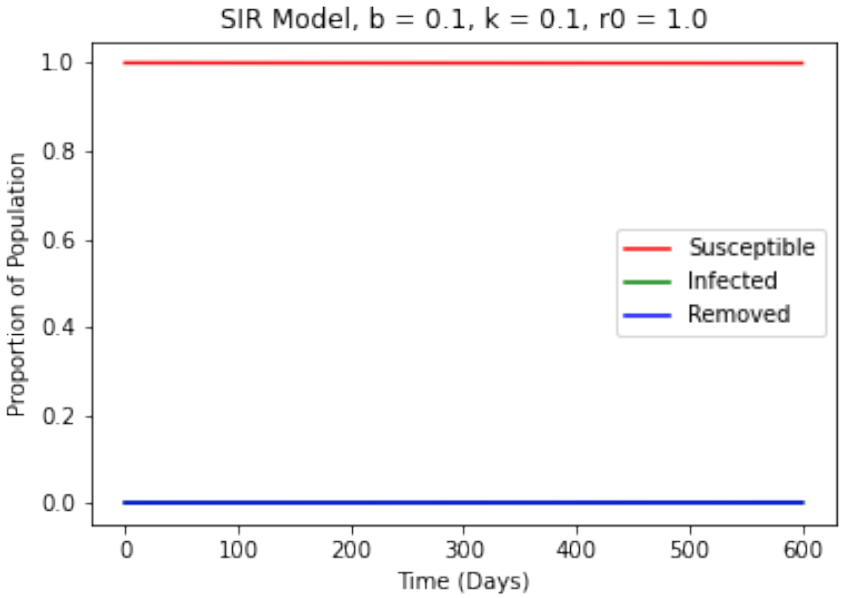

Figure 6: When $\frac{k}{b} = \frac{1}{R_0} \geq 1$ (as shaded in red in the Phase Diagram above) we have that the infected population immediately tends towards zero, leaving a susceptible proportion of the population ≈ 1 and a removed proportion of the population ≈ 0.

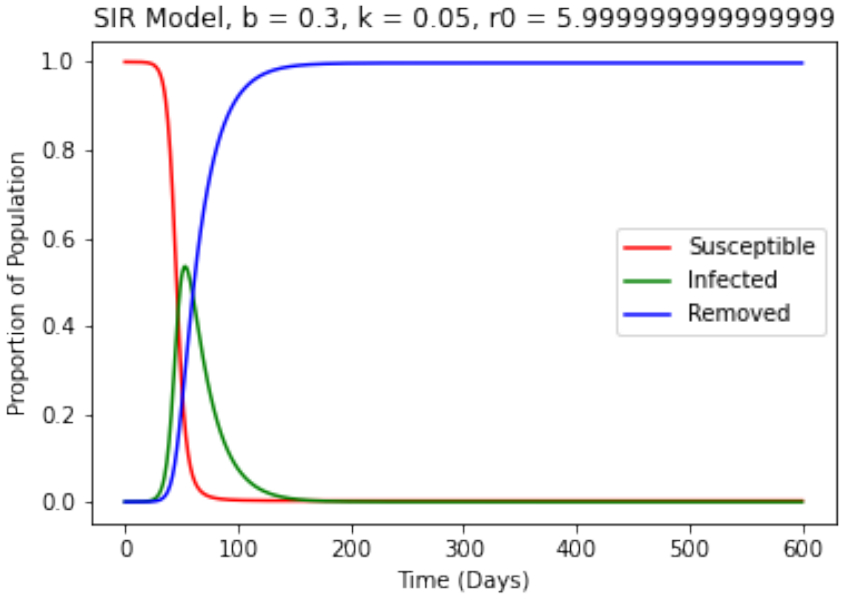

Figure 7: When $R_0 = \frac{b}{k} \geq 3$ (as shaded in green in the Phase Diagram above) we have that the entire population contracts the disease, and hence the removed proportion of the population tends towards one, leaving a susceptible proportion of the population ≈ 0.

