

# Preliminary Report

Yujin Liu, Helen Jin, Jiangmeng Li

CAAM37830

University of Chicago

November 13, 2020

---

# 1 Introduction to the SIR Model and Notation

The susceptible-infected-removed (SIR) model is built based on the first-order derivative to model the spread of infectious disease, where the time dependent variables  $S$ ,  $I$ ,  $R$  each represent the following populations:

$S$ (susceptible): number of individuals who are not infected but could become infected

$I$ (infected): number of individuals who are already infected and can spread disease

$R$ (removed): number of individuals who are either recovered and immune or have died

Additionally,  $s$ ,  $i$ ,  $r$  are used to represent the proportion of susceptible, infected and removed individuals among the population. Note that this model assumes the susceptible population  $S$  decreases monotonically towards 0, and the population size is fixed, and the duration of infectivity is same as length of the disease.

There are two parameters  $b$  and  $k$  in the model, where  $b$  indicates the number of interactions each day that could spread the disease (per individual) and  $k$  indicates the fraction of the infectious population which recovers each day. The model consists of the following system of the nonlinear ordinary differential equations, where  $t$  is time:

$$\frac{ds}{dt} = -b * s(t) * i(t)$$

$$\frac{dr}{dt} = k * i(t)$$

$$\frac{di}{dt} = b * s(t) * i(t) - k * i(t)$$

## 2 Python Package 'sir'

## 3 Results of Simulations

To investigate how a new disease spreads, we run the simulations of both the discrete agent-based model and the continuous ODE model, starting with 100 infectious individuals out of the total population 20000.

### 3.1 Discrete Agent-Based Model

We generate several plots to show how  $s$ ,  $i$ , and  $r$  change over the length of the simulation for a couple different choices of  $k$  and  $b$ . First we check the plots for  $b = 2$ , which means we assume each individual has two interactions each day that could spread the disease, with different  $k$  values.

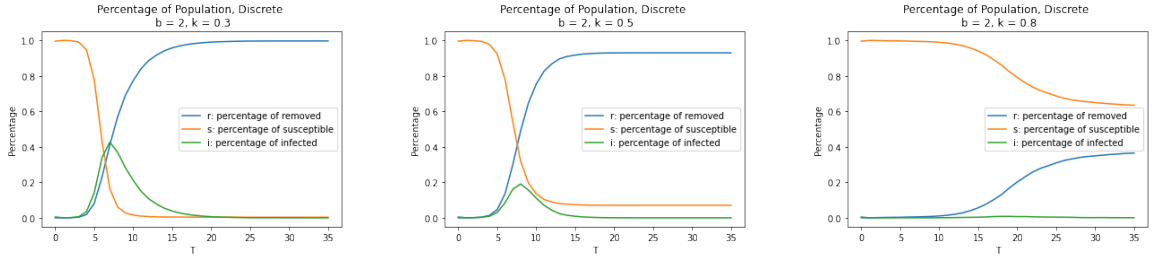


Figure 1: Percentage of population susceptible ( $s$ ), infected ( $i$ ), and removed ( $r$ ), discrete

From the plots, we find that when the recovery rate is low (0.3), finally all people are removed, which indicates that all people were infectious and either have recovered and are now immune, or have died. Within the first 20 days, the percentage of infected population increases then decreases, and the percentage of susceptible population decreases dramatically to 0, while the percentage of removed population quickly increases to 1. As the recovery rate increases to a high value (0.8), the trends change obviously. The percentage of infected people does not increase, and the percentage of removed people increases to approximately 0.4 while the percentage of susceptible people decreases to 0.6. Finally, both trends of removed and susceptible people stays constant. In this case, there are fairly large proportion of individuals who are never infected. Thus, when the recovery rate  $k$  is large enough (larger than 0.6) and  $b$  is relatively small (less than 3), then the herd immunity is eventually achieved. The herd immunity occurs when a large portion of a community becomes immune to a disease, making the spread of disease unlikely, so the whole community becomes protected. Next, we increase the value of  $b$  and see whether the trends change.

