

AMATH 383 Term Paper

Epidemic Modeling

Nan Tang, Rebecca Zhao

March 12, 2020

Contents

1	Abstract	2
2	Introduction	2
2.1	Compartmental Model	2
2.2	General Assumption & Notation	3
3	SIR Model without Natural Vital Dynamics	3
3.1	Formula	3
3.2	Analysis	4
4	SIR Model with Natural Vital Dynamics	5
4.1	Formula	5
4.2	Analysis	6
5	Case Study: COVID-19	6
5.1	Background	6
5.2	SEIR Model	7
5.3	Modified SEIR Model	7
5.4	Estimation of Parameters	8
6	Result & Discussion	10
6.1	Prediction Result	10
6.2	Discussion	12
7	Conclusion	12

1 Abstract

Outbreaks of epidemics caused by emerging infectious diseases (EIDs) have been evolving into a worldwide threat to public health and medical authorities. The outbreaks of SARS in 2003, influenza of H1N1 in 2009 and the ongoing concerns that related to respiratory illness, 2019 Novel Corona-virus, remind people of the urgent necessity to establish computational modeling and simulation system. Epidemiologists are indeed working on developing models that effectively predict short-term disease trends, assess expected impact of human interventions, and inform decision making. However, in consideration of the distinctive identities for various EIDs and limited availability to their characteristics during the outbreak, it is difficult to create a reliable and credible real-time modeling. Though with the help of computers, people can adjust parameters sooner to get a closer simulation of real virus development, uncertainties such as mutation present obstacles to achieving such purpose.

For the first half of this article, we will introduce two traditional epidemic modeling approach with both formula and result analysis. In the second half, we will bring in 2019 novel corona-virus as a specific case, develop a modified model and use the latest accessible data from January 15th to early February to explain in detail how the corresponding model works. At the end of the article, we will estimate the necessary model parameters and predict the trend of outbreak for the disease.

2 Introduction

2.1 Compartmental Model

There are three types of compartmental models for EIDs which are spread by direct person-to-person contact in population. Compartmental model divides the complex dynamic problem into separable clusters, groups of people who are susceptible, infected, and recovered. Compartmental models are also known as SIR model, presented by Kermack and Mckendrick in 1927. The first and simplest SIR model, lays on assumptions of constant population and confers complete immunity, where recovered people will get rid of the disease. We will also include a more complex SIR model, which takes into account long-term vital dynamics. We will also introduce the adjusted SEIR model which matches the prediction of the most recent novel corona-virus outbreak.

These traditional models are formulated as systems of ordinary differential equations with initial conditions and are analyzed mathematically. Although they are simple and mathe-

matically elementary, these models present concepts, intuition, and foundation for nowadays refined models.

2.2 General Assumption & Notation

Let N denotes total population, and assume the population is large enough so that number of people in each cluster could be regarded as continuous variables. Let $s(t)$ denotes the proportion of population who are susceptible at time t , $i(t)$ as proportion of infectious people, and $r(t)$ as proportion of recovery, so that $s(t) + i(t) + r(t) = 1$ and $s(t) \cdot N$ is population who are susceptible to the disease. We assume β , the infectious rate and γ , the recovery rate is constant. Here are some other general assumptions applied to all the compartmental models. For all SIR models, Kermack and Mckendrick assumed that any death during the epidemic is unrelated to infectious state. The model also assumes recoveries have immunity, in other words, people recovered from the disease will not return to susceptible (S) group.

3 SIR Model without Natural Vital Dynamics

3.1 Formula



We assume the epidemic disease occurs during a relative short period (less than a year) so the model does not include natural birth and death, i.e. population N is constant in this case. Base on the figure above, we have differential equations:

$$(Ns(t))' = -\beta s(t)i(t)N \quad (1)$$

$$(Ni(t))' = \beta s(t)i(t)N - \gamma i(t)N \quad (2)$$

$$(Nr(t))' = \gamma i(t)N \quad (3)$$

With initial conditions $s(0) = s_0$, $i(0) = i_0$, $r(0) = r_0$.

$$Ns(0) = Ns_0 > 0$$

$$Ni(0) = Ni_0 > 0$$

$$Nr(0) = Nr_0 \leq 0$$

Divide each equations (1) - (3) by constant population N , we can get differential equations for fractions of susceptible and infectious.