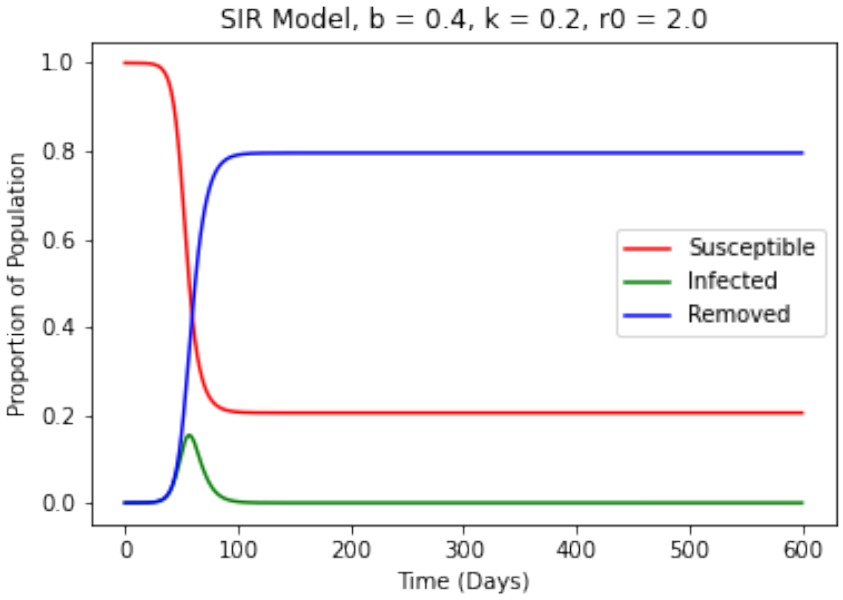

Figure 8: With these parameters, it takes 57 days to reach the maximum active cases, where 15.4% of the population was infected. Eventually, the virus goes away after 91 days total, in the process removing 79.5% of the population.

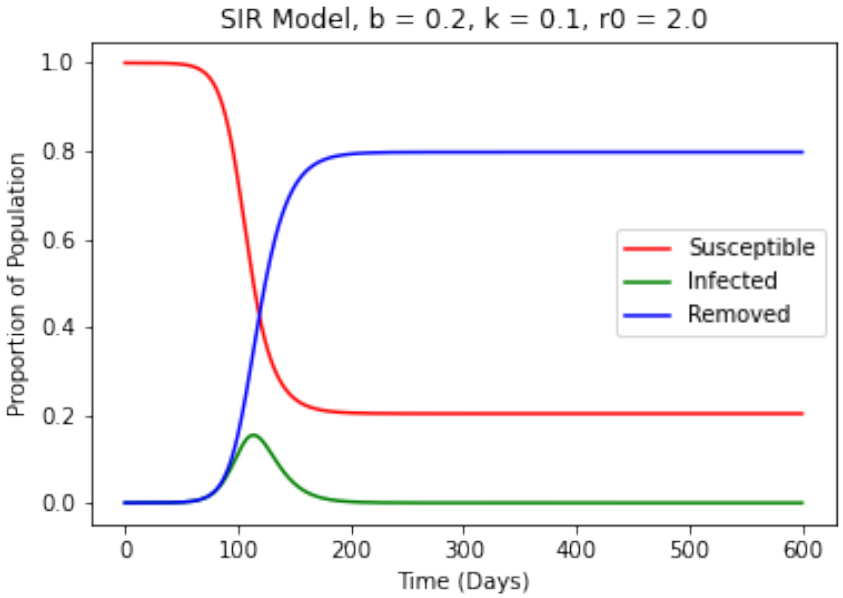

Figure 9: With these parameters, it takes 114 days to reach the maximum active cases, where 15.4% of the population was infected. Eventually, the virus goes away after 180 days total, in the process removing 79.7% of the population. Note that this has the same R_0 as the previous figure, but because the parameters are both smaller in magnitude, the disease spread and recovery is more spread out over time.