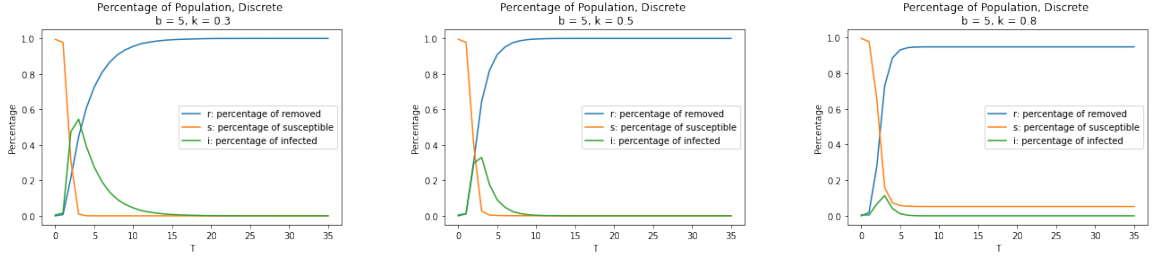


Figure 2: Percentage of population susceptible ( $s$ ), infected ( $i$ ), and removed ( $r$ ), discrete

When  $b = 5$ , each individual has five interactions each day that could spread the disease. Figure 2 suggest that if the recovery rate is 0.3, then finally all people are removed after 20 days; if the recovery rate is 0.5, all people are removed after 10 days; even if the recovery rate is 0.8, less than 10% of individuals will remain susceptible after 10 days. As a result, when  $b$  is large enough, no matter how large  $k$  is, almost all people will be removed eventually, which means that all people will be infected and left with different results. Hence, people should avoid contacting in person with infectious people. Some suggestions include staying at home as much as possible and avoiding visiting areas where the disease is prevalent.

Now, we want to investigate some qualitative behaviors of the simulations based on the parameters  $b$  and  $k$  in the phase diagrams. The simulations are done with  $T$  values 10, 20 and 30 days for both discrete and continuous cases.

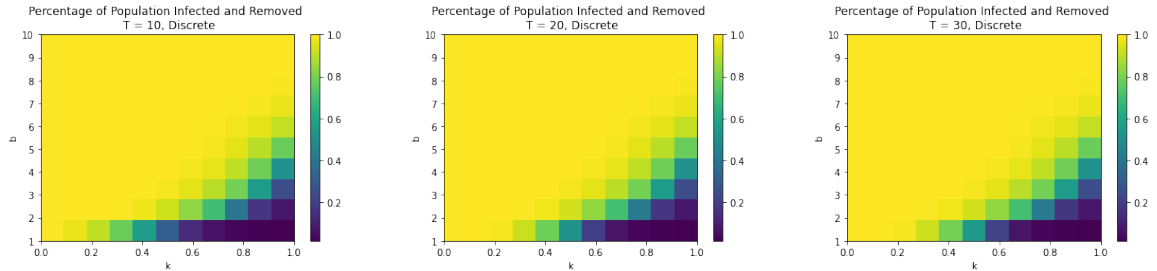


Figure 3: Phase diagrams for total percentage of population infected ( $i+r$ ), discrete

In figure 3, three phase diagrams illustrate the percentage of population infected and removed, i.e., the total percentage of population infected. It depends on both  $b$ , the number of interactions per day per individual that could spread the disease, and  $k$ , the recovery rate. The diagrams look

very similar for different lengths of simulation (10, 20, and 30 days). It is clear that as  $k$  increases and  $b$  decreases, the total percentage of population infected decreases. If each individual only has less than two interactions per day that could spread the disease and the recovery rate is higher than 0.6, then almost nobody will be infected over time. The yellow area represents that everyone is eventually infected. When  $k$  is less or equal to 0.2, then everyone is eventually infected. Also, when  $b$  is larger than 6, everyone is eventually infected. Furthermore, when  $b$  is relatively small (less than 7) but  $k$  is relatively large (larger than 0.3), everyone is eventually infected. More precisely, the cut edge between the area where all individuals are infected and the remaining area is approximately linear, so that we may estimate the edge by  $b = 10k - 3$ . Therefore, when  $b - 10k + 3 > 0$ , everyone is eventually infected.