$$s(t)' = -\beta s(t)i(t)$$

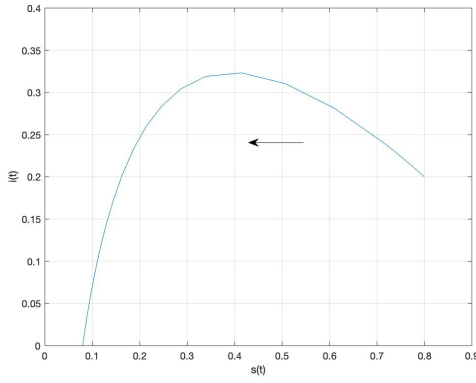
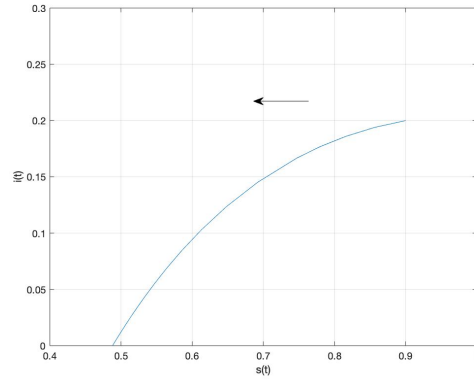
$$i(t)' = \beta s(t)i(t) - \gamma i(t)$$

Since $r(t) = 1 - s(t) - i(t)$, to solve this dynamic problem, it is sufficient to consider the phase plane of $s(t)$ and $i(t)$.

3.2 Analysis

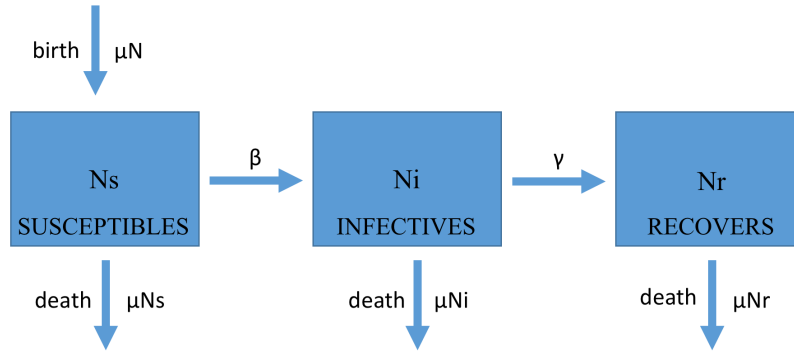
We define $\sigma = \beta/\gamma$ as basic reproductive number, also known as contact number, and $P_0 = s_0\sigma$ as effective reproductive number, which determines the number of secondary infections caused by a single infected person before his death or recovery. $s(0)$ is the initial value of susceptible fraction of total population. If $P_0 \leq 1$, the $I(t)$ decrease to zero as $t \rightarrow \infty$. If $P_0 > 1$, then $I(t)$ starts increasing, after it reaches maximum value, and then decrease to zero as $t \rightarrow \infty$.

Due to the closed population without vital dynamics and non-zero recovery rate, $I(\infty)$ will eventually go to zero. Figure 1 and Figure 2 prove this theorem since in both cases, $i(t)$ ends up begin 0 with growing time. In the first case when $P_0 < 1$, the fraction of recovery rate to infection rate is greater than the initial proportion of susceptible population, therefore the disease will not be an epidemic. In other words, the disease dies out before infecting all susceptible people. For the second case, when $P_0 > 1$, the fraction of recovery rate to infection rate is smaller than the initial proportion of susceptible people. As a result, the disease will outbreak and infect a large proportion of the susceptible population.