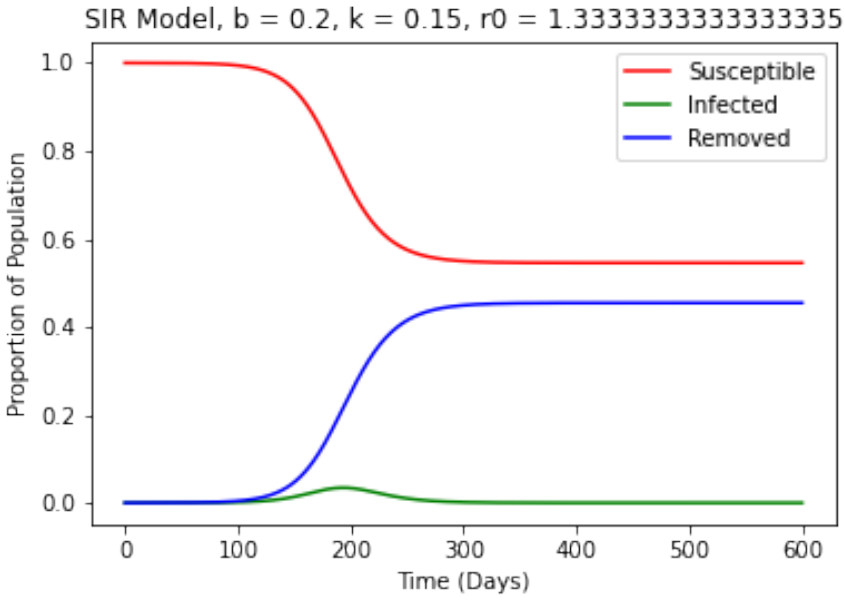

Figure 10: With these parameters, it takes 193 days to reach the maximum active cases, where 3.4% of the population was infected. Eventually, the virus goes away after 231 days total, in the process removing 35.5% of the population.

3.3 Comparison between the Models

Based on the visualizations above, we find that our SIR curves share similar attributes. In both discrete and ODE settings, the susceptible population is strictly non-increasing, the recovered population is strictly non-decreasing. At all points of the discrete system, S + I + R = N (the total population), meanwhile at all points of the ODE system, x(t) + y(t) + z(t) = 1 (100% of the population).

When we run both settings with a parametric combination b, k such that the infection curve is non-trivial, we find that the virus is eradicated after a similar amount of days - between 200 and 350 days on average. This said, the parameters b, k are not equal between our two methods. This can be visualized by the differences in our phase diagrams. In our discrete solution, our infection rate parameter b must be small to allow our infection curve to build up to a defined infection peak. On the other hand, the ODE solution is not as sensitive to the magnitude of b and k. In both cases, the relationship between b and k strongly establish the behavior of our systems.

4 Suggested Extensions

4.1 The Stochastic SIR Model

4.1.1 Motivation

The mechanisms behind the spread of infectious diseases has become an important research question now more than ever, with the coronavirus pandemic. The SIR model has proven useful and informative in ascertaining the rate of growth and the size of an epidemic. However, the model ignores the apparent stochastic nature of epidemic growth and spread; individuals are expected to interact with a random number of people everyday. This group would comprise of a core subgroup (like family) and a random group whose number changes everyday (like a flight attendant). Therefore, a simple extension of the SIR model would be to include a probabilistic element of social interaction i.e. to modify it as a simple discrete time stochastic epidemic model. Although this is a simple extension, it could reveal some underlying mechanisms of disease spread that the simple SIR model fails to capture.