We set  $b$  is larger than 1 in the above simulations. However,  $b$  can be less than 1. For instance,  $b = 0.5$  is equivalent to say, each individual has 50% chance to be infected by one interaction with infectious people per day.

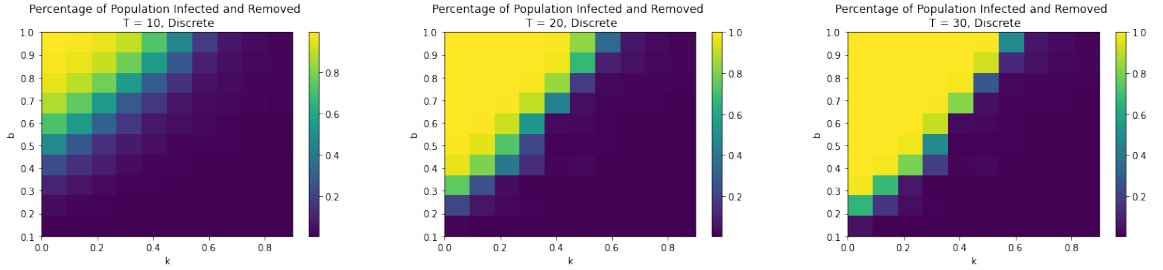


Figure 4: Phase diagrams for percentage of population infected and removed ( $i+r$ ), discrete

There are some interesting findings in figure 4. When  $T = 10$ , only when  $b$  is larger than 0.7 and  $k$  is smaller than 0.3, everyone is eventually infected. However, as  $T$  increases, the yellow area takes more space in the phase diagram, which means the range of parameters  $b$  and  $k$  is wider corresponding to where everyone is eventually infected. We do not observe effect of  $T$  in figure 3, but here it seems that for  $T = 30$  even if  $b$  is really small and  $k$  is moderate, there still exists chance that everyone is infected.

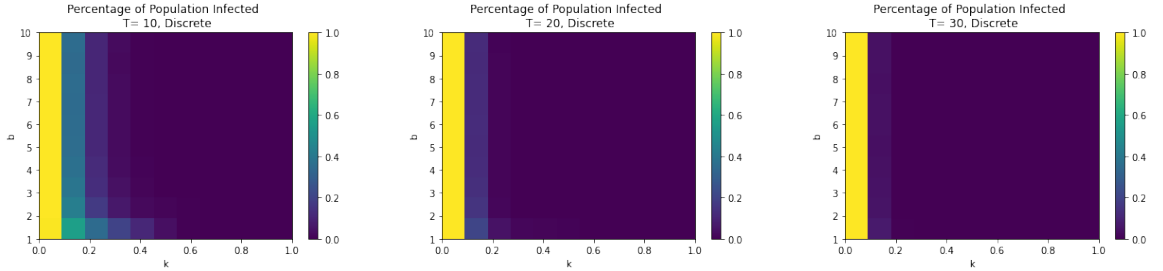


Figure 5: Phase diagrams for percentage of population infected ( $i$ ), discrete

Then, we try to discover the parameter regimes that  $i$  will quickly go to 0. Figure 5 shows the phase diagrams for only infected population. We conclude that the percentage of infected people will quickly go to 0, if the recovery rate  $k$  is large. Specifically, our simulations demonstrate when  $k$  is larger than 0.3,  $i$  will quickly go to 0 regardless of  $b$ .

Based on figure 6, when  $b$  is really small, the percentage of infected people can go to 0 even if the recovery rate  $k$  is small.

