Piecewise SEIUR model for the spread of COVID-19

A Thesis

Presented to the Faculty of the Department of Mathematical Sciences

Middle Tennessee State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Mathematical Sciences

---

by

Ziren Chen

Spring 2021

Thesis Committee:

Dr. Abdul Khaliq, Chair

Dr. Zachariah Sinkala

Dr. Wandi Ding

APPROVAL

This is to certify that the Graduate Committee of

Ziren Chen

met on the

[7] day of [April, 2021].

The committee read and examined his/her thesis, supervised his/her defense of it in an oral examination, and decided to recommend that his/her study should be submitted to the Graduate Council, in partial fulfillment of the requirements for the degree of Master of Science in Mathematics.

Dr. Abdul Khaliq
Chair, Graduate Committee

Dr. Zachariah Sinkala

Dr. Wandi Ding

Dr. James Hart
Graduate Coordinator,
Department of Mathematical Sciences

Dr. David Chris Stephens
Chair,
Department of Mathematical Sciences

Signed on behalf of
the Graduate Council

Dr. David Butler
Dean,
School of Graduate Studies

DEDICATION

Dedicated to my father Jie, mother Jing and girlfriend Lin for their love and support.

ACKNOWLEDGMENTS

My deepest gratitude goes first and foremost to my research advisor Dr. Abdul Khaliq for his guidance and support over the last 3 years. I would like to thank Dr. James Hart, Director of the mathematical science program for his help. I gratefully acknowledge the help of Dr. Zachariah Sinkala and Dr. Wandi Ding, who sat on my committee. I also would like to thank Dr. Khaled Furati and Dr. Harold A. Lay, Jr. who spent time reviewing this research and providing their feedback.

Last my thanks would go to my beloved family for their love and support. I want to thank my father Jie and mother Jing for their support, concern, and encouragement all the time. I want to thank my girlfriend Lin for her company and love.

ABSTRACT

Coronavirus disease 2019 (COVID-19) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that can be transmitted through human interaction. In this thesis, we present a Piecewise Susceptible-Exposed-Infectious-Unreported-Removed model (SEIUR) for infectious diseases and discuss it qualitatively and quantitatively. The parameters are explored by mathematical and statistical methods. Numerical simulations of these models are performed on COVID-19 US data, and Python is used in the visualization of the simulation results. Outbreak factor is generated by piecewise SEIUR model to explore the future trend of the US pandemic. Several error metrics are given to discuss the accuracy of these models. The main achievement of this thesis is to propose the SEIUR model and piecewise SEIUR model and to find the relationship between the spread of the pandemic and control strategies by observing the results of the numerical simulations. Performance analysis of SEIUR model and piecewise SEIUR model is presented based on COVID-19 data.

CONTENTS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

When the Ebola epidemic broke out in Africa in 2015, Bill Gates issued a warning in a speech. He said: "If anything kills over 10 million people in the next few decades, it's most likely to be a highly infectious virus rather than a war." [22]. Unfortunately, he became a prophet. Since the end of 2019, the COVID-19 pandemic [21][27] caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is raging around the world [6]. In December 2019, the first cases of COVID-19 was detected in Wuhan, Hubei Province, China. As of March 11, 2021, more than 118 million confirmed cases had been reported in 219 countries and regions, of which more than 2.621 million died and 66.926 million were cured. At present, the number of confirmed cases is still rising rapidly [25].

In the last two decades, human beings have gone through lots of severe infectious diseases, such as the SARS and Ebola [1]. In fact, from the perspective of history, the natural enemy of human beings - infectious diseases - have never stopped invading us. Plague, cholera, smallpox, syphilis and tuberculosis have all caused great harm to humans in history. In the course of struggling with infectious diseases, human beings have very limited capabilities. It was not until the last one hundred years that scientists gradually gained some scientific insight of infectious diseases. For instance, Yersinia pestis was separated out in 1894 [9], and Alexander Fleming, a British scientist, invented penicillin, the first antibiotic in the world in 1928 [20][10]. Viral infectious diseases are more formidable than bacterial infectious diseases. Scientists still have no progress to eliminate viral infectious diseases other than smallpox. In 1796, the British doctor E. Jenner carried out the first successful vaccination test in human body, so as to really control smallpox [8]. Until 1979, the WHO (World Health Organization) officially declared that smallpox was wiped out.

When we are stagnant in the field of biological research, we need to find a new

approach. Studying mathematical models of infectious diseases can also help scientists understand the various indicators of them so as to control their spread. Mathematical models could describe the transmission rate, spatial range, and transmission path of infectious disease. These models can provide a guidance on how to implement effective prevention and controls. Based on the types of infectious diseases, common infectious disease models can be divided into SI, SIR, SIRS, SEIR models, etc.. In addition, based on the propagation mechanism, they also can be divided into different types based on ordinary differential equations, partial differential equations, and network dynamics.

Infectious disease models have a long history, generally believed to begin with D. Bernoulli's study of vaccination against smallpox in one of his papers [13] in 1760. The advancement of infectious disease models is in the beginning of twenty century. W. H. Hamer, R. A. Ross, and other scientists had published several papers [17][14] related to the establishment of infectious disease models. It was not until 1927 that Kermack and McKendrick obtained the first compartmental deterministic model (SIR) proposed in the study of the Black Death in London [18][2]. Five years later, in 1932, the SIS model was given [19][3]. Along with the appear of these infectious disease models, the threshold theory was also presented [4]. The SIR model is the most basic compartmental model, and many models that later appeared were based on it. The SIR model divides the total population into the following three populations: susceptible S, which represents the individuals who are not infected but are likely to be infected by this type of disease; infectious I, which represents the individuals who have been infected as patients and have the capability of transmitting the disease; and removed R, which stands for the individuals who have been removed from the infected. The establishment of the SIR model is based on the following three assumptions [16]: a. The population dynamics such as birth, natural death, and mobility of the population are not considered. The total population always maintains a constant, that is, $S(t) + I(t) + R(t) = N$. b. Contacts between the infected and susceptible

individuals are sufficient to spread the disease. c. Removed individuals will not be infected again. As time passes by, the SEIR model [5] draws more attention. Nowadays, the SEIR model is popularly used to analyze infectious diseases. In this model, people experience an incubation period in the process of transiting from susceptible compartment to infected compartment (we use Exposed (E) to represent people in this process). We assume that the individual in the incubation period (E) is infected but not yet infectious. For instance, SARS and Ebola diseases have an incubation period where the individual cannot yet transmit the disease to others.

As the COVID-19 continues to spread, it is crucial to find a reasonable mathematical model to simulate this spread. Previous work for simulating the spread of COVID-19 has been developed using many different methods, such as the deep learning neural networks, polynomial regression, SEIR model, and ARIMA [11]. Neural networks are not stable due to their always reaching a local optimum for data with irregular growth. Polynomial regression with third degree end up over-fitting the data, which leads to a very high bias when making predictions [12]. This is because the COVID-19 data has a different trend of spread during different periods. The most commonly used method to simulate COVID-19 data is the SEIR model. But Exposed individuals in the SEIR model do not have the capability of transmitting the disease, which is inconsistent with COVID-19. The parameters of SEIR model, such as transmission rate β and removed rate γ are not easy to derive. In this thesis, we propose the piecewise SEIUR model and use a type of least-squares method to do the parameter estimation, which overcomes the main difficulty in the SEIR model.

This thesis contains five chapters. Chapter 1 is the introduction. Chapter 2 introduces the SEIR model. In chapter 3, a modified SEIR model, called SEIUR model, is introduced and discussed. We use the least square method to do the parameter estimation for the model. In chapter 4, numerical simulations and error analysis of SEIUR model and piecewise SEIUR model are performed based on COVID-19 data and the Outbreak factor is calculated. In the last chapter, we make the conclusion of

whole research and do a discussion about our future work.

CHAPTER 2

SEIR MODEL FOR INFECTIOUS DISEASES

2.1 Introduction

The main consideration and research aspect of epidemiology is to study the spread of diseases. Epidemiology is mainly responsible for tracking and analyzing the factors that cause the disease to spread over time, and in this process, it tries to explore possible control methods. In addition to epidemiology, mathematical modeling is also an effective method for us to conduct research on infectious diseases. Mathematical modeling helps us analyze and predict infectious disease behavior. We have been studying mathematical epidemiological models for a long time. The first epidemic infectious disease compartment model was proposed by Kermarck and McKendrick. Epidemiological models usually divide the total population into several categories or compartments. In order to study the dynamic changes of infectious diseases, the SEIR epidemic model divides the population into four categories: susceptible S, exposed E, infectious I, and removed R individuals. In this basic model, the movement of individuals between compartments depends on their personal resistance, resilience, and their interaction with the infected person.

In this chapter, the properties of the SEIR models are explored. We represent the SEIR model by a diagram containing the four compartments, and we explore the relationship and conversion between the compartments through the interpretation and analysis of the four compartments and parameters.

2.2 Susceptible-Exposed-Infectious-Removed (SEIR) Model

Nowadays, the SEIR model is widely used to analyze infectious diseases. The SEIR model divides the total population into the following four populations: susceptible

S, which represents the people who are not infected but are likely to be infected by this type of disease; exposed E, which represents the people who have been infected but are not yet infectious; infectious I, which represents the people who have been infected as patients and has capability of transmitting the disease; and removed R, which stands for the people who have been removed from the infected.

The establishment of the SEIR model is based on the following assumptions [7]:

- The population dynamics such as birth, nature death, and mobility of the population are not considered.
- Contacts between the infected and susceptible individuals are sufficient to spread the disease.
- Removed individuals will not be infected again.
- Only infectious individuals are capable of transmitting the disease.

The SEIR model is as follows;

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

$$\frac{dE(t)}{dt} = \frac{\beta S(t)I(t)}{N} - \sigma E(t) \tag{2}$$

$$\frac{dI(t)}{dt} = \sigma E(t) - \gamma I(t) \tag{3}$$

$$\frac{dR(t)}{dt} = \gamma I(t) \tag{4}$$

The meaning of each parameter and variable is shown below:

N: Total population.

$S(t)$: The number of susceptible individuals at time t.

$E(t)$: The number of exposed individuals at time t.

$I(t)$: The number of infected individuals at time t.

$R(t)$: The number of individuals who have been removed from the infected at time t.

β: Transmission rate.

γ: Removed rate.

σ: Incubation rate.

The SEIR diagram below shows how individuals move through each compartment in the model:

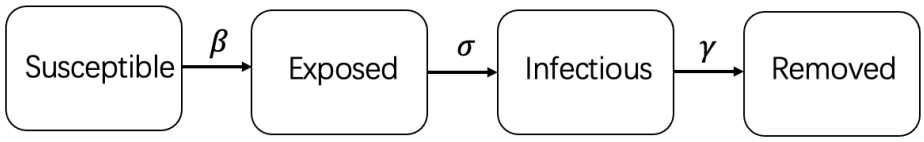

Figure 1: SEIR model

The SEIR model can be applied to most infectious diseases, but there are some limitations when used in the COVID-19 data. Due to the characteristics of COVID-19, exposed individuals also have the capability of transmitting the disease. The SEIR model is the neglect of infectivity of the people during the incubation period. The SEIR model also ignores the group of unreported cases. Due to the lack of medical resources and variability in the testing policy, there is no guarantee that all infected people are tested and reported under such a huge pandemic. This group of people is likely to become spreaders hidden in the population. The SEIR model with constant parameters cannot be used for long-term simulation. This is because the transmission rate β and removed rate γ must be changed in a long time.