Figure 1: $\sigma = 0.4, s_0 = 0.8$ Figure 2: $\sigma = 1, s_0 = 0.8$

4 SIR Model with Natural Vital Dynamics

4.1 Formula



For some of the real cases, the disease will behave in a population for a long period of time, 10 to 20 years. Such disease is called endemic, in which births as a new source of susceptible and natural death will be included. For simplicity, we assume the birth rate is equal to the death rate, and let positive number μ denotes the constant. Adding these term to previous equations, we get a new set of equations for adjusted SIR model:

$$(Ns(t))' = -\beta s(t)i(t)N + \mu N - \mu Ns(t) \quad (4)$$

$$(Ni(t))' = \beta s(t)i(t)N - \gamma i(t)N - \mu Ni(t) \quad (5)$$

$$(Nr(t))' = \gamma i(t)N - \mu Nr(t) \quad (6)$$

Apply similar procedure and initial condition as above, we simplify (4) to (6).

$$\begin{aligned}s(t)' &= -\beta s(t)i(t) + \mu - \mu s(t) \\ i(t)' &= \beta s(t)i(t) - \gamma i(t) - \mu i(t) \\ s(0) &= s_0 > 0, i(0) = i_0 \leq 0\end{aligned}$$

4.2 Analysis

The change in differential equations will change our defined threshold condition for stability. Now we define $\sigma = \beta/(\gamma + \mu)$. If the contact number σ is less than one, so that an infected one replaces himself with less than one new infected case. The disease will die out. Inversely, if the infection rate is high enough to cross the threshold quantity, i.e. $\sigma > 1$, the disease will outbreak. Similar to the previous case, for both $\sigma > 1$ or $\sigma < 1$, the disease will die out as time goes on, since we are assuming the disease has zero lethal rates.

5 Case Study: COVID-19

5.1 Background

The corona-virus(COVID-19) is an ongoing outbreak of disease spreading rapidly and globally. As of the beginning of March, nearly a hundred thousand people are infected and more than 3,400 people died with corona-virus. 87 countries have been affected, with main outbreaks in China, Italy, South Korea, and Iran. The first case was identified in Wuhan, Hubei Province, China in December 2019. On January 20, Dr. Nanshan Zhong, an epidemiologist in china who also fought against SARS in 2003, confirmed that the novel virus can spread human by human. On January 23 morning, Wuhan was quarantined and a day after, all cities in Hubei province were quarantined. However, due to the large immigration for spring festival and school break, COVID-19 has been spreading out around the world. On 30 January, the World Health Organization declared a public health emergency of international concern for the outbreak of COVID-19. After a month of aggressive measures, China has slowed the corona-virus spreading. Based on the data collected and analysis from Dr. Nanshan Zhong's team, we can use mathematical modeling to predict the progression and take further responding properly.

5.2 SEIR Model

Since COVID-19 has a long incubation period which may be as short as two days to as long as 14 days, we introduce SEIR model for more accuracy. There are four groups of population in the SEIR model: susceptible(S), exposed(E), infectious(I) and recovered(R) so that the total population(N) equals the sum of them. Compared with the basic SIR model, SEIR adds a term to describe the latent population who are infected but not yet infectious and infectious(I) refers to those who are both symptomatic and infectious. Since the corona-virus shows no sign of long-term duration, we assume new birth and natural death will have no impact on each group of people.

Let β and γ still stands for the ratio of getting infected and recovery. The incubation rate σ denotes the rate by which asymptomatic patients ($E(t)$) develops into symptomatic ones ($I(t)$). Here is the base SEIR model without vital dynamic, note that different from previous formulas, we use functions of population in each group instead of proportion:

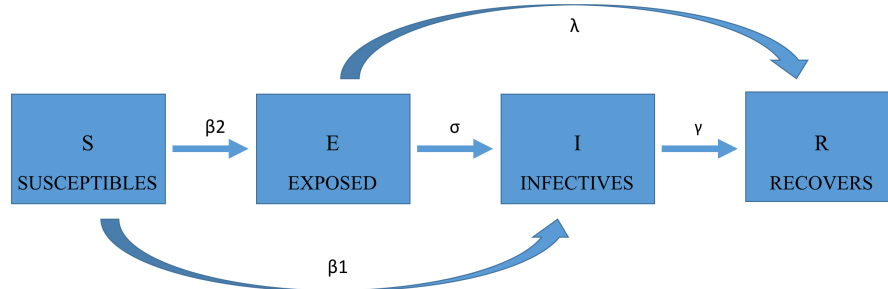
$$\frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N} \quad (7)$$