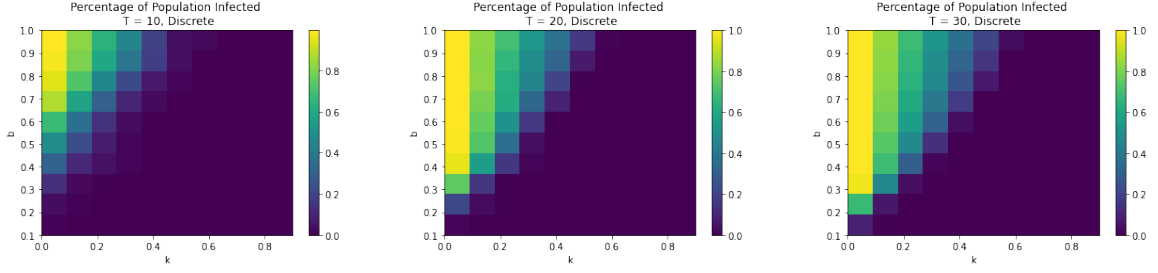


Figure 6: Phase diagrams for percentage of population infected (i), discrete

### 3.2 Continuous/ODE-Based Model

Similarly, we run simulations of the continuous ODE model, and make some plots to see how s, i, and r change with different values of k and b.

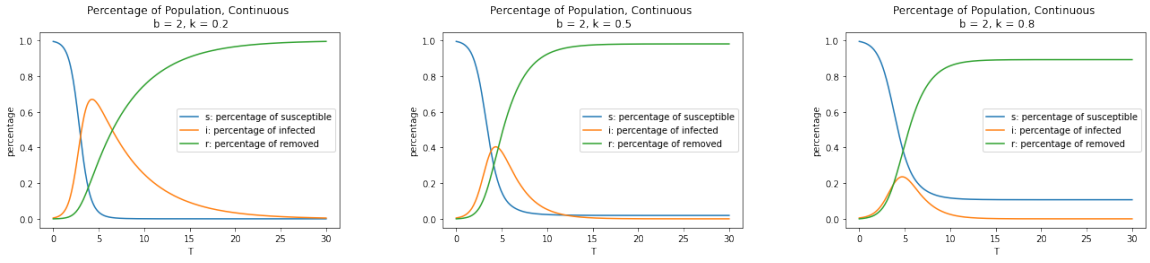


Figure 7: Percentage of population susceptible (s), infected (i), and removed (r), continuous

Figure 7 shows the percentage trends of ODE simulations. For  $b = 2$ , if the recovery rate  $k$  is 0.2, finally all individuals are removed. Within the first 10 days, the percentage of infected population increases then decreases, and the percentage of susceptible population decreases dramatically to 0, while the the percentage of removed population quickly increases to 1. When the recovery rate increases to a high value (0.8), the percentage of infected people finally goes to 0, and the percentage of removed people is over 0.9 while only very small proportion of individuals remain susceptible eventually.

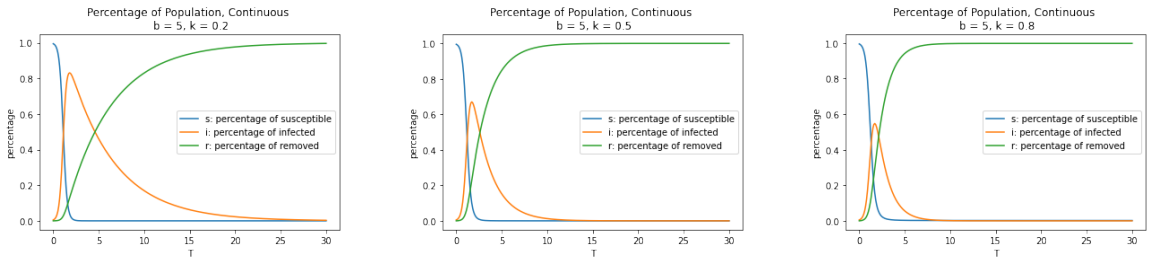


Figure 8: Percentage of population susceptible (s), infected (i), and removed (r), continuous

Figure 8 provides the similar results as figure 2 of discrete model. When  $b = 5$ , regardless of the recovery rate, all people are removed eventually. Then generate the phase diagrams for more information. The phase diagrams in figure 9 show the percentage of population infected and removed. When  $b$  is less than 2, it depends on  $k$ , and the total percentage of infected population decreases