2.3 Signals for Outbreak and end of the Epidemic

In an epidemic, we can obviously judge whether the epidemic is breaking out or disappearing through the change rate of the number of infected people. In the SEIR model, both E and I represent the number of infections. The difference is that those in E are in the incubation period while those in I have been diagnosed.

Thus, we can conclude that the epidemic is still in outbreak when the change rate of the summation of E and I is greater than 0, that is,

$$\frac{d(E+I)}{dt} > 0$$

which implies

$$\frac{dE}{dt} + \frac{dI}{dt} > 0$$

By the SEIR model, we have

$$\frac{\beta S(t)I(t)}{N} - \sigma E(t) + \sigma E(t) - \gamma I(t) > 0$$

It follows that

$$\frac{\beta S(t)I(t)}{N} - \gamma I(t) > 0$$

Because of $I(t) > 0$, we can imply

$$\frac{\beta S(t)}{N} - \gamma > 0$$

Consequently, We can simplify the above inequality to

$$\frac{\beta S(t)}{N} - \gamma > 0$$

Since $\beta > 0$ and $\gamma > 0$, it is equivalent to

$$\frac{\beta S(t)}{\gamma N} > 1$$

When the rate of change of $E + I < 0$, we have that the virus is disappearing.

By the similar steps,

$$\frac{d(E+I)}{dt} < 0$$

$$\frac{dE}{dt} + \frac{dI}{dt} < 0$$

$$\frac{\beta S(t)I(t)}{N} - \sigma E(t) + \sigma E(t) - \gamma I(t) < 0$$

$$\frac{\beta S(t)}{N} - \gamma < 0$$

$$\frac{\beta S(t)}{\gamma N} < 1$$

When the rate of change of $E + I = 0$, we have that the virus is under control and may coexist with humans for a long time.

We do not repeat the proof again; we can derive $\frac{\beta S(t)}{\gamma N} = 1$ by several steps.

In the next section, we use the results obtained in this section to derive the outbreak factor.

2.4 The Outbreak Factor generated by SEIR Model

In the last section, we implied that $\frac{\beta S(t)}{\gamma N}$ can determine the trend of epidemic. There are three situations:

- $\frac{\beta S(t)}{\gamma N} > 1$, which indicates that the epidemic is still in outbreak;
- $\frac{\beta S(t)}{\gamma N} = 1$, which indicates that the virus is under control and may coexist with humans for a long time;
- $\frac{\beta S(t)}{\gamma N} < 1$, which indicates that the virus is disappearing.

For the SEIR model, we can define $\frac{\beta S(t)}{\gamma N}$ as the outbreak factor at time t, denoted as $O(t)$, that is,

$$O(t) = \frac{\beta S(t)}{\gamma N}$$

Solving the $O(t)$ can help to generate the control strategies and explore the future trend of the epidemic.

2.5 Numerical Solution of SEIR Model using Python

There are many ways to solve the numerical solution of differential equations. Here we choose the forward Euler method and choose $\Delta t = 1$. Although the theoretical

model changes continuously, the real-life data are updated once a day. Therefore, using the forward Euler method with $\Delta t = 1$ can better fit the data we selected, and more information about the data is introduced in the following chapters.

Before using the forward Euler method, we can get the discrete model of SEIR with $\Delta t = 1$. It is shown as:

$$S[t+1] = S[t] - \frac{\beta S[t](I[t])}{N}$$

$$E[t+1] = E[t] + \frac{\beta S[t](I[t])}{N} - \sigma E[t]$$

$$I[t+1] = I[t] + \sigma E[t] - \gamma I[t]$$

$$R[t+1] = R[t] + \gamma(I[t])$$

Note that: $S[t]$ is different than $S(t)$ mentioned in the previous section. $S(t)$ is a continuous function, and t can be any non-negative real number. But here in $S[t]$, t can only could be an integer. $E[t]$, $I[t]$ and $R[t]$ are similar.

$S[t]$: The number of susceptible individuals on day t from the initial point.

β: Transmission rate.

N: Total population

$E[t]$: The number of exposed individuals on day t from the initial point.

σ: Incubation rate.

$I[t]$: The number of infected individuals on day t from the initial point.

γ: Removed rate.

$R[t]$: The number of removed individuals on day t from the initial point.

Now we can start to use the forward Euler method at $t = 0$. When we know the initial value of each variable at t=0, including $S[0]$, $E[0]$, $I[0]$ and $R[0]$ (in fact, this is simple, usually we set $S[0] = N - 2, E[0] = I[0] = 1, R[0] = 0$), and all parameters β, γ and σ (the determination of the parameters is introduced in the next section), we can obtain the value of each variable at $t = 1$ through one iteration, which is $S[1]$, $E[1]$, $I[1]$ and $R[1]$.

They are shown as:

$$S[1] = S[0] - \frac{\beta S[0](I[0])}{N}$$

$$E[1] = E[0] + \frac{\beta S[0](I[0])}{N} - \sigma E[0]$$

$$I[1] = I[0] + \sigma E[0] - \gamma I[0]$$

$$R[1] = R[0] + \gamma(I[0])$$

Similarly, we can generate the $S[2]$, $E[2]$, $I[2]$ and $R[2]$ by performing the second iteration:

$$S[2] = S[1] - \frac{\beta S[1](I[1])}{N}$$

$$E[2] = E[1] + \frac{\beta S[1](I[1])}{N} - \sigma E[1]$$

$$I[2] = I[1] + \sigma E[1] - \gamma I[1]$$

$$R[2] = R[1] + \gamma(I[1])$$

Repeating this step, we can generate the value of each variable at any time t. We use $S[t]$, $E[t]$, $I[t]$ and $R[t]$ to represent them respectively, and they satisfy the following equations.

$$S[t] = S[t-1] - \frac{\beta S[t-1](I[t-1])}{N}$$

$$E[t] = E[t-1] + \frac{\beta S[t-1](I[t-1])}{N} - \sigma E[t-1]$$

$$I[t] = I[t-1] + \sigma E[t-1] - \gamma I[t-1]$$

$$R[t] = R[t-1] + \gamma(I[t-1])$$

Where $S[t-1]$, $E[t-1]$, $I[t-1]$ and $R[t-1]$ are all from the result of the previous iteration.

In summary, the algorithm is shown as Algorithm 1.

Algorithm 1 Determining the numerical solution of basic SEIR model at all time $t \in \{1, 2, \ldots, n\}$

Input:
The initial value of variables: $S[0]$, $E[0]$, $I[0]$ and $R[0]$;
The optimal parameters: β, γ and σ;
Total population of some area: N;
The number of iterations or days: n;
Output
$S[t]$, $E[t]$, $I[t]$ and $R[t]$ at all time $t \in \{1, 2, \ldots, n\}$

Procedure
For i in 1 to n
$S[i] = S[i-1] - \frac{\beta S[i-1](I[i-1])}{N}$
$E[i] = E[i-1] + \frac{\beta S[i-1](I[i-1])}{N} - \sigma E[i-1]$
$I[i] = I[i-1] + \sigma E[i-1] - \gamma I[i-1]$
$R[i] = R[i-1] + \gamma(I[i-1])$
Return $S[i]$, $E[i]$, $I[i]$ and $R[i]$

2.6 Parameter Estimation for SEIR Model

In the SEIR model, there are three unknown parameters that need to be estimated: β, σ, and γ. The incubation rate of the COVID-19 has been estimated precisely, which is between 2 to 11 days(2.5^{th} to 97.5^{th} percentile) and that the mean incubation period was 6.4 days(95% CI: 5.6 - 7.7) [15]. Thus, we assume $\sigma = \frac{1}{6}$. The least square method is applied to find the optimal parameters of β and γ. The algorithm is shown as Algorithm 2.

Algorithm 2 Parameter Estimation for SEIR model

Input

Sequence $\{\beta_j\}_{j=0}^{99} = \{0,\ 0.01,\ 0.02, \ldots,\ 0.99\}$;
Sequence $\{\gamma_j\}_{j=0}^{99} = \{0,\ 0.01,\ 0.02, \ldots,\ 0.99\}$;
Parameters: $\sigma = \frac{1}{6}$;
Numerical Solution for I at day t, denoted as $I(t, \beta, \gamma)$;
True value for I at day t, denoted as $AI(t)$
The number of iterations or days: n;

Output

Optimal Parameters: β, γ

Procedure

For i in 0 to 99
    For j in 0 to 99
        $m_{ij} = \frac{1}{n}\Sigma_{t=0}^{n}(AI(t) - I(t, \beta_i, \gamma_j))^2$
Find $\min\{m_{ij},$ for i, j in 0 to $99\}$ and corresponding indices a and b
Return β_a and γ_b

CHAPTER 3

SEIUR MODEL

3.1 Introduction

As we mentioned in chapter 2, we know that the SEIR model has two limitations: one is the neglect of infectivity, and the other is that it does not consider the unreported cases.

We expect to complement these two weaknesses in the SEIUR model. To overcome the second weakness, the SEIUR model divides the infection equation (3) from SEIR model into two new parts, namely, unreported infection equation (14) and reported symptomatic infection equation (13). The new variables U and I represent people who have been infected but have not been tested for various reasons and the reported/confirmed infected people respectively. Let f be the proportion from E to I, and $1-f$ be the proportion from E to U. To handle the first weakness, the infectious population I in equation (11) and (12) is replaced by the sum of exposed population E and unreported infectious cases U, because the individuals in groups U and E have the ability to spread the virus to others during the infection. On the contrary, people in group I were required to be quarantined at home or in a hospital and thus could not infect others.

The establishment of the SEIUR model is based on the following assumptions:

- Keep first three assumptions same as SEIR model
- Exposed and Unreported infected individuals are capable of transmitting the disease
- Reported infectious individuals do not have capability of transmitting the disease

- Unreported infected individuals cannot become the reported infected individuals

The SEIUR model is given by;

$$\frac{dS(t)}{dt} = -\frac{\beta S(t)(E(t) + U(t))}{N} \tag{5}$$

$$\frac{dE(t)}{dt} = \frac{\beta S(t)(E(t) + U(t))}{N} - \sigma E(t) \tag{6}$$

$$\frac{dI(t)}{dt} = \sigma f E(t) - \gamma I(t) \tag{7}$$

$$\frac{dU(t)}{dt} = \sigma(1 - f)E(t) - \gamma U(t) \tag{8}$$

$$\frac{dR(t)}{dt} = \gamma(I(t) + U(t)) \tag{9}$$

Keep $S(t)$, $E(t)$, $R(t)$, β, γ, and σ same as SEIR Model

$I(t)$: The number of reported infected individuals at time t

$U(t)$: The number of unreported infected individuals at time t

f: The proportion of Exposed individuals that become reported infected individuals (Reported rate)

The transition diagram is shown as:

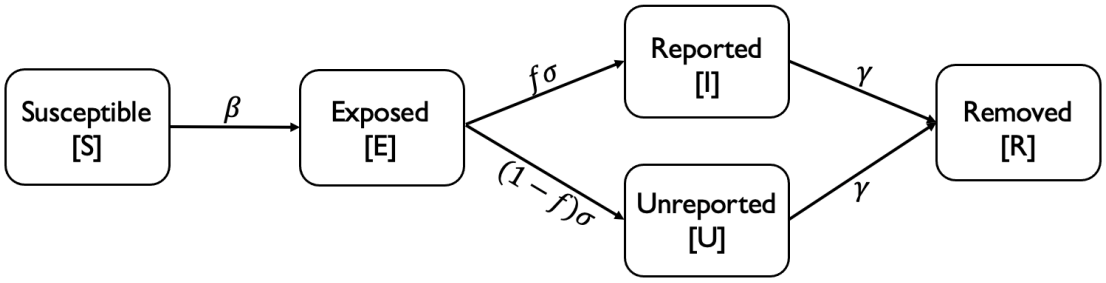

Figure 2: SEIUR model

Because real data based on SEIR is updated daily, we use the discrete model with $\Delta t = 1$. It is shown as:

$$S[t+1] = S[t] - \frac{\beta S[t](E[t] + U[t])}{N}$$

$$E[t+1] = E[t] + \frac{\beta S[t](E[t]+U[t])}{N} - \sigma E[t]$$

$$I[t+1] = I[t] + \sigma f E[t] - \gamma I[t]$$

$$U[t+1] = U[t] + \sigma(1-f)E[t] - \gamma U[t]$$

$$R[t+1] = R[t] + \gamma(I[t]+U[t])$$

3.2 Reported and Unreported Cases

At the early stage of the outbreak, test reagents for COVID-19 were not fully effective, and the number of reagents was not enough to cope with the huge test population. So it is certain that at the beginning of the pandemic, a significant number of people infected with COVID-19 could not be tested. In the middle stage of the pandemic, although various states had gradually opened up testing, due to the variability of COVID-19 test policy and the imperfect reagents, many people still could not be tested or received false negative results. And until the past six months, with the continuous improvement of the range and accuracy of the test, this has made the average test rate higher than before.