$$\frac{dE(t)}{dt} = \frac{\beta S(t)I(t)}{N} - \sigma E(t) \quad (8)$$

$$\frac{dI(t)}{dt} = \sigma E(t) - \gamma I(t) \quad (9)$$

$$\frac{dR(t)}{dt} = \gamma I(t) \quad (10)$$

5.3 Modified SEIR Model



In real studies, especially during the outbreak of the novel corona-virus, when different provinces in China are facing significantly distinctive scenarios, prediction models are built base on provincial data. Some modifications will be applied to the traditional SEIR model by

adding a few variables. Note that the ratio of transmission from susceptible(S) to infected(I) is not the same as transmission to exposed(E). Therefore, we divided β into β_1 and β_2 as the ratio of transmission to infectious and exposed. In this modified SEIR model, we primarily focus on daily changes within each group of people, in other words, equations start with $S(t+1) - S(t)$ instead of $\frac{d(S(t))}{dt}$. The modified model is shown below:

$$\begin{aligned} S(t+1) &= S(t) - \frac{k_1(t)bI(t)S(t)}{N(t)} - \frac{k_2(t)bE(t)S(t)}{N(t)} \\ E(t+1) &= E(t) + \frac{k_1(t)bI(t)S(t)}{N(t)} + \frac{k_2(t)bE(t)S(t)}{N(t)} - \sigma E(t) - \lambda E(t) \\ I(t+1) &= \sigma E(t) + I(t) - \gamma I(t) \\ R(t+1) &= \gamma I(t) + R(t) + \lambda E(t) \end{aligned}$$

Where $k_1(t), k_2(t)$ are time-dependent variables indicating number of contacts per person per day, Due to control policies, $k_2(t) = 15$ before January 22, and $k_2(t) = 3$ after that day. b is the probability of infection and λ is the rate that asymptomatic patients are self recovered. We will discuss details of these variables and parameters in next section.

5.4 Estimation of Parameters

Since we have obtained the fresh data of infected and recovery people in some provinces of China during January and February, we will try to estimate the exact number of parameter in these provinces base on the given conditions. In order to apply SEIR model on Hubei Province, the hardest-hit area by the disease, we are going to estimate infectious quantity β and recovery rate γ . β is controlled by number of infected people that a person exposed to and the probability of successful transmission when exposed to. Let k denotes number of infected people contacted each day and b denotes the probability, then $\beta = k \cdot b$. γ denotes average combined rate of recovery and death, and can be calculated via the equation $\gamma = \frac{1}{D} + \text{death rate}$, where D is the average or median duration of infection.

Base on relevant reports, the incubation period of coronavirus ranges from 2 to 14 days with medium 7 (Yang et al., 2020). We choose the medium 7 as the duration of infection, in order to avoid impact from outliers. The average mortality rate of Hubei province is 3%, therefore, the decreasing factor $\gamma = \frac{1}{7} + 0.03 = 0.173$. For sigma, which is the rate of transmission form exposed group(E) to infections(E), we choose the most common incubation period 5 days so that $\sigma = 1/5$. There are some typical cases showing that some people are asymptomatic and

Date	Cumulative Infections
Jan 16, 2020	45
Jan 17, 2020	62
Jan 18, 2020	121
Jan 19, 2020	198
Jan 20, 2020	270
Jan 21, 2020	375
Jan 22, 2020	444
Jan 23, 2020	549

can be self recovered due to their strong immune system, but during the incubation period they still carry the virus and the disease can be spread by human to human. We can assume that if someone is not symptomatic after the longest incubation period which is 14 days, then he will be considered as recovered so that the rate λ will be $1/14$ under this assumption.

A few conditions are need before estimating β 's. At $t = 0$, the beginning of outbreak, $I(t) = 1$, and $N(t) \approx S(t)$. The function that represents change in infectious group at early stage of outbreak can be written as:

$$\begin{aligned}\frac{dI(t)}{dt} &= \frac{I(t)S(t)}{N(t)} - \gamma I(t) \approx (\beta - \gamma)I(t) \\ I(t) &= e^{(k \cdot b - \gamma)t}\end{aligned}$$

Government intervention on migration control and enforced segregation of suspected populations have been applied to Hubei province since January 23, 2020. Due to such external force, the number of people contacted by both exposed and infectious decreased. For precision of estimation, we will fit the equation using data from January 16 to January 23.