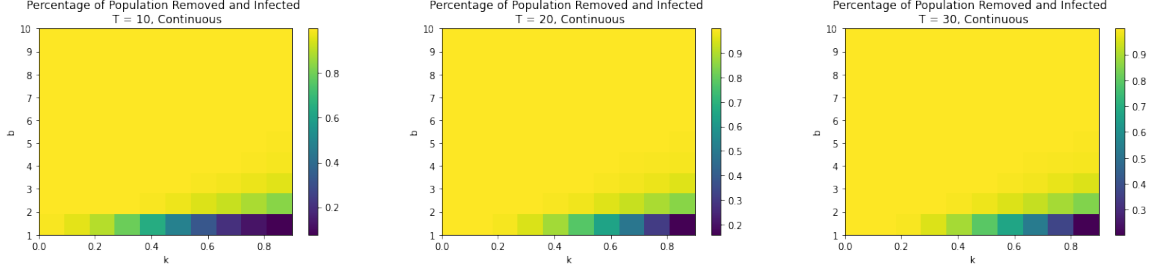


Figure 9: Phase diagrams for percentage of population infected and removed (i+r), continuous

as k increases. When b is larger than 2, then everyone is eventually infected. For cases with  $b < 1$ , figure 10 illustrates that similar patterns as figure 4 of the discrete model. When  $T = 30$ , if b is really small and k is relatively small, there still exists chance that all individuals are finally infected.

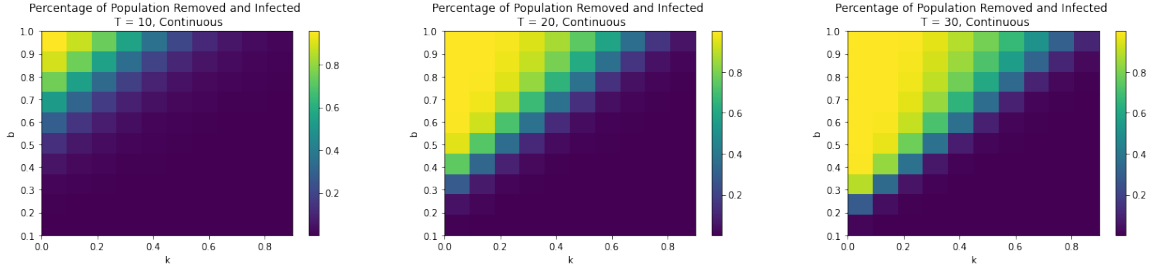


Figure 10: Phase diagrams for percentage of population infected and removed (i+r), continuous

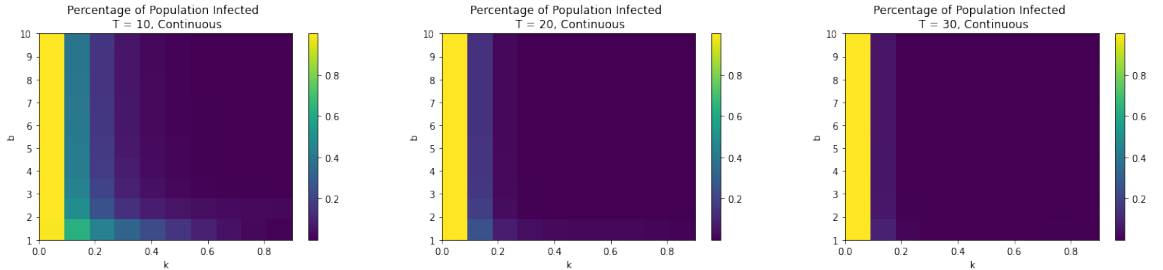


Figure 11: Phase diagrams for percentage of population infected (i), continuous

As for the parameter regimes that i will quickly go to 0, figure 11, the phase diagrams for only infected population, shows it occurs when the recovery rate k is larger than 0.3. In addition, figure 12 shows that when b is less than 1, the percentage of infected people is not 0 only if the recovery rate k is very small (less than 0.1).