Now, our main goal is to estimate an appropriate value of the proportion f. Recalling the definition explained in the previous section, f stands for the proportion of E that become I. Let $E(t)$ represent the exposed individuals in time t, γ represent the removed rate, σ is the incubation rate, $n(t)$ represent the daily COVID-19 tests per thousand people at time t, $p(t)$ stand for the daily positive rate (the share of COVID-19 tests that are positive) at time t, and N is the total population. And we assume that all individuals in $E(t)$ are newly infected, then the proportion $f(t)$ at time t can be defined as (10).

$$f(t) = \frac{(\frac{1}{\sigma} + \frac{1}{\gamma})(\frac{n(t)p(t)N}{1000})}{E(t)} \tag{10}$$

Since the number of exposed people cannot be directly counted, it is almost impossible to find relevant statistical data on E. Here we use the SEIR model to generate historical data of $E(t)$. σ is $\frac{1}{6}$ assumed in section 2.6. γ is obtained by the least square method shown in Algorithm 2. N is the total population of United States. Regarding the data of $n(t)$ and $p(t)$, we use data from 'Our World in Data' [23], which provides the daily COVID-19 tests per thousand people and the daily positive rate. Now, we can generate the proportion $f(t)$ at any day t.

For simplicity, we define $f = \overline{f(t)}$ as the constant proportion rate for all data.

By the formula (10) and the data we have, we estimate $f = \overline{f(t)} = 0.5937$

3.3 Piecewise SEIUR Model based on Control Strategies

No matter how the SEIR model changes, the most core parameters are always β and γ. Therefore, the estimation of β and γ is particularly vital for both SEIR model and SEIUR model.

The basic SEIR model has a fixed value for β and γ for the entire time period. Since the values of β and γ are affected by many factors, such as government epidemic prevention measures, quarantine, and vaccination, it is unreasonable to set β and γ as constants. In addition, another popular method is to use a function to fit β and γ respectively. This approach makes the values of β and γ continuous and the changes are too frequent. Therefore, this thesis combines these two ideas, assuming that the value of β and γ are constant in a period of time, and when entering the next period, the values of β and γ are updated. The selection of update points is based on the time when government issued the relevant strategies or policies and the phase between two consecutive update points is called the "period" in this thesis. The strategies release timeline is shown in Figure 3.

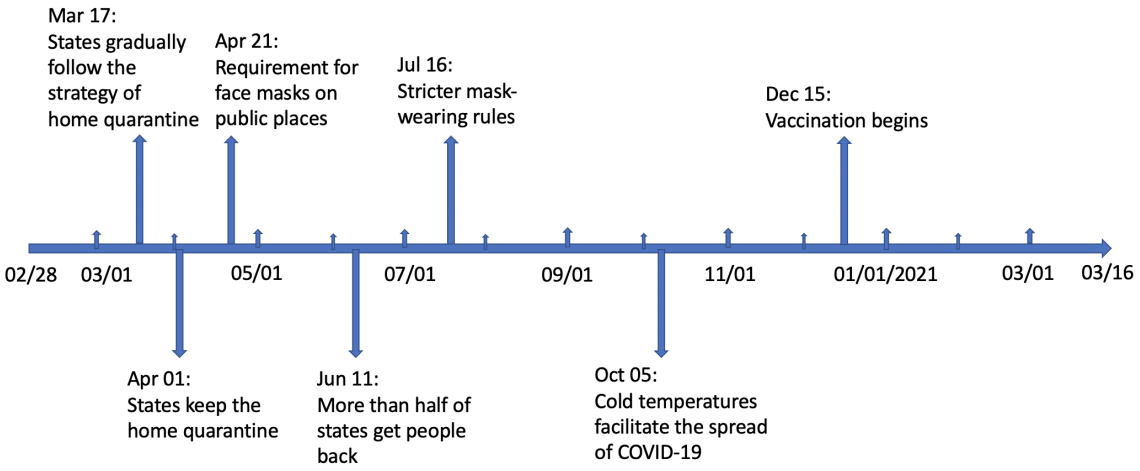

Figure 3: Strategies[26] release timeline

By the Figure 3, we separate the whole time period from 2/28/2020 to 3/16/2021 to 8 different parts, which are a. From 2/28/2020 to 3/16/2020 b. From 3/17/2020 to 3/31/2020 c. From 4/1/2020 to 4/20/2020 d. From 4/21/2020 to 6/10/2020 e. From 6/11/2020 to 7/15/2020 f. From 7/16/2020 to 10/4/2020 g. From 10/5/2020 to 1/15/2021 h. From 1/15/2021 to 3/16/2021. For each period, we have independent SEIUR model, and corresponding specific parameters β and γ. The piecewise SEIUR model is shown below:

$$\frac{dS(t)}{dt} = -\frac{\beta_i S(t)(E(t)+U(t))}{N} \tag{11}$$

$$\frac{dE(t)}{dt} = \frac{\beta_i S(t)(E(t)+U(t))}{N} - \sigma E(t) \tag{12}$$

$$\frac{dI(t)}{dt} = \sigma f E(t) - \gamma_i I(t) \tag{13}$$

$$\frac{dU(t)}{dt} = \sigma(1-f)E(t) - \gamma_i U(t) \tag{14}$$

$$\frac{dR(t)}{dt} = \gamma_i(I(t)+U(t)) \tag{15}$$

β_i: Transmission rate during the period i.

γ_i: Removed rate during the period i.

Other notations have the same meaning as before and the transition diagram for period i is shown in Figure 4.

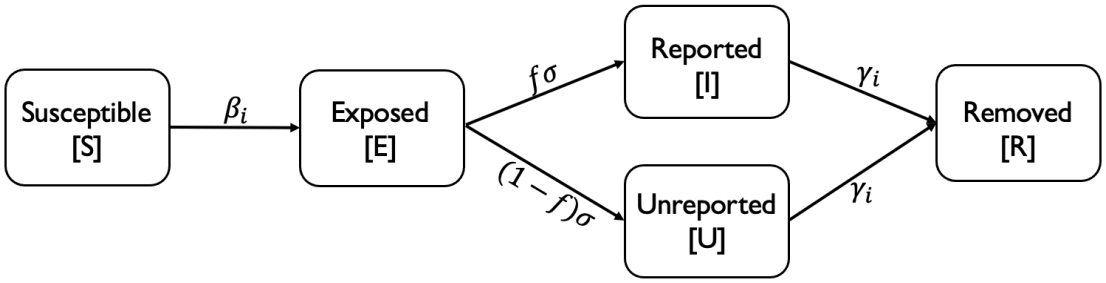

Figure 4: SEIUR model for period i

Because real data based on SEIUR is updated daily, for the epidemic of period i, we use the discrete model with $\Delta t = 1$, which is shown as:

$$S[t+1] = S[t] - \frac{\beta_i S[t](E[t] + U[t])}{N}$$

$$E[t+1] = E[t] + \frac{\beta_i S[t](E[t] + U[t])}{N} - \sigma E[t]$$

$$I[t+1] = I[t] + \sigma f E[t] - \gamma_i I[t]$$

$$U[t+1] = U[t] + \sigma(1-f)E[t] - \gamma_i U[t]$$

$$R[t+1] = R[t] + \gamma_i(I[t] + U[t])$$

3.4 The Outbreak Factor generated by Piecewise SEIUR Model

Recall Section 2.3 and 2.4 about the outbreak factor $O(t)$ for the SEIR model. We know that $O(t) = \frac{\beta S(t)}{\gamma N}$ which can be used to learn the future trend of the epidemic. If $O(t) > 1$, it indicates that the epidemic is still in outbreak. If $O(t) < 1$, then the virus is disappearing. If $O(t) = 1$, the virus is under control and may coexist with humans for a long time.

In this section, we discuss the outbreak factor $O(t)$ produced by the piecewise SEIUR model. We know that the value of outbreak factor $O(t)$ is related to the transmission rate β and the removed rate γ. Section 3.3 shows that different time periods have distinct parameters β and γ, thus, the outbreak factor $O(t)$ is also different for each period.

In piecewise SEIUR model, for each period i, when the rate of change of $E+I+U > 0$, we have that the epidemic is still in outbreak, that is,

$$\frac{d(E+I+U)}{dt} > 0$$

which implies

$$\frac{dE}{dt} + \frac{dI}{dt} + \frac{dU}{dt} > 0$$

By SEIUR model

$$\frac{\beta_i S(t)(E(t)+U(t))}{N} - \sigma E(t) + \sigma E(t) - \gamma_i(I(t)+U(t)) > 0$$

It follows that $\sigma E(t)$

$$\frac{\beta_i S(t)(E(t)+U(t))}{N} - \gamma_i(I(t)+U(t)) > 0$$

Since γ_i, $I(t)$, $U(t)$ are all positive, we have

$$\frac{\beta_i(E(t)+U(t))S(t)}{\gamma_i(I(t)+U(t))N} > 1$$

When the rate of change of $E+I+U < 0$, we have that the virus is disappearing.

By the similar steps, we have that

$$\frac{\beta_i(E(t)+U(t))S(t)}{\gamma_i(I(t)+U(t))N} < 1$$

When the rate of change of $E+I+U = 0$, we have that the virus is under control and may coexist with humans for a long time.

we have

$$\frac{\beta_i(E(t)+U(t))S(t)}{\gamma_i(I(t)+U(t))N} = 1$$

By mathematical analysis above, in each period i, the formula of outbreak factor at time t generated by piecewise SEIUR model is defined by

$$(O(t))_i = \frac{\beta_i(E(t)+U(t))S(t)}{\gamma_i(I(t)+U(t))N}$$

And we define $O_i = \overline{O(t)_i}$ to represent the average value of outbreak factor in period i.

3.5 Numerical Solution of SEIUR Model and Piecewise SEIUR Model

Section 2.4 introduces the numerical solution of the SEIR model. Similarly, we can also obtain the numerical solution of the SEIUR model and piecewise SEIUR model. The algorithms are shown as Algorithm 3 and 4.