After fitting the function of $I(t)$ with the distribution of data from the table, researchers have determined the expected value of probability of infection b be 0.0525, with 95% confidence interval 0.0507 to 0.0543 (Yang et al.). Here we assume that symptomatic infectious patients (I) are quarantined, therefore $k_1 = 3$, while asymptomatic people (E) have normal contact with people, therefore $k_2 = 15$ before January 23. After the publish of segregation policy, the majority of people, includes asymptomatic patients, who have contact records with confirmed cases are sent to the isolation facility. Such policy limited their contact with susceptible population, therefore, k_2 after January 23 can be considered as small as $k_1 = 3$.

Now we can estimate the infection quantity:

$$\begin{aligned}\beta_1 &= k_1 \cdot E(b) = 3 \cdot 0.0525 = 0.1575 \\ \beta_{2\text{pre}} &= k_2 \cdot E(b) = 15 \cdot 0.0525 = 0.7873 \\ \beta_{2\text{post}} &= k_1 \cdot E(b) = 3 \cdot 0.0525 = 0.1575\end{aligned}$$

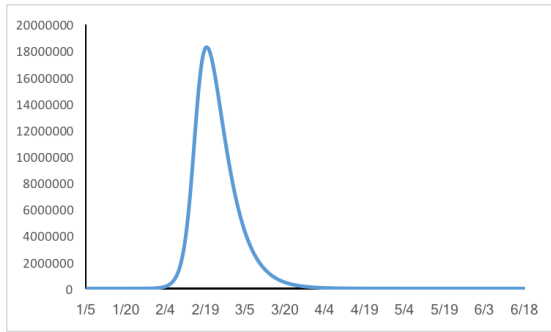
6 Result & Discussion

6.1 Prediction Result

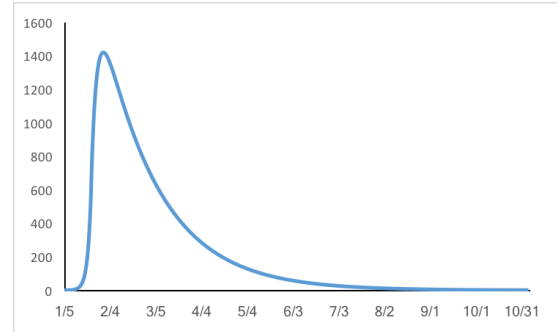
With the parameters estimated above, we can predict the epidemic progression of COVID-19 in Hubei Province with the modified SEIR model. To simplify the problem, we assumed that the resident population in Hubei Province is constant, which is approximately equal to 60 million. It is hard to tell the exact date of t_0 , and also we don't know how many people were carrying the virus when the first case was confirmed. Here the model was set up with an initial value E_0 is 1, which means that the first person was infected by some unclear sources but stayed asymptomatic when $t = 0$. S will be $N - 1$ with both I and R are zero. Then we can calculate $S(t + 1)$, $E(t + 1)$, $I(t + 1)$, $R(t + 1)$ respectively with equations showing above by iteration. Corresponding the date with infections number, we can predict when will the outbreak ends. With different conditions, it will result in different numbers of total infections and epidemic progression.

In the first case, the value of β is not changed. The model predicts infectious disease dynamics without any government intervention and public self prevention. People constantly contact with an average of 15 people every day until the virus dies out. Based on the infectious number in Hubei, when $I(t)$ is close to 45, t is set as January 16. Using excel and plotting data calculated from the SEIR model, we can see there will be a peak of infections that will reach 18 million on February 20. After that, the number of infectious people will start to decrease and finally, the virus will die out in mid-June. From the analysis of the basic SIR model, it shows that an epidemic disease will disappear anyway. In this situation, there will be nearly 59 million people infected, which indicates that 98% population in the province will suffer from this epidemic disaster. Obviously it is not realistic and the world will not let it happen.