Hence, the ODE continuous simulation and the discrete simulation totally agree in results.

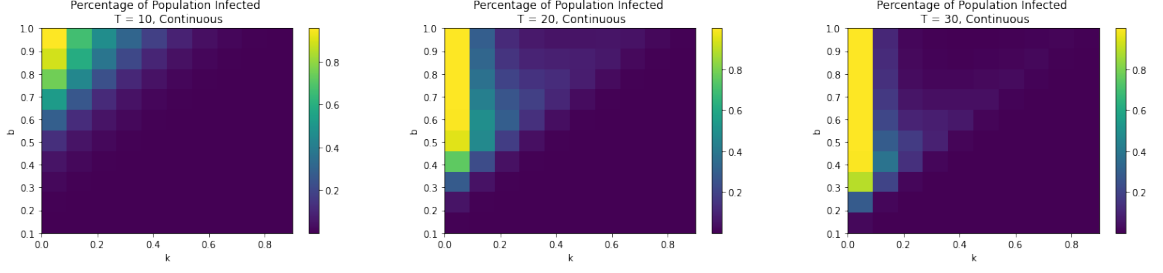


Figure 12: Phase diagrams for percentage of population infected (i), continuous

## 4 Variations and Improvements

### 4.1 Modelling the usage of mask

In this section, we study the effectiveness of mask in the spreading of the novel COVID-19 disease and develop a variation of the susceptible-infected-removed (SIR) model that provides a theoretical framework to investigate its spread within a community. More specifically, we will investigate how the use of mask will influence the percentage of deaths and hospitalizations with respect to different infectious contact rate. The model is based on the paper "To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic" by Steffen E. Eikenberry et al. Although it is now recommended and a common practice to wear mask to limit the spread of COVID 19, it is very controversial at the beginning of the pandemic that whether it is necessary and efficient to wear mask. So, we believe it is meaningful to understand quantitatively more about this topic. In this model, we will introduce seven variables.  $S(t), E(t), I(t), H(t), A(t), R(t)$  each denote susceptible, exposed, symptomatic infectious, hospitalized, asymptomatic infectious, and recovered classes, with an assumption that people progress from

$S \rightarrow E \rightarrow A \rightarrow I \rightarrow H$

At each state, infectious people will be able to recover, and only people in severe condition will go to hospital and may die.

$D(t)$  is also included to track cumulative deaths. Two sets of ODE equations will be used during simulation to find how use of mask will impact the spread of disease. We first consider a baseline model for the case where no masks are used. In the second set of odes, we will introduce more variables  $S_U(t), E_U(t), I_U(t), H_U(t), A_U(t), R_U(t), S_M(t), E_M(t), I_M(t), H_M(t), A_M(t), R_M(t)$  to divide the population in two part: The population wearing masks and the one not wearing masks Here are two sets of Odes:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta(t)(I + \eta A) \frac{S}{N} \\
\frac{dE}{dt} &= \beta(t)(I + \eta A) \frac{S}{N} - \sigma E \\
\frac{dI}{dt} &= \alpha \sigma E - \phi I - \gamma_I I \\
\frac{dA}{dt} &= (1 - \alpha) \sigma E - \gamma_A A \\
\frac{dH}{dt} &= \phi I - \sigma H - \gamma_H H \\
\frac{dR}{dt} &= \gamma_I I + \gamma_A A + \gamma_H H \\
\frac{dD}{dt} &= \sigma H
\end{aligned}
\tag{1}$$

We will run the simulation and observe what will happen in several days. Here,  $\beta$  is the infectious contact rate,  $\sigma$  is the transition exposed to infectious,  $\eta$  is the infectiousness factor for asymptomatic carrier,  $\alpha$  is the fraction of infections that become symptomatic,  $\phi$  is rate of hospitalization,  $\gamma_A$  is the recovery rate for asymptomatic,  $\gamma_I$  is the recovery rate, symptomatic,  $\gamma_H$  is the recovery rate for hospitalized.  $\sigma$  is the death rate in hospital.  $\epsilon_0$  is the outward efficiency of the masks while  $\epsilon_i$  is the inward efficiency of the masks.