Algorithm 3 Determining the numerical solution of SEIUR model at all time $t \in \{1, 2, \ldots, n\}$

Input:
The initial value of variables: $S[0]$, $E[0]$, $I[0]$, $U[0]$ and $R[0]$;
The optimal parameters: β, γ, f and σ;
Total population of some area: N;
The number of iterations or days: n;
Output
$S[t]$, $E[t]$, $I[t]$, $U[t]$ and $R[t]$ at all time $t \in \{1, 2, \ldots, n\}$

Procedure
For i in 1 to n
$S[i] = S[i-1] - \frac{\beta S[i-1](I[i-1])}{N}$
$E[i] = E[i-1] + \frac{\beta S[i-1](I[i-1])}{N} - \sigma E[i-1]$
$I[i] = I[i-1] + \sigma f E[i-1] - \gamma I[i-1]$
$U[i] = U[i-1] + \sigma(1-f)E[i-1] - \gamma U[i-1]$
$R[i] = R[i-1] + \gamma(I[i-1] + U[i-1])$
Return $S[i]$, $E[i]$, $I[i]$, $U[i]$ and $R[i]$

Algorithm 4 Determining the numerical solution of piecewise SEIUR model at time $t \in \{p_i, p_i + 1, \ldots, p_{i+1} - 1\}$ for $i = 1 \ldots n$
where p_i is the start point of period i and $p_1 = 0$

Input:
The initial value of variables: $S[0]$, $E[0]$, $I[0]$, $U[0]$ and $R[0]$;
The optimal parameters for each period i: β_i, γ_i, f and σ;
Total population of some area: N;
The number of iterations or days: n;
Output
$S[t]$, $E[t]$, $I[t]$ and $R[t]$ at all time $t \in \{1, 2, \ldots, n\}$

Procedure
For i in 1 to n
  For j in p_i to $p_{i+1} - 1$
    $S[j] = S[j-1] - \frac{\beta_i S[j-1](I[j-1])}{N}$
    $E[j] = E[j-1] + \frac{\beta_i S[j-1](I[j-1])}{N} - \sigma E[j-1]$
    $I[j] = I[j-1] + \sigma f E[j-1] - \gamma_j I[j-1]$
    $U[j] = U[j-1] + \sigma(1-f)E[j-1] - \gamma_j U[j-1]$
    $R[j] = R[j-1] + \gamma_j(I[j-1] + U[j-1])$
    Return $S[j]$, $E[j]$, $I[j]$, $U[j]$ and $R[j]$

3.6 Parameter Estimation for SEIUR Model and Piecewise SEIUR model

In the SEIUR model, there are four unknown parameters that need to be estimated: β, σ, f, and γ. We have shown that the proportion of E to I is 0.6, that is, $f = 0.6$, from section 3.2. We still assume $\sigma = \frac{1}{6}$ for the same reason in section 2.6. The least square method is employed to find the optimal parameters of β and γ. The algorithm is shown as Algorithm 5.

Algorithm 5 Parameter Estimation for SEIUR model

Input
Sequence $\{\beta_j\}_{j=0}^{99} = \{0,\ 0.01,\ 0.02, \ldots,\ 0.99\}$;
Sequence $\{\gamma_j\}_{j=0}^{99} = \{0,\ 0.01,\ 0.02, \ldots,\ 0.99\}$;
Parameters: $f = 0.6$, $\sigma = \frac{1}{6}$;
Numerical Solution for I at day t, denoted as $I(t, \beta, \gamma)$;
True value for I at day t, denoted as $AI(t)$
The number of iterations or days: n;
Output
Optimal Parameters: β, γ

Procedure
For i in 0 to 99
  For j in 0 to 99
    $m_{ij} = \frac{1}{n}\Sigma_{t=0}^{n}(AI(t) - I(t, \beta_i, \gamma_j))^2$
Find $\min\{m_{ij},$ for i, j in 0 to 99$\}$ and corresponding indices a and b
Return β_a and γ_b

In the piecewise SEIUR model, we also have $\sigma = \frac{1}{6}$ and $f = 0.6$. Parameter estimation of β_i and γ_i for each period i is shown as Algorithm 5.

Algorithm 6 Parameter Estimation for piecewise SEIUR model

Input
The number of periods: k;
Sequence $\{\beta_j\}_{j=0}^{99} = \{0,\ 0.01,\ 0.02, \ldots,\ 0.99\}$;
Sequence $\{\gamma_j\}_{j=0}^{99} = \{0,\ 0.01,\ 0.02, \ldots,\ 0.99\}$;
Parameters: $f = 0.6$, $\sigma = \frac{1}{6}$;
Numerical Solution for I at day t, denoted as $I(t, \beta, \gamma)$;
True value for I at day t, denoted as $AI(t)$;
The number of iterations or days: n;
Note that: $\{0, 1, 2, ..., n\} = U_{i=1}^{k}\{p_i, p_i + 1, \ldots, p_{i+1} - 1\}$;
where p_i is the start point of period i and $p_1 = 0$;
Output
Optimal Parameters: (β_i, γ_i) for all i in 0 to k

Procedure
For i in 1 to k
  For j in 0 to 99
    For g in 0 to 99
      $m_{ijg} = \frac{1}{p_{i+1} - pi}\Sigma_{t=p_i}^{p_{i+1}}(AI(t) - I(t, \beta_j, \gamma_g))^2$
  Find $\min\{m_{ijg},$ for j, g in 0 to 99$\}$ and corresponding indices a_i and b_i
  Return β_{a_i} and γ_{b_i}

CHAPTER 4

PERFORMANCE ANALYSIS OF SEIUR AND PIECEWISE SEIUR FOR COVID-19

4.1 Introduction

For a better understanding of the mechanics of our model, the implementation of SEIR model and piecewise SEIUR model using Python for United Stated COVID-19 data is presented in this chapter. Data comes from the reference website 'Worldmeter' [24]. We use a Python program to capture the active cases, the total cases and the total deaths as raw data and then perform data preprocessing on them. We generate the removed cases and total recoveries from the raw data. These five groups of data can be applied to estimate parameters in the SEIUR model Using Python program. Since piecewise SEIUR model is a policy-based piecewise model, we need to divide the whole set of data into several different parts, which represent various control strategies. We define each part as a period. After obtaining the piecewise data, the optimal parameters of piecewise SEIUR model can be obtained using the parameter estimation algorithm introduced in chapter 3 by a Python program. Finally, combining the optimal parameters and numerical solutions of the SEIUR model and piecewise SEIUR model generated in previous chapters, the numerical simulation of the active cases of each model at different time t is obtained. Error analysis is shown in the last section of this chapter.

4.2 Data

This section briefly introduces the source of the database used in this thesis, the content of the data, and how we preprocessed it. Finally, a graphical representation of data is provided.

4.2.1 Data source

The data used in this thesis is from the reference website 'Worldmeter' [24]. This website provides access to a variety of real-time statistical data. The website belongs to Dadax, an independent digital media company in the United States. This website records all pandemic data in the United States from February 15, 2020 to the present and data is updated daily.

4.2.2 Data content

The data consists of three different contents. The first is active cases starting from 2/15/2020 to 3/16/2021. The second and third contents are in same period as the first one, and they include the total cases and total deaths respectively. The detail of each content is shown in Table 1:

Table 1: Data content of total cases, active cases and total deaths

Data content	Data description	Data period	Data size
Total Cases	Total number of infections (including deaths and recoveries)	2/15/2020 to 3/16/2021	(1, 396)
Active Cases	Current number of infections (excluding deaths and recoveries)	2/15/2020 to 3/16/2021	(1, 396)
Total Deaths	Total number of deaths	2/15/2020 to 3/16/2021	(1, 396)

4.2.3 Data preprocessing

So far, we have obtained the total cases, the active cases and the total deaths from 2/15/2020 to 3/16/2021. In this section we discuss how we preprocess this raw data into what we need. First, we shift the initial data point from 2/15/2020 to 2/28/2020.

Although the United States announced the first case of coronavirus on 1/21/2020, the lack of effective detection methods and insufficient understanding of the coronavirus have led to the low accuracy of the data during the early stage of the pandemic. Thus, we discarded the data from the early period of the pandemic and set 2/28/2020 as the new initial data point.

Secondly, in addition to the above three sets of data, this thesis also needs the total removed cases and total recovered cases. And these two sets of data can be obtained by simple calculations between total cases, active cases, and total deaths. The formulas are as follows:

$$\text{Total Removed Cases} = \text{Total Cases} - \text{Active Cases}$$

$$\text{Total Recovered Cases} = \text{Total Removed Cases} - \text{Total Deaths}$$

Table 2 shows the data content after processing,

Table 2: Data content of total cases, active cases, total deaths removed cases and total recoveries

Data content	Data period	Data size
Total Cases	2/28/2020 to 3/16/2021	(1, 383)
Active Cases	2/28/2020 to 3/16/2021	(1, 383)
Total Deaths	2/28/2020 to 3/16/2021	(1, 383)
Removed Cases	2/28/2020 to 3/16/2021	(1, 383)
Total Recoveries	2/28/2020 to 3/16/2021	(1, 383)

4.2.4 Separate data in different period

In Section 3.3, we clearly explained the reasons for constructing the piecewise model and how we separate it. This section mainly shows the details after separation at the

data level.

Table 3 shows the details about each time period.

Table 3: Split data according to strategies [26]

Period #	Time period	Data size	Related policy
Period 1	2/28/2020 to 3/16/2020	(1, 18)	No control measures implemented
Period 2	3/17/2020 to 3/31/2020	(1, 15)	States gradually follow the strategy of home quarantine
Period 3	4/1/2020 to 4/20/2020	(1, 20)	States keep the home quarantine
Period 4	4/21/2020 to 6/10/2020	(1, 51)	Requirement for face masks on public places
Period 5	6/11/2020 to 7/15/2020	(1, 35)	More than half of states get people back to work
Period 6	7/16/2020 to 10/4/2020	(1, 81)	Stricter mask-wearing rules
Period 7	10/5/2020 to 12/14/2020	(1, 71)	Cold temperatures facilitate the spread of COVID-19
Period 8	12/15/2020 to 3/16/2021	(1, 92)	Vaccination begins

4.2.5 Training data and test data

According to the piecewise SEIUR model, in each period i, there is a corresponding SEIUR model. In addition to showing the accuracy of SEIUR model in each period i, this thesis also illustrates the performance of this model in the short future term. In other words, we obtain the SEIUR model in period i, and use it to estimate the data at the beginning of period $i+1$. For that reason, the training data for SEIUR model in period i is all data from itself, and takes the data of the first 7 days from period $i+1$ as the test data. For the last period, since the next period data is not available, we divide the period itself into training data and test data, and the number of test data is 7 days.