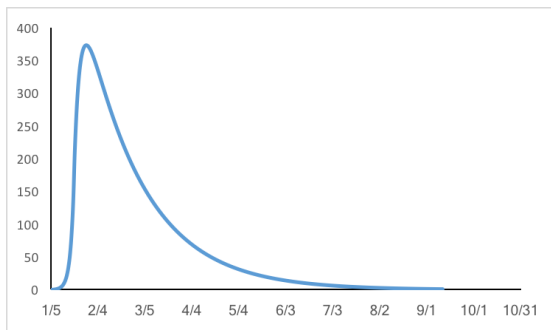
After the confirmation that the corona-virus can spread human by human, the public starts to concern and avoid contact with others. At the same time, the government implements



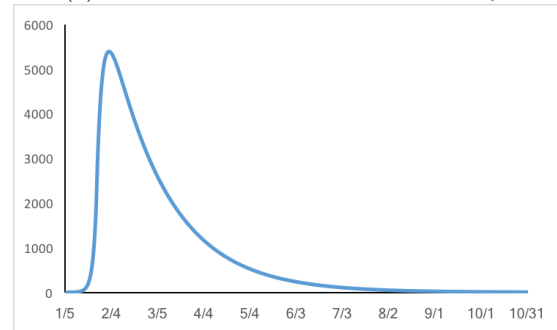
I(t) without any control



I(t) with intervention on January 23

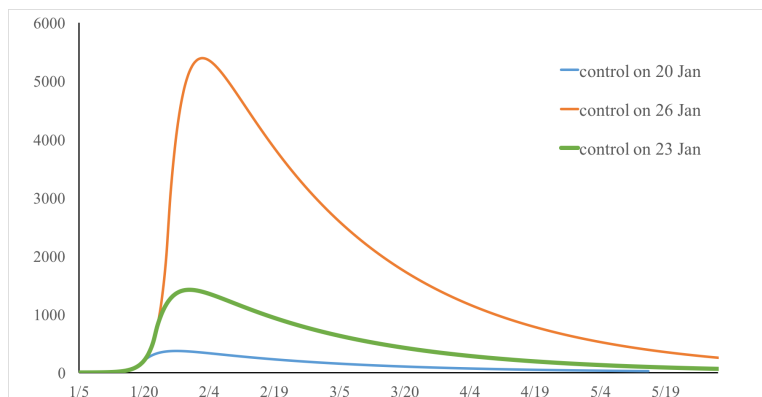


I(t) with intervention on January 20



I(t) with intervention on January 26

policies restricting the movement of people. We can assume that the contact rate after January 23 decreases to 3 people per day and people are quarantined at home. And then β_2 will equal to β_1 as the same as the transmission rate from susceptibles(S) to infectives(I). So after January 23, β_2 is set as $3 \cdot b$ and β_1 is $1 \cdot b$ in equals of $S(t+1)$ and $I(t+1)$. We considered that the outbreak ends when $I(t)$ is smaller than 1, also the cumulative infections are calculated by $N - S$ at this time. The epidemic peak is the point when $I(t+1) - I(t)$ is negative, which means that the number of infected people begins to decrease. Rebuilt up the model, the new plot shows that the peak will appear at the beginning of February with 1400 cases and the outbreak ends around November with a total size 14,000.



If the government implements 3 days earlier, the disease will completely end two months earlier and reach an epidemic peak with no more than 400 infections. The total size of infections is around 3,600. Instead, a 3-day delay in government control will lead to 50,000 cumulative cases in total. And then the corona-virus will die out at the end of this year. It is cleared shown that the epidemic disease dynamic is highly sensitive to the date when the transmission rate of β is changed if we keep other variables constant. The size of the infection will be effectively suppressed if the public start to avoid interaction and movement earlier.