We will find the default estimated values for all parameters with population masked and not masked. Here, an assumption is made that the inward efficiency and outward efficiency of masks are the same ( $\epsilon_0 = \epsilon_1$ ). We will draw a 2 dimensional phase diagram to show how the efficiency of mask (homemade mask or surgery mask) and the percentage of masked people will influence the hospitalization as well as the total deaths among the population.

## 4.2 Fitting SARS data with SIR model

In this section we will try to fit the SIR model with the SARS coronavirus. We try to use the SARS dataset in China from the link:[3]

Generally, We collect the dataset from the 2003-03-17 to 2003-06-01. It was mainly spreaded in China. To be specific, On April 22nd a hospital was established in Beijing and most of infected people are sent to the hospital. After that, the SARS coronavirus was controlled. In this model, we will try to simulate parameters and explore how building the hospital controlled the coronavirus. We use the ode simulation to simulate the most appropriate parameters  $b, k$  from data in 2003-03-17 to 2003-04-22. Then we add a variable  $\gamma_H$  inside the dataset. And we establish the following ode:

$$\frac{dS}{dt} = -\beta I * \frac{S}{N} * (1 - \gamma_H) \tag{3}$$

$$\frac{dI}{dt} = \beta * \frac{S}{N} * I * (1 - \gamma_H) - k * I \tag{4}$$

$$\frac{dR}{dt} = k * I \tag{5}$$

$$\frac{dD}{dt} = \gamma_H * I * \delta \tag{6}$$

$$\tag{7}$$



---

Using the first 30 days from 2003-03-17 to 2003-04-22, we may find appropriate  $\beta$  and  $k$  as well as the value  $\delta$ , representing the death rate. To do so, we will experiment with lots of  $\beta, k, \delta$  starting from common default values and try to find the most appropriate values such that MSE between the simulated data and real data are small ( $S-t, R-t, D-t$ ). Then, we will find appropriate  $\gamma_H$  so that the trend of data will be corresponded with the later data set (from 2003-04-22 to 2003-06-01). As what we know, China almost sent all the illness people to the hospital located in Beijing. And we try to see whether  $\gamma_H$  is close to 1. Finally, we will build a 1-d plot showing the relationship between the length of SARS coronavirus (end point is no infected people) with the percentage of people that we send to hospital. ( $\gamma_H$ )

### 4.3 Data visualization of Covid-19

In this part we will do the visualization to reflect the increasing trend and distribution in terms of Covid-19 in United States. The data set is from the data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) and we will mainly use the daily reports of US from April to present [1]. Firstly, we can make some time series plots to reflect the trend of several variables related to Covid-19 for all states, such as confirmed cases, deaths, recovered, test rate, etc. We consider to attain the total deaths as well as the confirmed cases every 7 days in order to observe potential peaks and try to analyze the reasons. Furthermore, we can use the python packages `mplcursor` and `matplotlib` to move the cursor along the lines in different days and relevant information (deaths and confirmed cases each day) will be presented.

Apart from that, we will implement several maps to better visualize the density of variables in each state. Specifically, we may collect the monthly data of confirmed cases, deaths, and recovered cases, and then produce choropleths using the `geoplotlib` package. There are longitude and latitude information from the same data set so that we can build the maps. We may use the packages `matplotlib`, `plotly` or `bokeh` to make the graph clear. Therefore, this visualization will reflect the distribution of Covid-19 through the United States.

## 5 Bibliography

The paper that we need to replicate for mask usage [2], the SARS dataset [3], and the covid-19 dataset [1].

## References

- [1] *CSSEGISandData*. 2020. URL: <https://github.com/CSSEGISandData?tab=overview&from=2020-07-01&to=2020-07-31>.
- [2] S. Eikenberry et al. *To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic*. 2020.
- [3] D. kp. *SARS 2003 Outbreak Dataset*. 2020. URL: <https://www.kaggle.com/imdevskp/sars-outbreak-2003-complete-dataset>.