4.2.6 Visualize data using Python

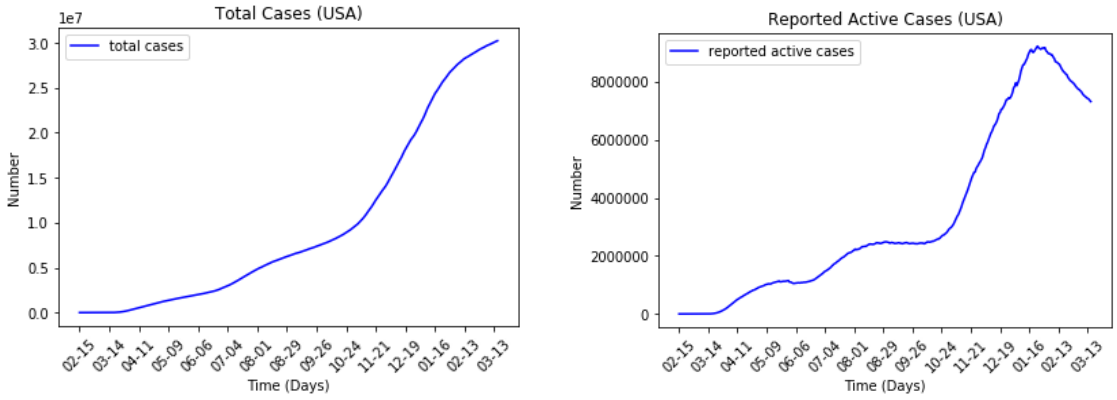

Figure 5: Total Cases and Reported Active Cases in US

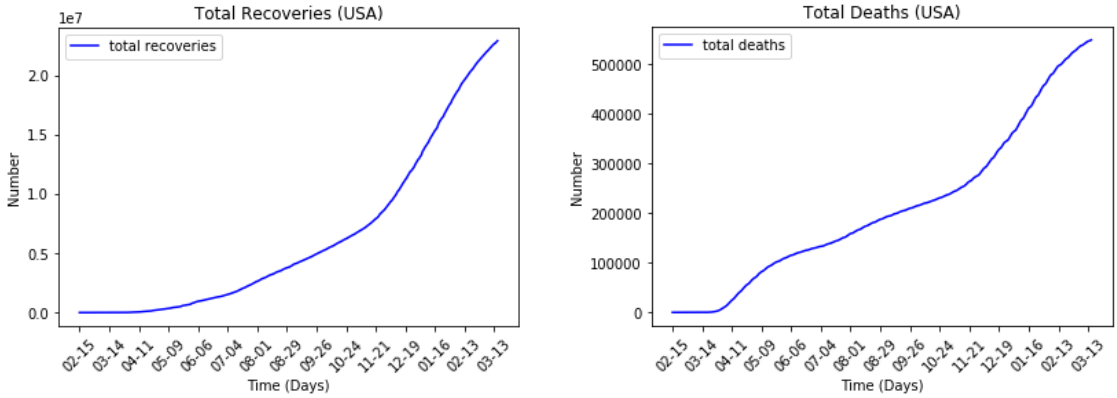

Figure 6: Total Recoveries and Total Deaths in US

Figure 5 shows the total cases and reported active cases. Combining the two figures, we can see that the total cases in the United States has exceeded 30 million since March. This means that nearly ten percent population in the United States has been infected with COVID-19, but the good news is that as the availability of the vaccine grows, the reported active cases has continued to decrease since February. We also found that starting from October 2020, due to policy changes, the total cases and the

reported active cases have both increased sharply. Figure 6 shows the total recoveries and the total deaths. We can see that the curves of the total deaths and the total recoveries have similar patterns.

4.3 Performance analysis of SEIUR and Piecewise SEIUR Models using Python

Before implementing the model, we need to estimate the parameters of SEIUR model and piecewise SEIUR model respectively.

4.3.1 Parameter Estimation for SEIUR model

Based on the parameter estimation algorithm 4 discussed in section 3.6, we estimate the parameters β and γ for SEIUR model in this section. Figure 7 is the 3D error graph with different β and γ.

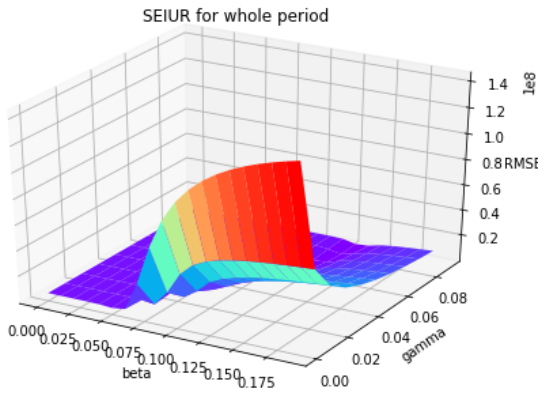

Figure 7: 3D error graph with different β and γ for SEIUR model

Optimal parameters: $\beta = 0.13$, $\gamma = 0.08$

4.3.2 Parameter Estimation for Piecewise SEIUR model

Based on the parameter estimation algorithm 5 discussed in section 3.6, we estimate the parameters β_i and γ_i for piecewise SEIUR model in each period i. Figures 8 to 11 are the 3D error graphs for each of period.

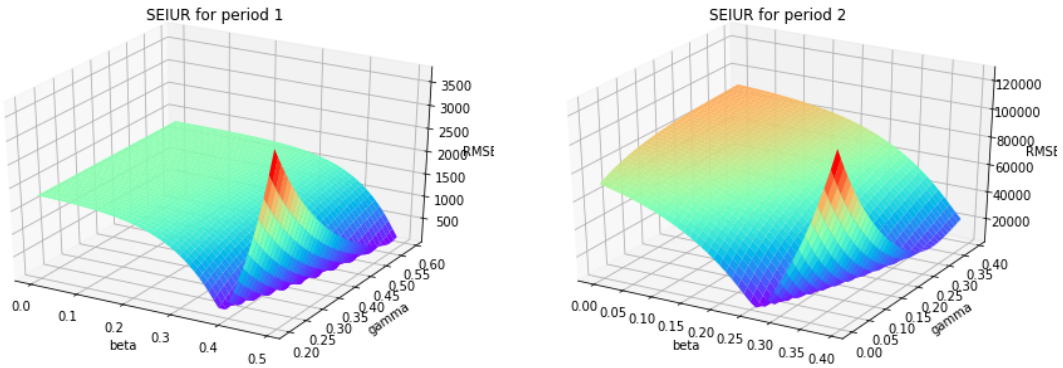

Figure 8: 3D error graph for different β and γ for period 1 and 2

Optimal parameters for period 1: $\beta = 0.44$, $\gamma = 0.36$

Optimal parameters for period 2: $\beta = 0.26$, $\gamma = 0.1$

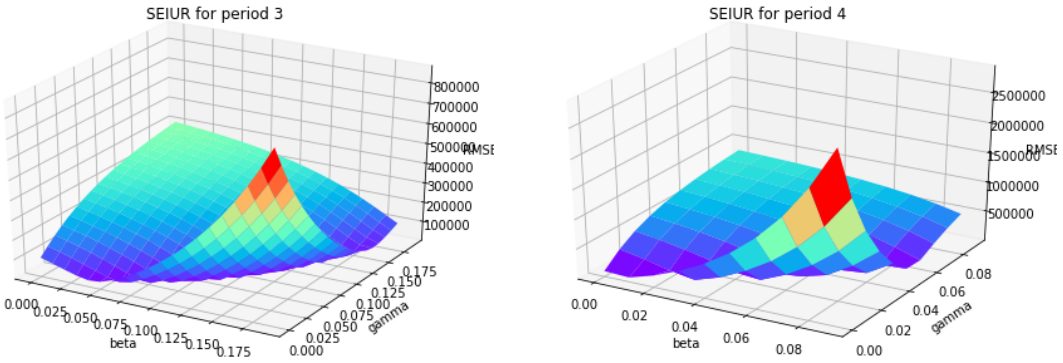

Figure 9: 3D error graph for different β and γ for period 3 and 4

Optimal parameters for period 3: $\beta = 0.1$, $\gamma = 0.04$

Optimal parameters for period 4: $\beta = 0.03$, $\gamma = 0.01$

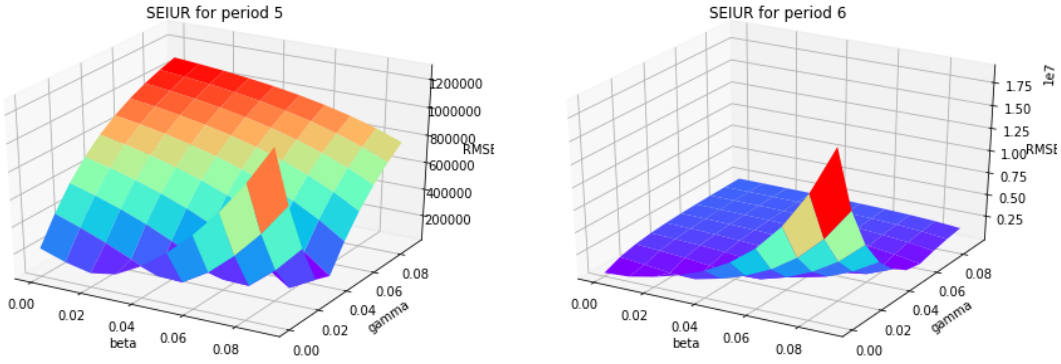

Figure 10: 3D error graph for different β and γ for period 5 and 6

Optimal parameters period 5: $\beta = 0.08$, $\gamma = 0.03$

Optimal parameters period 6: $\beta = 0.06$, $\gamma = 0.03$

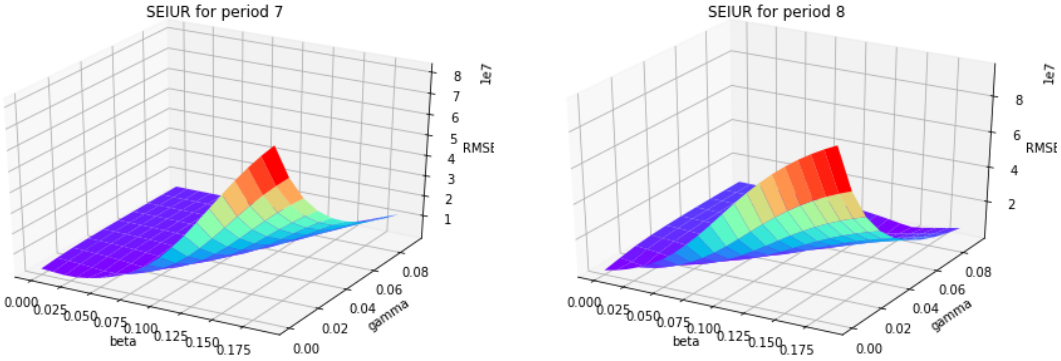

Figure 11: 3D error graph for different β and γ for period 7 and 8

Optimal parameters period 7: $\beta = 0.11$, $\gamma = 0.06$

Optimal parameters period 8: $\beta = 0.1$, $\gamma = 0.05$

4.3.3 Numerical simulation for SEIUR model

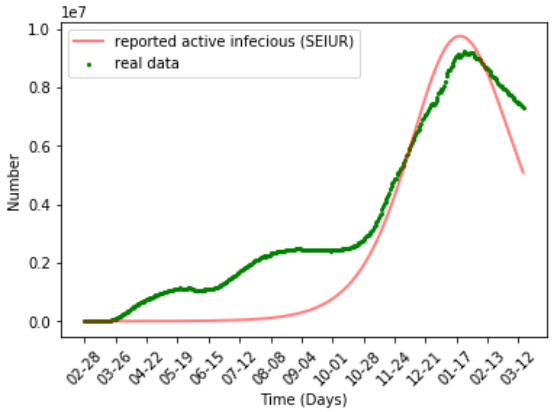

Figure 12: Numerical simulation for SEIUR

Figure 12 shows the numerical simulation results of SEIUR model on all data. We find that when simulating all of the data at once, the SEIUR model can successfully simulate the trend of data, but it cannot accurately reflect the true value of the data at each point. We have a more intuitive reflection in the error analysis in the next section.