6.2 Discussion

However, as of March 10, there are more than 6,700 people in Hubei province infected by COVID-19, much higher than the SEIR model predicts. If we plot the real data compared with the modeled data, real number of infection(I) is growing faster with a larger and steeper slope. There are a couple of reasons that lead to this gap. Our model is built on multiple ideal situations. To begin with, we applied the assumption that the population in Hubei province is constant since it is difficult to capture real-time migration patterns. If inflow is not equal to outflow population, the dynamic model will come up with a different solution. In fact, due to spring festival and school break, the migration index inflow is greater than outflow in Hubei Province in January, which means N is increasing and the real number of susceptibles is greater than what we got from the model. Also, r could be greater than 15 because people attend outside activities more often to celebrate the new year in December and January. Secondly, the first case was confirmed in mid-December, but the model shows the first asymptotic case that appeared on January 5. Insufficient medical equipment and laboratory tests, plus the novelty of corona-virus, are main reasons causing the delay of cases' confirmation and reporting. Furthermore, for all the model parameters, we are using the sample mean values as estimators. Such estimation heavily relies on the assumption that these model parameters such as the probability of contact infection b and recovery rate γ follow a normal distribution, otherwise, the foundation of the model will be violated. In addition to model limitations, we excluded the possibility of virus mutation. In fact, the virus' abilities of spread, legality, and resistance keep changing all the time.

7 Conclusion

In the above sections, we discussed in detail two traditional compartmental epidemic models that can be applied to predict outbreaks of most epidemic cases. For each compartmental

model, we also analyzed necessary conditions to determine if the disease will outbreak into EID. To better illustrate the magic of epidemic modeling, we introduced the case study of the 2019 novel corona-virus. During the study, we learned the procedure of model measure, model simplification, estimation of parameters and prediction of result. The key to building an accurate compartmental model is to identify separable compartments and specify inter-relationships between them. To optimize the accuracy of prediction, the following approaches may be effective. In traditional models, basic model parameters are infection rate and recovery rate, however statistically, adding more parameters and variables (along with equations) always increases the fitness to real situation. From the section that displays prediction results, we can perceive that initial conditions and assumptions are critical to the outcome. Due to the property of dynamic model, minor changes could lead to a significant difference in result. Therefore, checking initial values and validity of assumptions is important before solving the problem.

References and Appendix

1. Hethcote, Herbert W. “Three Basic Epidemiological Models.”
[Http://Www.mtholyoke.edu/ Ahoyerle/math333/ThreeBasicModels.pdf](http://www.mtholyoke.edu/ahoyerle/math333/ThreeBasicModels.pdf).
2. Weiss, Howard. “The SIR Model and the Foundations of Public Health.”
[Http://Mat.uab.cat/Matmat/PDFv2013/v2013n03.Pdf](http://Mat.uab.cat/Matmat/PDFv2013/v2013n03.Pdf).
3. Naji, et al. “The Dynamics of Epidemic Model with Two Types of Infectious Diseases and Vertical Transmission.” *Journal of Applied Mathematics*, Hindawi, 5 Jan. 2016, www.hindawi.com/journals/jam/2016/4907964/model-formulation.
4. Iannelli, Mimmo. “THE MATHEMATICAL MODELING OF EPIDEMICS.”
[Http://Itech.fgcu.edu/Faculty/Pfeng/Teaching/Epidemics.pdf](http://Itech.fgcu.edu/Faculty/Pfeng/Teaching/Epidemics.pdf).
5. Yang, Zifeng, et al. “Modified SEIR and AI Prediction of the Epidemics Trend of COVID-19 in China under Public Health Interventions.” *Journal of Thoracic Disease*, jtd.amegroups.com/article/view/36385/pdf.
6. Wang, Huwen, et al. “Phase-Adjusted Estimation of the Number of Coronavirus Disease 2019 Cases in Wuhan, China.” *Nature News*, Nature Publishing Group, 24 Feb. 2020, www.nature.com/articles/s41421-020-0148-0.
7. Coronavirus disease 2019 (COVID-19) Situation Report–25 2020. Available online: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200214-sitrep-25-covid-19.pdf?sfvrsn=61dda7d2>
8. “Confirmed COVID-19 Cases Globally.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 11 Mar. 2020, www.cdc.gov/coronavirus/2019-ncov/locations-confirmed-cases.html.
9. “Coronavirus Cases.” Worldometer, www.worldometers.info/coronavirus/.
10. Baidu qianxi (in Chinese) 2020 Available online: <https://qianxi.baidu.com/>
11. “Coronavirus Updates Live: U.S. Cases Top 1,000 as Spread Slows in China.” *NBC-News.com*, NBCUniversal News Group, 11 Mar. 2020, www.nbcnews.com/health/health-news/live-blog/coronavirus-updates-live-u-s-cases-top-1-000-spread-n1155241.