4.1.2 The Model

The time stochastic element can be introduced through a chain-binomial model. This means that individuals remain infectious for R time units, after which they become removed or immune. Individual transition probabilities from susceptible to diseased states are given in terms of the binomial distribution. The discrete time and discrete state space property would imply that the model is Markovian. Moreover, the total population is constant and individuals meet a random number of other individuals at each time step. This model is based on Tuckwell and Williams (2007). A thorough explanation will be given in the final report.

4.1.3 Implementation Logistics

This extension does not rely on data; it is based on running simulations on a modified SIR model. This would entail modelling the binomial distribution in a Markovian state space, and to run multiple trials for every set of parameters to analyze the distribution of variables like the number of infectives and the duration of the pandemic.

4.2 Spatial Considerations

4.2.1 Motivation

By nature, people are necessarily located in space. So far, we have considered a model in which the location of an agent does not determine their likelihood of getting sick. For example, in our population array, an agent interacts with all other agents without consideration of their position (in the case of our current model, a single dimensional array). In reality, the locational similarity of agents is highly important in determining their likelihood of getting sick. As this project is being completed, for example, Covid-19 rages in the Midwest while coastal cities remain far less encumbered by the virus.

Further, the notion of space can be extended to describe other similarities, such as ideology or the behaviour of certain age demographics. For this section of the project, we will incorporate notions of space to add new dimensions to our model.

4.2.2 The Model

For the discrete model, we will consider the location of agents on a 2-D array as being a determinant of the likelihood of their getting sick. In order to make this more realistic, we will consider both sparse and dense population types and the degree to which population density affects viral transmission, which will naturally give rise to consideration regarding sparse matrix computation. Further, we will attempt to extend the idea of space to other interpretations, such as those listed above.

4.2.3 Implementation Logistics

This part of the project will make extensive use of Anselin (2013) in consideration of spatial statistics and theory, as well as Perez and Dragicevic (2009), which discusses the incorporation of spatial considerations into agent-based modeling in the context of infectious disease.

4.3 Fitting with Covid/Other Disease Data

4.3.1 Motivation

COVID-19 is an ongoing pandemic. We seek to understand the virus more by estimating parameters b, k from the active cases curves of various countries. From this, we can understand that, given the same virus is global, the parameters are then differing due to data quality/standards, as well as country-specific reaction and social measures (shut downs, wearing masks, etc). We also seek to compare this outbreak to the historical SARS 1 outbreak of 2003, particularly in China.

4.3.2 The Model

We would look to estimate parameters b and k from the exponential growth and decay of the measured active cases curves of the COVID-19 virus from various countries with respect to time (days). From these sampled countries' b and k, we can construct a b and k range, from which we can explore. We can then pull virus data from historical datasets relating to the SARS 1 virus from 2003 and discuss the implications of the differing parameters as they relate to attributes of the virus. For example, SARS1 typically yielded infected people becoming symptomatic before being able to infect other people - this can be parameterized in our model as a factor of the infection rate, because there was easier action that could be taken so as to not infect others.

4.3.3 Implementation Logistics

We would pull COVID data per country from *Worldometers.info* (N.d.). We might not be able to find clean dated active cases for the SARS1 virus. For that, we may have to look at general data points of total number cases (total removed) and maximum cases and back-construct a curve from our ODE system to estimate parameters b, k.

References

Anselin, Luc. 2013. *Spatial Econometrics: Methods and Models*. Vol. 4 Springer Science & Business Media.

Perez, Liliana and Suzana Dragicevic. 2009. "An agent-based approach for modeling dynamics of contagious disease spread." *International journal of health geographics* 8(1):50.

Tuckwell, Henry C and Ruth J Williams. 2007. "Some properties of a simple stochastic epidemic model of SIR type." *Mathematical biosciences* 208(1):76–97.

Weiss, Howard Howie. 2013. "The SIR model and the foundations of public health." *Materials matematics* pp. 0001–17.

Worldometers.info. N.d. https://www.worldometers.info/coronavirus/.