4.3.4 Outbreak factor and Numerical simulation for piecewise SEIUR model

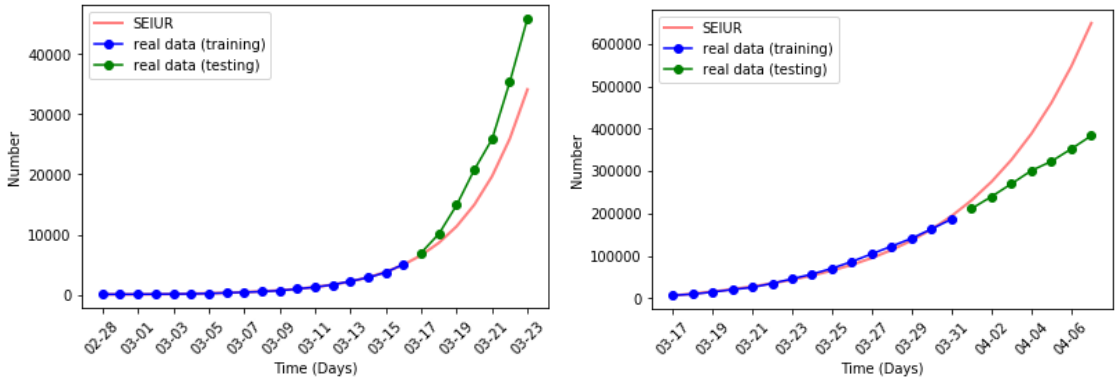

Figure 13: Numerical simulation in period 1 and 2

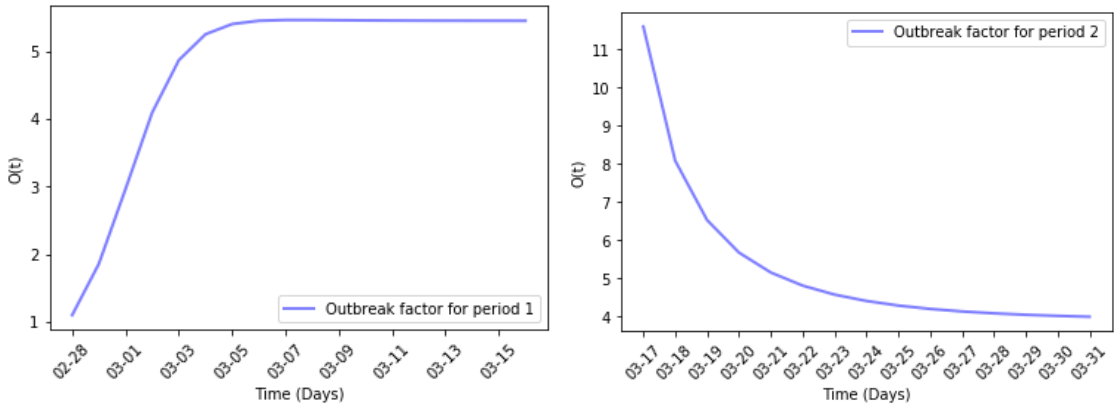

Figure 14: Outbreak factor in period 1 and 2

Figure 13 shows the estimated reported active cases and real data in period 1 and 2, and the blue dots are training data, green dots are test data. Figure 14 shows the daily outbreak factor in period 1 and 2. In period 1, no control measures were implemented which led to an exponential outbreak, and we calculate that the $O_1 = 4.9461$ in period 1, which is much greater than 1. This is consistent with what we proved in Section

2.3: When $O(t)$ is much greater than 1, the pandemic is in a major outbreak stage. Observing the simulation results on the test data, the SEIUR model in the first period can still simulate the data at the beginning of the second period very well. This also means that at the beginning of the second period, the home quarantine strategy has not yet contributed to a significant impact, and the pandemic was still breaking out. The $O(t)$ of the entire second period also confirmed this point. In this period, the $O_2 = 4.8743$, which is slightly lower than that of the period 1 but still far greater than 1. It indicates that in the middle and late stages of the period 2, the home quarantine strategy had achieved initial results, but the epidemic was still in outbreak. The test data of the period 2 showed that at the beginning of the period 3, we can hardly simulate the data by the model in period 2. The true value of the test data is significantly smaller than the result obtained by SEIUR model generated in period 2, which implies that the home quarantine strategy implemented in the period 2 achieved a significant effect in the period 3.

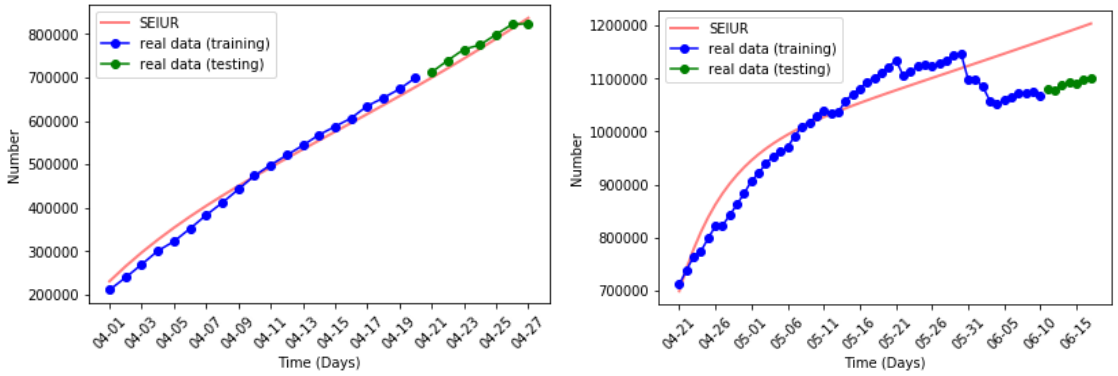

Figure 15: Numerical simulation in period 3 and 4

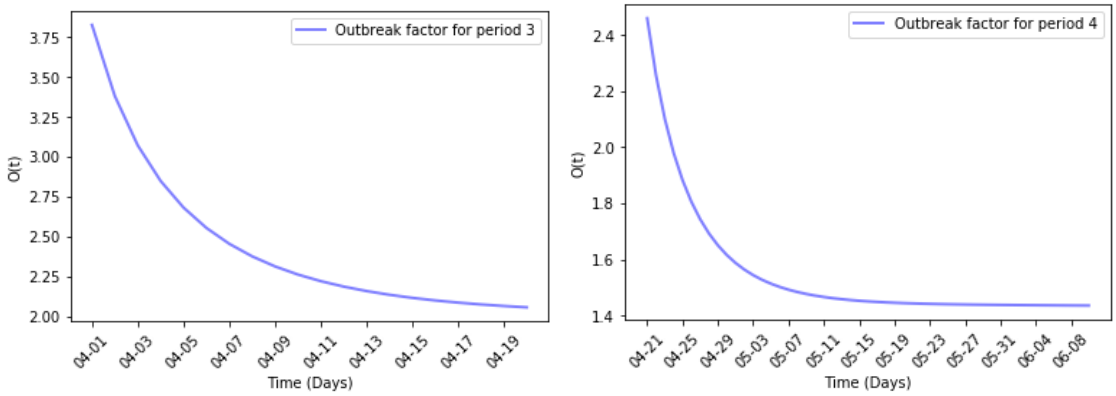

Figure 16: Outbreak factor in period 3 and 4

Figure 15 and 16 shows the results of numerical simulations and outbreak factor in period 3 and 4. $O_3 = 2.3412$ in period 3, which has a 52% decrease compared to period 2. This confirms the analysis made in the discussion of period 2. The home quarantine strategy has slowed the spread of the epidemic. In period 4, mask wearing further reduced the spread of pandemic. The figure about period 4 illustrates the growth rate of the active cases has decreased compared to period 3 and reached the first turning point at the end of May. Quantitatively, $O_4 = 1.5362$ in period 4, which is a decrease of 34% compared to period 3.

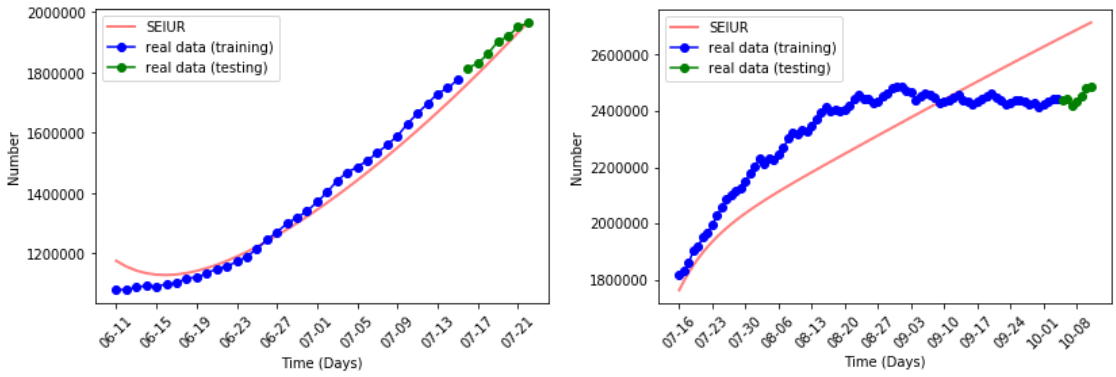

Figure 17: Numerical simulation in period 5 and 6

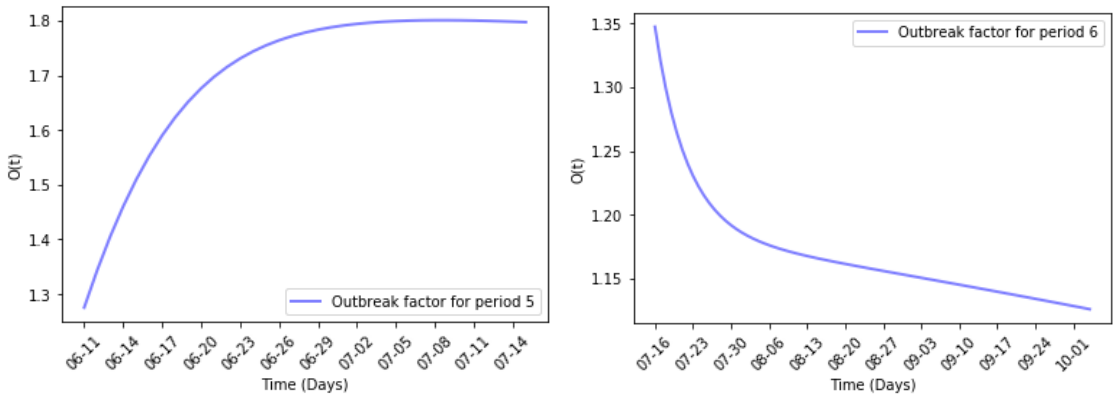

Figure 18: Outbreak factor in period 5 and 6

Figure 17 and 18 shows the results of numerical simulations and outbreak factor in period 5 and 6. In period 5, many states are relaxing their restrictions, with more than half set to be partially reopened, which led to the pandemic to rebound rapidly after reaching the first turning point. The number of active cases start to rise again. Average outbreak factor also rose from 1.5362 to 1.7176. In period 6, stricter mask-wearing rules slowed the spread of pandemic, the second turning point was reach at the end of August and $O_6 = 1.1660$.

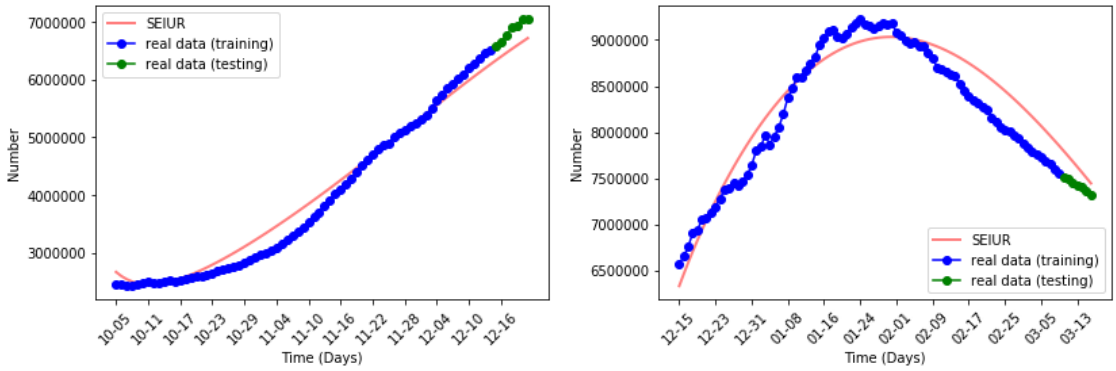

Figure 19: Numerical simulation in period 7 and 8

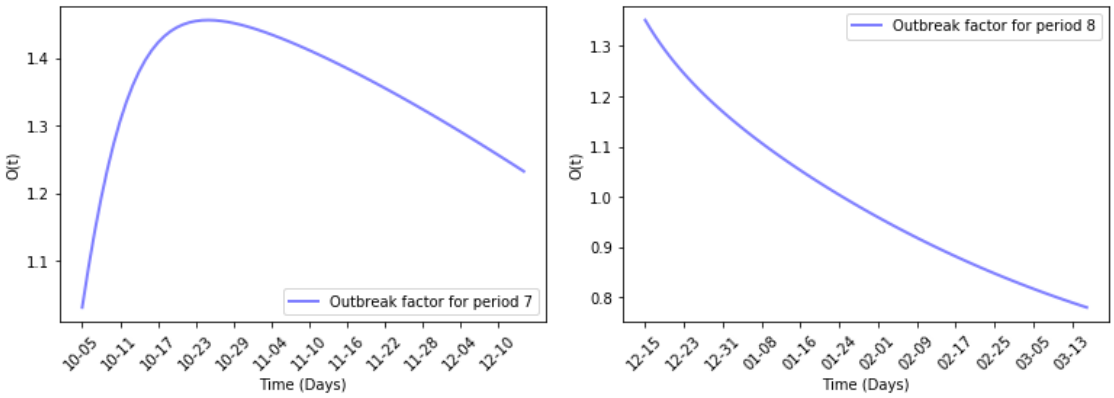

Figure 20: Outbreak factor in period 7 and 8

Figure 19 and 20 show the results of numerical simulations and outbreak factor in period 7 and 8. As winter brings shorter days and lower temperatures to the United Stated this facilitates the spread of COVID-19. Period 7 shows this process. With the advent of winter, active cases began to increase again after reaching the second turning point and O_7 rose back to 1.3414. Period 8 is the latest period so far, and vaccinations began on the first day of this period. The graph in period 8 shows that the vaccination quickly controlled the spread of the pandemic, and the third turning point appeared at the end of January 2021. So far, the reported active cases has continued to decline. And in period 8 $O_8 = 0.9991$, which is less than 1. If this continues, the epidemic is likely to be completely controlled in the next few months.

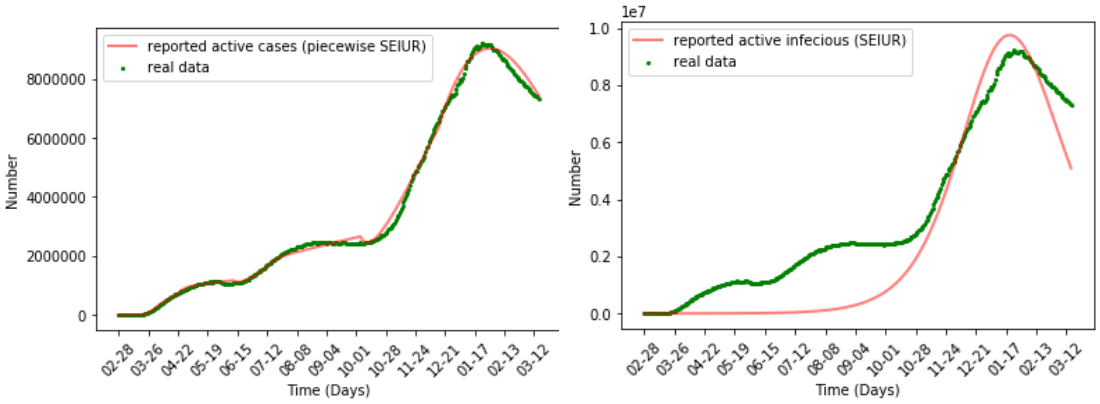

Figure 21: Numerical simulation for piecewise SEIUR and SEIUR

Figure 21 shows the numerical simulation results of piecewise SEIUR model and SEIUR model, where the result of piecewise SEIUR is obtained by combining the graphs of 8 periods. It can be seen from the figure that the piecewise SEIUR model performs better than the SEIUR model. This also means that it is reasonable to divide a long period of time into different sub-period according to policy dates, and then use the SEIUR model to simulate each period respectively.

4.3.5 Error analysis

After performing the numerical simulation of the model, error analysis is an effective way to evaluate the performance of the model. There are many error metrics, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE). In this section we discuss several representative error metrics for error analysis.

The Mean Absolute Percentage Error (MAPE) [28] is one of the most commonly used metrics to measure model accuracy. It is the average of the relative error and is

given by the following formula:

$$\text{MAPE} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{real_i - estimate_i}{real_i} \right| \times 100\%$$

The Mean Absolute Error (MAE) [28] is a popular error metric to measure model accuracy. As the name implies, it is the average of the absolute error. The formula is shown below:

$$\text{MAE} = \frac{1}{n} \cdot \sum_{i=1}^{n} |real_i - estimate_i|$$

The Mean Squared Error (MSE) [28] is an error metric to measure model accuracy. Since MSE is a continuous function, it is often used together with the least square method and gradient descent method. It is defined as the average squared error. The formula is shown below:

$$\text{MSE} = \frac{1}{n} \cdot \sum_{i=1}^{n} (real_i - estimate_i)^2$$

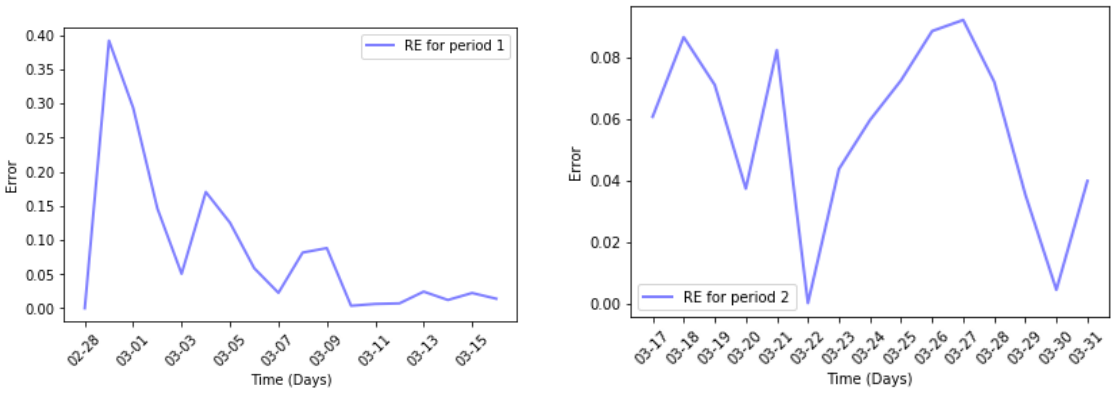

Figure 22: Relative error in period 1 and 2

Figure 22 shows the relative error of each data point in period 1 and period 2. By the formula of MAPE, we have that:

$$\text{MAPE}_{\text{period 1}} = 8.436\%$$

$$\text{MAPE}_{\text{period 2}} = 5.645\%$$

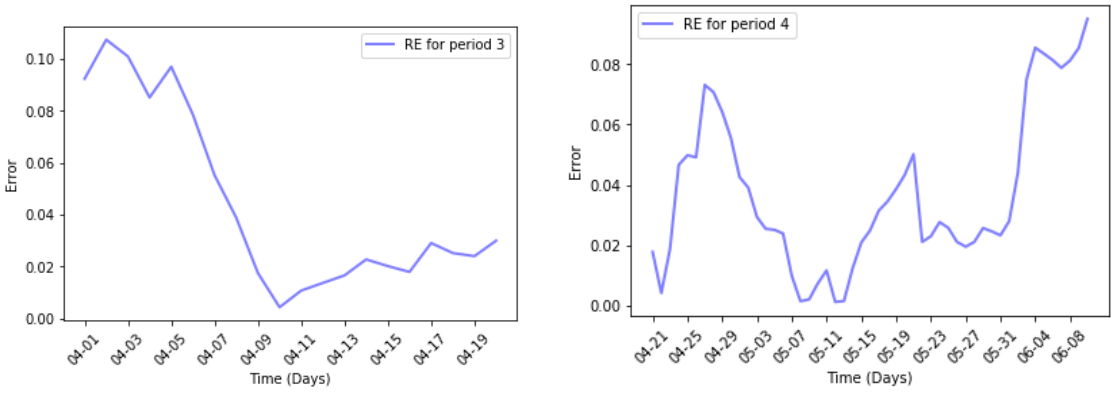

Figure 23: Relative error in period 3 and 4

$$\text{MAPE}_{\text{period 3}} = 4.443\%$$

$$\text{MAPE}_{\text{period 4}} = 4.721\%$$

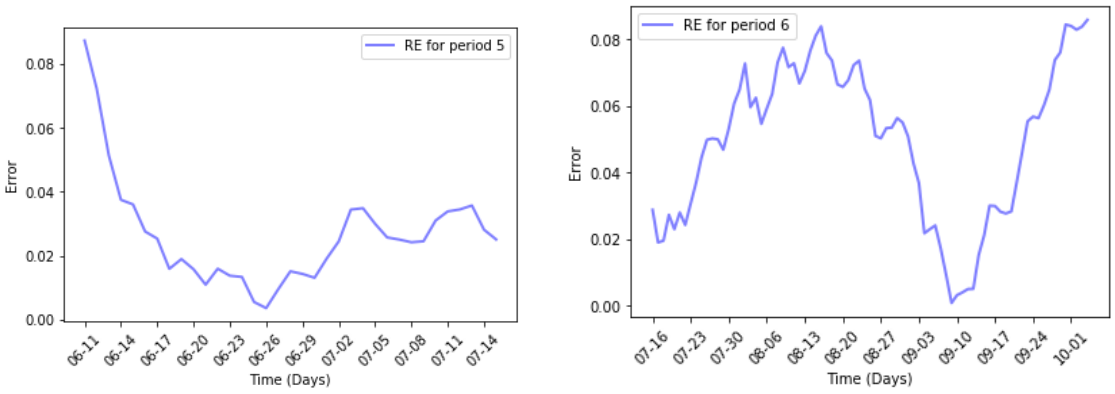

Figure 24: Relative error in period 5 and 6

$$\text{MAPE}_{\text{period 5}} = 2.665\%$$

$$\text{MAPE}_{\text{period 6}} = 4.961\%$$

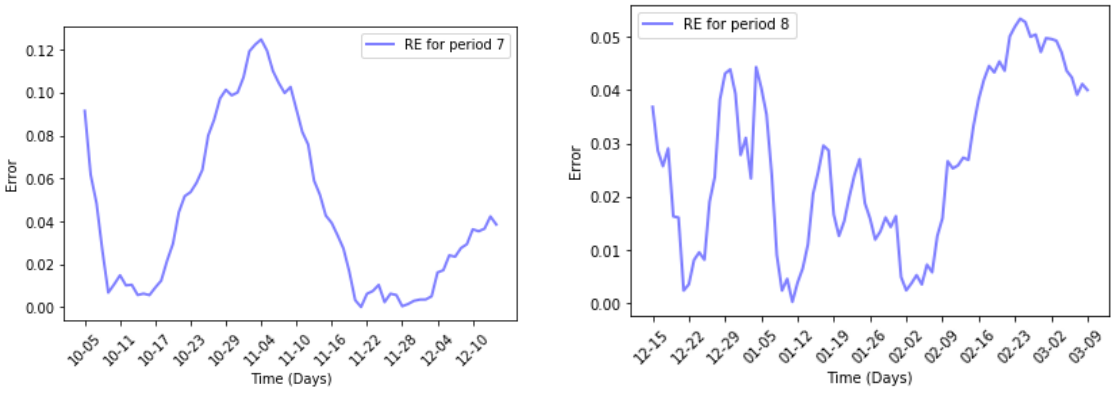

Figure 25: Relative error in period 7 and 8

$$\text{MAPE}_{\text{period 7}} = 4.404\%$$

$$\text{MAPE}_{\text{period 8}} = 2.615\%$$

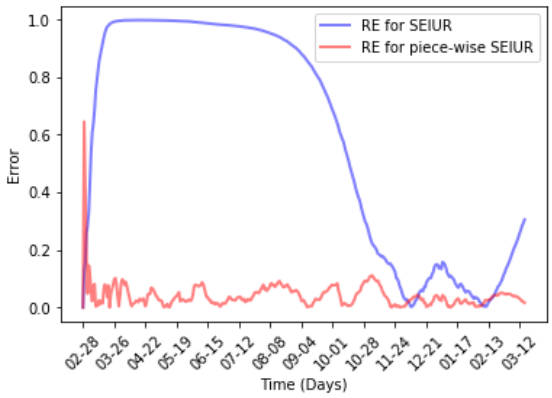

Figure 26: Relative error for Piecewise SEIUR and SEIUR

$$\text{MAPE}_{\text{PW-SEIUR}} = 4.083\%$$

$$\text{MAPE}_{\text{SEIUR}} = 59.526\%$$

From the results shown in Figure 26, it can be seen that the piecewise SEIUR model has small MAPE in each period and entire period. Specifically, the MAPE from period 3 to period 8 is lower than 5%, the MAPE in all periods is lower than 10%, and the MAPE in the entire period is only 4%. In contrast, the MAPE of SIEIUR model exceeds 50%.

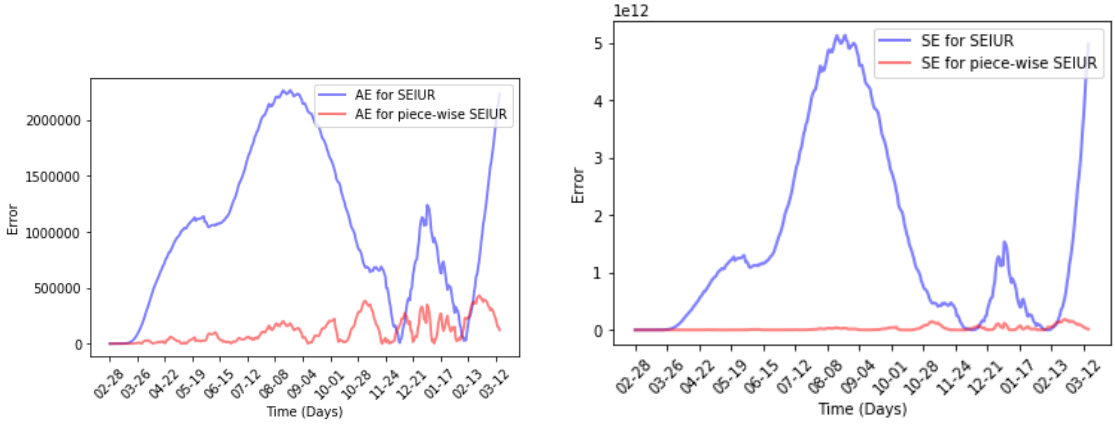

Figure 27: Absolute error and Square error for Piecewise SEIUR and SEIUR

$$\text{MAE}_{\text{PW-SEIUR}} = 1.12 \times 10^5.\ \text{MSE}_{\text{PW-SEIUR}} = 2.50 \times 10^{10}$$

$$\text{MAE}_{\text{SEIUR}} = 1.041 \times 10^6,\ \text{MSE}_{\text{SEIUR}} = 1.55 \times 10^{12}$$

Figure 27 shows that MAE and MSE also gave the same results. No matter which error metric is used, the piecewise SEIUR model has a better performance than the SEIUR model. Quantitatively, the error of piecewise SEIUR model under the MAE metric is only one-tenth of that of the SEIUR, and under the MSE metric it is only one-sixth.

CHAPTER 5

CONCLUSION AND FUTURE WORK

The main achievement of this thesis is to incorporate unreported cases into the SEIR model to obtain the SEIUR model. On this basis, we considered piecewise parameters β and γ and derived the piecewise SEIUR model. These two models are applied to explore the COVID-19 data in the United States. The main conclusion of this thesis is that the piecewise SEIUR model does have higher simulation accuracy than the SEIUR model. The MAPE of the piecewise SEIUR model is only 4%. The error of the piecewise SEIUR model under the MAE and MSE metrics is far less than that of SEIUR model. The outbreak factor generated by the piecewise SEIUR model can be applied to explain the changes of the epidemic prevention strategy in each period. This also provides a mathematical tool for future research on the impact of different strategies on the epidemic.

We introduced the pandemic that affected the world, research background of this thesis, and some methodologies are used. We have developed the improvement methods and reasons for the deficiencies of SEIR model mentioned and introduced the SEIUR model. We discussed the core part of the SEIUR model, that is, the reported and Unreported Cases. But for numerical simulations with a long time period, the model with constant parameters is always difficult to cope with changing situations. Therefore, we gave the concept of a piecewise model, which performs parameter estimation and numerical simulation for each period independently. Due to the modification of the model, the expression of outbreak factor, the form of numerical solution, and the algorithm of parameter estimation is partially changed. We discussed these changes and obtained suitable forms for the piecewise SEIUR model.

We discussed the performance of the SEIUR model and the piecewise SEIUR model for the COVID-19 data. This chapter is mainly divided into two parts. The first part discussed the data, and the second part showed the implementation of the

models. We discussed the data source, the data content, the data preprocessing, the training data and the test data, and the data visualization. In the implementation part, we performed parameter estimation, numerical simulation, and error analysis for the SEIUR model and piecewise SEIUR model respectively.

The difficulty in implementing the piecewise SEIUR model is to determine the data division strategy. Using different division strategies could result in different parameter estimations and numerical simulation results. Therefore, a specific algorithm to find the division strategy of the data is my future work. I plan to use neural networks and deep learning methods to achieve this. The model that can automatically determine the division strategy can be more conveniently used for analysis in more complex situations.

All the code written in Python language. The core code has been uploaded to my github: `https://github.com/ZirenChen/Thesis_Code/blob/main/Python%20code`

REFERENCE

[1] G. Wong, W. J. Liu, Y. X. Liu, B. P. Zhou, Y. H. Bi, *MERS, SARS, and Ebola: The Role of Super-Spreaders in Infectious Disease*, Elsevier Inc., **18**(4), (2015), 398-401.

[2] F. Brauer, *Mathematical epidemiology: Past, present, and future*, Infect Dis Model., **2**(2), (2017), 113-127.

[3] J. Yang, Y. Chen, F. Xu, *Effect of infection age on an SIS epidemic model on complex networks*, J. Math. Biol., **73**, (2016), 1227-1249.

[4] C. I. Siettos, L. Russo, *Mathematical modeling of infectious disease dynamics*, Virulence., **4**(4), (2013), 295-306.

[5] J. M. Carcione, J. E. Santos, C. Bagaini, J. Ba, *A Simulation of a COVID-19 Epidemic Based on a Deterministic SEIR Model*, Frontiers in Public Health, **8**, (2020), 230-243.

[6] Mohan BS, V. Nambiar, *COVID-19: An Insight into SARS-CoV-2 Pandemic Originated at Wuhan City in Hubei Province of China*, Mohan and Nambiar. J., **6**(4), (2020), 146-154.

[7] A. T. Porter, *A path-specific approach to SEIR modeling*, PhD (Doctor of Philosophy) thesis, University of Iowa, (2012).

[8] S. Riedel, *Edward Jenner and the history of smallpox and vaccination*, Proc (Bayl Univ Med Cent), **18**(1), (2005), 21-25.

[9] T.Butler, *Plague history: Yersin's discovery of the causative bacterium in 1894 enabled, in the subsequent century, scientific progress in understanding the disease and the development of treatments and vaccines*, Clinical Microbiology and Infection, **20**(3), (2014), 202-209.

[10] S. Y. Tan, Y. Tatsumura, *Alexander Fleming (1881–1955): Discoverer of penicillin*, Singapore Med J., **56**(7), (2015), 366-367.

[11] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, M. Ciccozzi, *Application of the ARIMA model on the COVID-2019 epidemic dataset*, Data in Brief, **29**, (2020).

[12] Y. Hao, T. Xu, H. P. Hu, P. Wang, Y. P. Ba, *Prediction and analysis of Corona Virus Disease 2019*, PLoS ONE, **15**(10), (2019).

[13] D. Bernoulli, *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour le prévenir*, Histoire et Mémoires de l'Académie des Sciences, (1760), pp. 1-79.

[14] R. A. Ross, *The prevention of malaria*, London: John Murray, (1911), 651-686.

[15] J. Backer, D. Klinkenberg, J. Wallinga, *Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travelers from Wuhan, China*, Euro. Surveill., **25**(5), (2020), 20-28.

[16] N. Anand, A. Sabarinath, S. Geetha, S. Somanath, *Predicting the Spread of COVID-19 Using SIR Model Augmented to Incorporate Quarantine and Testing*, Trans Indian Natl. Acad. Eng., **5**, (2020), 141-148.

[17] W. H. Hamer, *Epidemic disease in England - the evidence of variability and of persistence*, The Lancet., **167**, (1906), 733–738.

[18] W. O. Kermack, A. G. McKendrick, *Contributions to the mathematical theory of epidemics*, Proc R Soc A, **115**, (1927), 700–721.

[19] W. O. Kermack, A. G. McKendrick, *Contributions to the mathematical theory of epidemics*, Proc R Soc A, **138**, (1932), 55–83.

[20] A. Fleming, *On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae 1929*, Bull World Health Organ, **79**, (2001), 780–790.

[21] https://en.wikipedia.org/wiki/COVID-19_pandemic

[22] https://www.ted.com/talks/bill_gates_the_next_outbreak_we_re_not_ready/transcript

[23] https://ourworldindata.org/coronavirus-testing

[24] https://www.worldometers.info/coronavirus/country/us/

[25] https://www.worldometers.info/coronavirus/

[26] https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html

[27] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen

[28] https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics