# DEPARTMENT OF SCIENCE

## MTH301 Final Year Project

## ANALYSE THE SPREAD OF COVID-19 IN UNITED STATES WITH AN EPIDEMIC DYNAMIC MODEL
## 使用流行病动力学模型分析 COVID-19 在美国的传播

Student name: Yipei.Guo

Student ID: 1823319

Supervisor: Mu He

Date: 4/May/2022

# Contents

# List of Figures

# Abstract

The COVID-19 outbreak has caused serious casualties and economic losses to the world. Many countries have adopted control measures such as blockade, isolation, and social distancing. The COVID-19 virus has a strong contagious ability and continues to produce mutant strains, which is more complicated and more contagious than its precedent virus. The complexity comes from asymptomatic infections, critically ill patients requiring medical assistance, and secondary infections in recovered patients. The emergence of these new features adds uncertainty to the basic SEIR model. The contribution of this article is that we use the optimal parameters learned from actual data to establish a nonlinear dynamic epidemic prediction model. The core of the model is to solve ordinary differential equations, and simulate the changes of the epidemic by solving ordinary differential equations. Simulate the U.S. COVID-19 outbreak by building a SEAIR model and an adaptive SEAIRD model. By comparing the prediction results of the SIR, SEAIR and adaptive SEAIRD models, it is found that the COVID-19 epidemic curve simulated by the adaptive SEAIRD model is most consistent with the actual data. The root mean square error of the adaptive SEAIRD model is smaller than that of the SEAIR model and the SIR model. Also observe the Quantile-Quantile map to observe whether the data iterated by the model is similar to the actual data. The adaptive SEAIRD model included 7 different populations representing different evolutionary trends during epidemics. Through the prediction results of the adaptive SEAIRD model, it is convenient to measure the severity of the epidemic, take different preventive measures, and provide considerations for the government to implement measures.

**Keywords:** SEAIR model; adapt SEAIRD model; Quantile-Quantile map; Root mean square error (RMSE)

# 摘要

COVID-19 的爆发对于世界造成了严重的人员伤亡和经济损失，很多国家采取封锁，隔离，保持社交距离等控制措施。COVID-19 病毒具有很强的传染能力，并且持续产生变异株，比其先例病毒更复杂，更具传染力。其复杂性来自于无症状感染者、需要医疗救助的危重患者以及康复患者的继发感染。这些新特征的出现增加了基本 SEIR 模型的不确定性。本文的贡献在于，我们利用从实际数据中学习获取的最优参数，从而建立非线性的动态传染病预测模型。该模型的核心是常微分方程的求解，通过求解常微分方程来模拟疫情的变化。通过构建 SEAIR 模型和自适应 SEAIRD 模型来模拟美国 COVID-19 的爆发，对比 SIR、SEAIR 和自适应 SEAIRD 模型的预测结果，发现自适应 SEAIRD 模型模拟的 COVID-19 流行曲线与实际数据最为吻合。自适应 SEAIRD 模型的均方根误差 (RMSE) 小于 SEAIR 模型和 SIR 模型。通过分位图观测模型迭代的数据是否与实际数据相似。自适应 SEAIRD 模型包括 7 个不同的人群，代表流行病期间的不同进化趋势。自适应 SEAIRD 模型的预测结果方便我们衡量疫情的严重程度，进而采取不同的预防措施，为政府实施措施提供参考。

**关键词：** SEAIR 模型；自适应 SEAIRD 模型；分位图；均方根误差

# 1  Introduction

The Coronavirus disease 2019 is referred to as COVID-19, and it has spread rapidly around the globe world since December 2019. More than 260 million confirmed cases have been found globally, with almost 5.188 million patients associated with deaths. [17] What's more, the Global case fatality rate is approximately 2% which is defined as the number of death divided by the number of confirmed cases. [13,21] Currently, the most severely affected regions worldwide include America, Europe, and South-East Asia. America is one of the most affected countries by COVID-19 in the world, and a total of 9.65 million people in the United States have had or are suffering from COVID-19. In addition, 34.59 million positive cases and 469 thousand deaths in India; 22.09 million positive cases and 614 thousand deaths in Brazil and so on in other countries. [11] COVID-19 has developed into a global pandemic, most countries have confirmed cases in the early stages of the epidemic, and many countries have not had time to implement mitigation measures. [1] Quarantine, keeping social distance, and isolation of the infected are effective measures to curb the spread of the epidemic, which can be exposed in the example of China. However, most countries are not able to implement the drastic measures that China finally adopted. [1] Therefore, the spreading speed and the severity of the epidemic in each region are dissimilar, and will be affected by local policies and social events.

In theory, if individuals' actions were restricted and they were completely isolated, the spread of the virus would be cut off. However, the implementation of these measures must inevitably affect the development of the economy to a great extent because people's movements are restricted and productivity is greatly reduced. [9] Therefore, it is necessary to find a balance between hindering economic development and controlling the spread of the epidemic, the purpose of which is to help society find appropriate control measures and solutions and prevent the collapse of the medical system. [9]

Mathematical model plays an important role in epidemiology, which can predict and estimate disease transmission, recovery rate, and other parameters. [7] A commonly used mathematical modeling method in epidemiology is the compartment model. The SIR model (Susceptible-infectious-recovered) is the most convenient model for epidemiological research and is one kind of compartmental model. SIR model is widely used to simulate the traveling of infectious disease after a big outbreak in epidemiological. SIR is the most basic epidemiology model, which describes how an infectious disease spreads from a small number of carriers to an increasing number of people at the beginning, until people become immune after infection, and the number of infected people gradually decreases. Many countries use the SIR model, or an extended adaptive model, to predict future infection numbers, as well as to inform decisions about implementing and lifting lockdowns. [9]

As writing this article, COVID-19 is still spreading wildly and endangering more lives. It has been observed that there have been some notable changes in the characteristics of COVID-19 as the virus has evolved since the first outbreak in Wuhan, China. [9] The most obvious are asymptomatic infections and reinfections in recovered patients. COVID-19 is highly mutated, and more and more mutated strains are emerging. Among them, two strains discovered in 2020

have been or are becoming the dominant strains of the global epidemic: Delta and Lambda. What is more, the new mutant strain Omicron that appeared in 2022 is also threatening the world under the epidemic. Delta, discovered in India in 2020, spread globally a year later and became the dominant variant in many parts of the world, with an astonishing speed and scope. The lambda strain first appeared in Peru in South America and spread rapidly around the world a year later. According to the available data, the Delta variant strain poses a greater threat to people's health than the original strain: the transmission rate is about 60% higher than that of the Alpha strain, while the Alpha strain has a higher transmission rate than the original strain of the new coronavirus 50%. Patients infected with the delta strain carry about 1,000 times the viral load of those infected with the original strain, and the combination of the shorter incubation period is likely to be the main reason for the unusually rapid spread of the delta strain. [14] Compared with other new coronavirus strains, the biggest difference between Delta is that there is no loss of smell and taste, so it is more like a bad cold. In conclusion, some basic epidemic models are no longer suited to simulate the evolution of SARS-CoV-2. Consequently, it is necessary to extend the traditional SIR model to build new and more complex epidemic models.

In the real world, the spread of COVID-19 may be affected by various events, such as strong government policy interventions, the development of vaccines, or the occurrence of large gatherings. Some events like Black Friday can be predicted, and we can prepare countermeasures in advance to avoid the occurrence of large-scale infectious disease outbreak chains. To address this issue, we extended the traditional SIR model by fitting the data of the different regions, and observing whether there is an event.
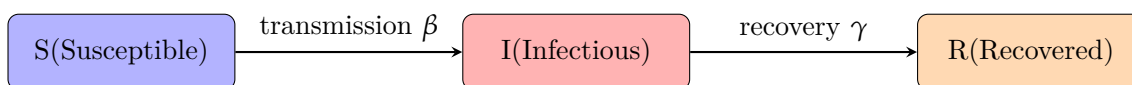
## 2 Literature Review

The earliest infectious data was published by John Grant (1620-1674) in his book "Natural and Political Observations Made upon the Bills of Mortality", which recorded the Mortality and causes of deaths weekly in London parishes between 1592 to 1603. However, the first model in mathematical epidemiology was usually described as the work of Daniel Bernoulli (1700-1782) on inoculation against smallpox. Bernoulli go to study whether Variolation was beneficial because it with a small risk of infection and death if you were vaccinated against smallpox. His approach was to assume that smallpox is eliminated as the cause of death, then the average life span of people will increase, and a brief outline in 1760 was published, followed by a complete exposition in 1766. [19] Before the germ theory of disease appeared, Jhon Snow (1813-1858) established the source of the spread of cholera in London was the public water pump on the Broad Street through the evidence presented by studying cholera in 1832, and he believed that the water source played an important role in the transmission of the bacteria. Their research laid the foundation for people to understand the spread of epidemics. [19]

The origin of the Mathematical model is the early 20th century, with critical research being that of Ronald Ross in 1916. In 1917 Ronald Ross and Hilda Phoebe Hudson published the work on epidemiology which was the SIR model to simulate the transmission mechanism of malarial infection. The transmission of infectious diseases between susceptible and infective was described

by A.G.McKendrick and W.O.Kermack from 1927 to 1933. [9] After that, the SIR model is explored by including more model variables according to the actual situation, and/or changing some assumptions. [15] Hethcote et al. discussed SI, SR with and without dynamic vitality for the long-term epidemic. [2] Berge et al. used the SIR model to simulate the traveling of the 2013 Ebola outbreak in Africa, and they described the endemic equilibrium which means that the infection disappears when susceptible die out. [4] Similarly, Rachah and Torres explain the spread of the Ebola outbreak using the SEIR (Susceptible–Exposed–Infectious–Recovered) model which is a variant of the SIR model, and prove that the full model only has one locally asymptotically stable equilibrium. If the Ebola virus sheds in the environment, it is globally asymptotically. [12] Mpshe et al. published the research based on the SIR model of Zika virus fever outbreak in 2015 in Cape Verde. This model predicts the basic reproduction number of the virus R0 and studies the relative influence of parameters on R0. The paper concluded that the purpose of controlling the epidemic can be achieved by increasing the mortality of Aedes mosquitoes and the recovery rate of patients. [10] Recently, Xian-Xian Liu et al. described SEAIRD predication model, and used it to inversely calculate the time-varying parameters of SARS-CoV-2 in process of spreading. Obtain the infection rate, exposure rate, removal coefficient, and other parameters in the process of American. [9] We can find that most mathematical epidemic dynamical models are used to study infectious disease outbreaks, and they are based on the basic SIR model and its variants, such as SIER, SIERS, etc.

## 2.1 SIR Model

The SIR model is a compartment model, usually expressed by ordinary differential equations. The population is assigned to compartments labeled S, I, or R (susceptible, infected, or recovering), and flows according to the flow pattern between the compartments.

$$\boxed{\text{S(Susceptible)}} \xrightarrow{\text{transmission } \beta} \boxed{\text{I(Infectious)}} \xrightarrow{\text{recovery } \gamma} \boxed{\text{R(Recovered)}}$$

### 2.1.1 Parameters

The model can predict the spread of a single species, and the infectious diseases that produce lasting resistance after recoveries, such as measles, mumps, and rubella. In order to indicate that the population of every compartment is changing over time, we usually express $S$, $I$, $R$ as $S(t)$, $I(t)$, $R(t)$. For a specific disease, we can calculate the outbreak of it and control it.

| Parameter | Meaning |
|---|---|
| (S)Susceptible | Individuals who have not yet been infected with the disease but are susceptible to infection in the future. |
| (I)Infectious | The infectious individual has been infected with disease, and is capable of spreading the disease on to susceptible individuals. |
| (R)Recovered | The recovered individual has had the disease in the past but is not long. |
| $\beta$ | Transmission rate: probability of transmission per contact per unit time per capacity. |
| $\gamma$ | Recovery rate: probability of recovery per contact per unit time per capacity. |

### 2.1.2 Assumption

SIR models usually have some assumptions to complete, we assume that the disease spreading process is deterministic, which means that history and the rules of the model completely determined the behavior of the population. We also assume that the transmission of the disease is modeled based on the derivative of each compartment which means the number of members in the compartment is a differentiable function of time. Following are some assumptions about the differential equation:

- Ignoring births, deaths, and immigration means that the total population remains constant.

- All individuals have the same chance of coming into contact. Therefore, one infectious individual will transmit the disease to $\beta(S/N)$ susceptible individuals in unit time. The total number of susceptible individuals that get infected in unit time is $\beta SI/N$, and move from the susceptible class to infectious class.

- Infectious individuals move to the recovered class with a fixed recovery rate $\gamma > 0$. It follows that the mean infectious period is $1/\gamma$. And after getting sick, antibodies will be produced, which means that individuals who are already sick will not get sick again.

### 2.1.3 Modelling

$$\frac{dS}{dt} = -\beta \frac{S}{N} I \tag{2.1}$$

$$\frac{dI}{dt} = +\beta \frac{S}{N} I - \gamma I \tag{2.2}$$

$$\frac{dR}{dt} = +\gamma I \tag{2.3}$$

This system is nonlinear, but we can drive its implicit solution. First of all, the total number of population is a constant:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \Rightarrow N(t) = S(t) + I(t) + R(t) = Constant \tag{2.4}$$

SIR model is a system of three equations, we analysis the equilibrium of Susceptible and Infectious of time:

$$\frac{dS}{dt} = -\beta \frac{S}{N} I = -\frac{\beta}{N} SI = 0 \tag{2.5}$$

$$\frac{dI}{dt} = +\beta \frac{S}{N} I - \gamma I = I(\frac{\beta}{N} S - \gamma) = 0 \tag{2.6}$$

we can get the nullcline diagram and the point of intersection of the lines of different colors is the point of equilibrium:

$$\frac{dS}{dt} = 0 \Rightarrow -\beta \frac{S}{N} I = 0 \Rightarrow \tag{2.7}$$

$$\boxed{\text{S=0 \quad or \quad I=0}}$$

$$\frac{dI}{dt} = 0 \Rightarrow +\beta \frac{S}{N} I - \gamma I = 0 \Rightarrow \tag{2.8}$$

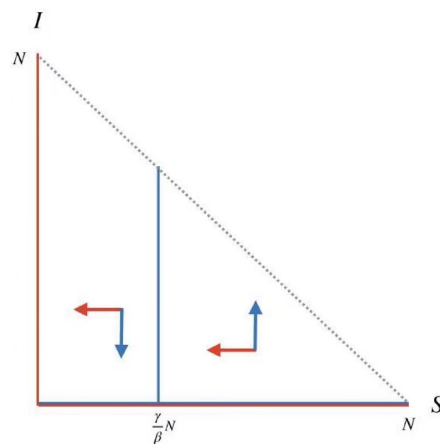$$\boxed{\text{S=}\frac{\gamma}{\beta}N \quad \text{or} \quad I = 0}$$



Figure 2.1:

we deduce the nullcline diagram that the point $S^* = \frac{\gamma}{\beta} N$ is the equilibrium of $I$.

- If $S_0 > S^*$ then the number of $I$ will increase. An epidemic occurs.

- If $S_0 < S^*$ then the number of $I$ will decrease. No epidemic occurs.

The number $S^*$ is called Epidemic threshold.

Using linearisation to classify the equilibrium and calculate the Jacobian matrix at the general point (S,I):

$$\mathbf{J}_{(}\mathbf{S}, \mathbf{I}) = \begin{pmatrix} -\frac{\beta}{N} I & -\frac{\beta S}{N} \\ \frac{\beta}{N} I & \frac{\beta S}{N} - \gamma \end{pmatrix}$$

And we can evaluate the Jacobian matrix at each equilibrium $(0,0)$ and $(S^*, 0)$:

$$\mathbf{J}_(\mathbf{0},\mathbf{0}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{J}_(\mathbf{S}^*,\mathbf{0}) = \begin{pmatrix} 0 & -\frac{\beta S}{N} \\ 0 & \frac{\beta S}{N} - \gamma \end{pmatrix}$$

The eigenvalues of $J_(S^*, 0)$ are 0 and $\frac{\beta S}{N} - \gamma$. The second eigenvalue are positive if $S > S_*$, then the equilibrium is unstable. And negative if $S < S_*$, and the equilibrium is stable.

### 2.1.4 Simulation

Through R, we can establish a graph of S, I, and R against time that can control the Transmission Rate, Recovery Rate, Initial Susceptible Individuals, and Initial Infected Individuals. Code see appendix.
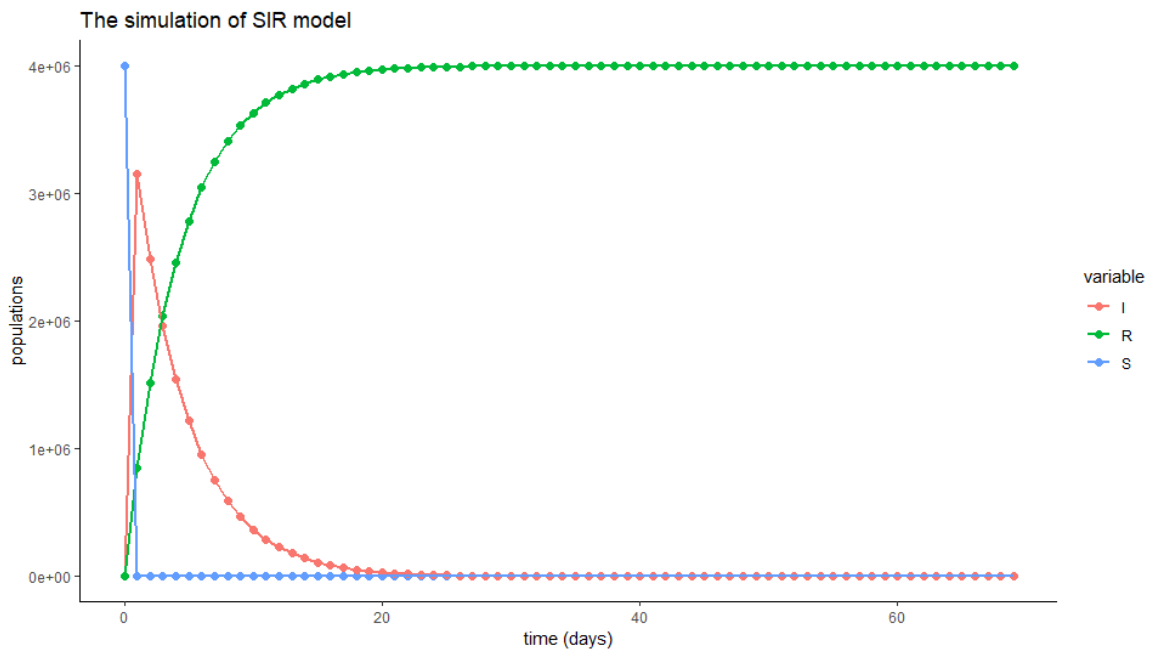


Figure 2.2:

# 3 Methodology

## 3.1 Background

As of the writing of the article, the cumulative number of cases in the United States has reached 80.9 million, and the death toll has reached 990,000. [16] If the U.S. takes appropriate protective measures in time, the infection process and the number of infections and deaths will have different results. The United States has collected and analyzed blood samples from 10,000 volunteers. The main purpose of serosurveys is to study the number of undetected cases of COVID-19 infection, both symptomatic and asymptomatic. In an unexpected result, 20% of New York Citizens may

have been infected with COVID-10. But these people who have been infected don't know that their bodies have faced SARS-CoV-2 and defeated them. [9] Since the outbreak of the new crown in Wuhan, the United States has adopted a series of lockdown measures against China. However, with the increase of time, the epidemic gradually became serious, and the United States implemented a series of policies to control the epidemic.

### 3.1.1  First stage response strategy

The main controls implemented in the United States are as follows, as shown in Figure 1, representing the national-level response:

1 Issue a health alert. On January 16, 2020, the U.S. State Department, in conjunction with the U.S. Centers for Disease Control and Prevention "Watch Level 1 Alert", issued a health alert update, and U.S. citizens in Wuhan, China are reminded not to touch animals and animal products. [18]

2 Continuously improve travel outbreak advice level, and screen flights from Wuhan to the United States. On January 17, the U.S. Centers for Disease Control and Prevention announced that it will carry out entry inspections for passengers traveling directly or connecting from Wuhan, China to the United States, including JFK International Airport, San Francisco International Airport, and Los Angeles International Airport, to test whether passengers have COVID-19. [18] On January 22, the Centers for Disease Control and Prevention announced that it raise the travel outbreak advice to Level 2. Atlanta and Chicago have also started Covid-19 screening for Wuhan flights, and 14 large-scale airports in the United States have been notified of a state of alert. [18] On January 23, 2020, the Centers for Disease Control and Prevention updated its travel outbreak advice to Level 3, advising travelers to avoid all non-essential travel to Wuhan.

3 The U.S. suspended the operation of the U.S. consulate in Wuhan on January 23, 2020, and arranged for planes to evacuate U.S. citizens in China. On January 28, 2020, the United States evacuated approximately 240 U.S. citizens in Wuhan, arranged for them to be quarantined in Alaska, and notified Alaska hospitals to prepare to isolate COVID-19 patients and suspected patients. [18]

The confirmed cases in the U.S. rapidly increased to more than a thousand in the next two months. As of March 13, 2020, the cumulative number of cases exceeded 2,100, and the cumulative number of cases on the 17th exceeded 5,100. West Virginia, the last state in the continental United States to fall, also had its first confirmed case. On March 19, there were already more than 10,000 cumulative confirmed cases of COVID-19 in the United States, about 40% of which were in New York State. [18]

4 On January 30, 2020, the State Department issued the highest level 4 travel alert, reminding U.S. citizens not to travel to mainland China for the time being. [18]

5 Suspend flights to China and deny entry to people who have visited mainland China. On January 31, the United States declared COVID-19 a national public health emergency and

announced anyone who has visited mainland China within 2 weeks will be denied entry unless they are U.S. citizens, green card holders or their relatives. The three major U.S. airlines (Delta Air Lines, United Airlines, and American Airlines) have announced that they will suspend flights to mainland China. A few days later, American Airlines and United Airlines announced the suspension of flights to and from Hong Kong from February 8 to 20. [18]

6 Total closure of consulates in China. Until February 3, according to the guidance of the People's Republic of China, the US embassy in China announced the closure of the embassy, and the consulates general in Wuhan, Chengdu, Guangzhou, Shanghai, and Shenyang will no longer operate externally, and only provide emergency services for US citizens. [18]

### 3.1.2 Second stage response strategy

On February 26, 2020, the President of the United States appointed Vice President Mike Pence to lead the national response to the outbreak. Pence oversees the overall response to COVID-19 and is cooperating with the government health agencies. [18]

7 Cut the benchmark interest rate and free up the Federal Reserve. On March 3, 2020, the Federal Reserve announced a 50 basis point cut in the benchmark interest rate. On March 13, the United States entered a state of national emergency. 50 billion in federal reserves will be released. [18]

8 On March 14, the U.S. House of Representatives passed a coronavirus relief package that includes: Free testing for all COVID-19, including those without insurance; Two weeks of paid sick leave and up to three months for patients; Greater Unemployment Insurance; The U.S. Supplemental Nutrition Assistance Program and other food relief programs for children and the elderly; Provides funding for insurance programs in states where low-income Americans live. [18]

9 The economy has been severely impacted, and many people are facing bankruptcy and unemployment. The U.S. government provides loans to all sectors of society. On April 9, the U.S. Federal Reserve announced that it would provide a total of 2.3 trillion dollars in loans to all sectors of society. [18]

10 Enter a major disaster state and provide a rescue plan. On April 11, U.S. President Donald Trump approved Wyoming as a "hardest hit area" for the 2019 coronavirus disease epidemic. For the first time in U.S. history, all 50 states have entered a "major disaster state." And subsequently put forward a series of food, economic and medical resources assistance programs. [18]

11 Accelerate the development and vaccination of vaccines. On May 15, President Trump officially announced the launch of Operation Warp Speed in the Rose Garden of the White House to accelerate the development of a vaccine. [18] On December 14, the Pfizer-BioNTech 2019-nCoV vaccine was approved by the US Food and Drug Administration, and the US 2019-nCoV vaccination program officially began. [18]

12 On January 21, 2021, after Joe Biden took office as president, he immediately issued several action orders to fight the epidemic, including wearing masks, compulsory quarantine for arrivals at airports, and increasing vaccine production and supply. After that, the United States extended the mask order, launched more economic rescue plans, and launched the "American Jobs Plan". [18]
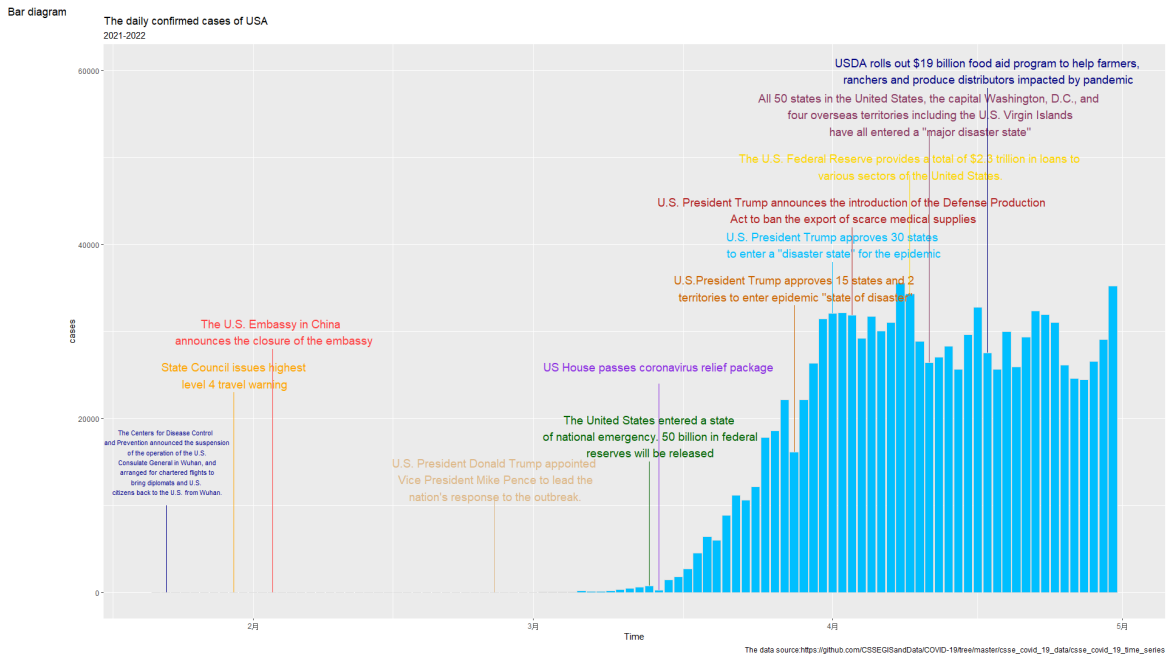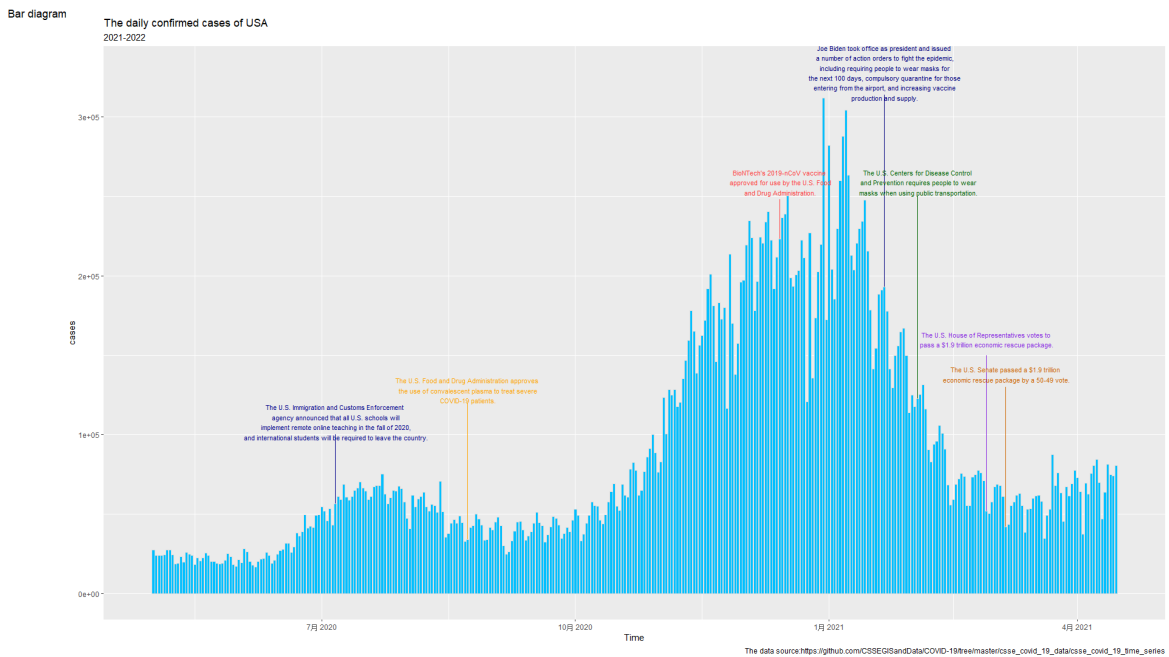


Figure 3.1:



Figure 3.2:

## 3.2 SEAIR model

Since the first outbreak of COVID-19 in Wuhan, China, the characteristics of SARS-CoV-2 have been gradually observed. The most obvious feature is the asymptomatic infection, that is, the patient who is infected with the disease and has the ability to contagious but does not have symptoms such as high fever and cough. Some of the people are infected without knowing it. In epidemiology, infection is without any obvious symptoms, and it is not uncommon for carriers to transmit diseases. Up to 25% of common flu infections have no symptoms. In this case, the virus can be spread through normal breath, which produces tiny droplets. If the cough and fever that are common features of COVID-19 are not shown, the carrier may go undetected and spread the virus to people where he goes, leading to a pandemic. Therefore, classical epidemic models like SIR, SEIR model are unsuitable for COVID-19，and it is recommended to extend the SEIR model to a new adaptive SEAIR model. [3]

Let variables S, E, A, I, R denote the proportions of susceptible individuals, exposed, asymptomatic, infected and recovering. The SEAIR model adds E and A to the SIR model. Group E is the exposed group, this group of people refers to a group of people who have been in contact with infected people but are temporarily unable to transmit the virus to other groups of people. Usually suitable for infectious diseases with a long incubation period.

Group A is the asymptomatic group. This group of people has been infected with infectious diseases but has no significant characteristics. They are more infectious, and this group of people also has the probability of turning into symptomatic infected people. In an asymptomatic state, individuals who spread the disease may have a higher ability to spread the infection, but they do not know that they are infected with the virus.

### 3.2.1 Assumption

In order to be more in line with the actual spread of COVID-19, some assumptions are made based on the SIR model：

- Births, deaths, and immigration are ignored, and the total population remains the same.

- Everyone has the same opportunities for exposure.

- Assuming that the exposure state corresponds to the incubation period of the host carrying the virus, but individuals can not infect other people.

- There are no specific drugs or vaccines to treat at this stage.

- Medical devices have certain restrictions, and some patients may not receive effective treatment.

- After being cured, the patient's body will produce antibodies so that there will be no secondary infection.

- Asymptomatic infections can spread to others.

### 3.2.2 Parameters

Following are some parameters.

| Parameter | Meaning |
| --- | --- |
| S | person who is not Susceptible to the disease, but lacks immunity and is Susceptible to infection after coming into contact with an infected person. |
| E | Exposed the person, refers to a person who have been in contact with an infected person but cannot transmit for the time being, usually appear in infectious diseases with a long incubation period. |
| A | Asymptomatic infections have no obvious symptoms, but they have the ability to spread the disease，so they are not isolated. |
| I | There is an Infectious disease that can be transmitted to a member of class S and become a member of class E or I; |
| R | Recovered, refers to people who have been isolated or Recovered from illness and are immune. If the immunity period is limited, members of class R may revert to class S. |
| $\beta$ | The rate of transmission for the susceptible to exposed. |
| $\xi$ | Rate of asymptomatic infections. |
| $\delta$ | The rate of transmission for the exposed to infectious or asymptomatic. . |
| $\sigma$ | The incubation rate. |
| $\gamma_1$ | Asymptomatic infection recovery rate. |
| $\gamma_2$ | Recovery rate for symptomatic infections. |

### 3.2.3 Modelling

Therefore, we can establish an intuitive SEAIR model. Susceptible people can become exposed people through contact, and a certain proportion of exposed people are infected with COVID-19. Whether they are asymptomatic or symptomatic, they will eventually be cured. This model does not take into account the problem of reinfection, the cured does not become susceptible again. Through the legend, the process of the model can be expressed more clearly and intuitively.
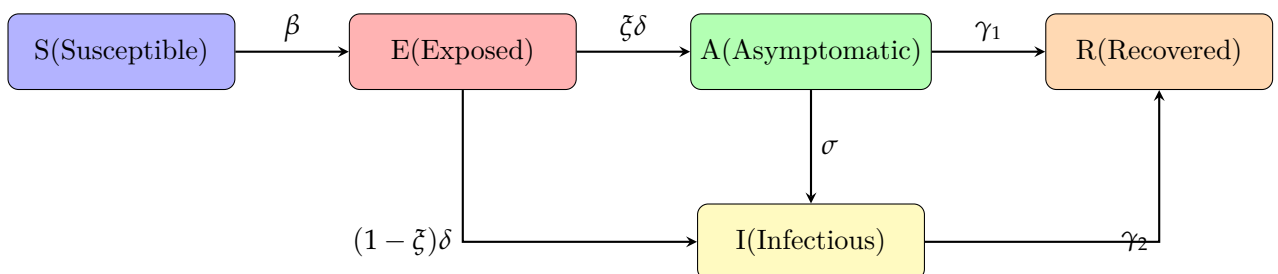


By making assumptions it is possible to list the ordinary differential equations of the SEAIR model, which are the most important part of the model. By solving these ordinary differential equations, we can perform machine learning to plot the predicted model curve.

$$\frac{dS}{dt} = -\beta SE \tag{3.1}$$

$$\frac{dE}{dt} = \beta SE - \delta E \tag{3.2}$$

$$\frac{dA}{dt} = \xi \delta E - (\gamma_1 + \sigma)A \tag{3.3}$$

$$\frac{dI}{dt} = (1 - \xi)\delta E + \sigma A - \gamma_2 I \tag{3.4}$$

$$\frac{dR}{dt} = \gamma_1 A + \gamma_2 I \tag{3.5}$$

$$\frac{dS}{dt} + \frac{dE}{dt} + \frac{dA}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \Rightarrow N(t) = S(t) + E(t) + A(t) + I(t) + R(t) = Constant \tag{3.6}$$

### 3.2.4 Simulating

Use the R language to build the model, and solve the ordinary differential equations by importing the deSolve package. The ordinary differential equation in the SEAIR model is solved by the ode function, in which the setting of the parameters is set according to the results obtained from the actual data, and there is an error within the allowable range. Finally, import the tidyverse package and use ggplot to draw the prediction curve of the SEAIR model. However, there is a certain error between the prediction curve and the actual data.

We suppose the initial susceptible population is S0=4e+06. [20] We start on November 24, 2021, as the initial date. And the exposed is set to E0=10. According to a study of 24 asymptomatic cases of COVID-19 in Nanjing, China, combining data from 24 cases, the median transmission time was 9.5 days. $\gamma_1$ is the recovery rate of asymptomatic infections, and we can thus set $\gamma_1$ to be 1/9.5. [23] Most patients with COVID-19 recover, according to the report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). About 80% of laboratory-confirmed patients had mild to moderate disease, including non-pneumonic and pneumonic cases. 13.8% had severe cases and 6.1% had critical cases. The median recovery time is 14 days for mild patients and about 21 to 42 days for severe patients. Therefore, we assume that the recovery time is 14 days for mild cases, 21 days for severe cases, and 42 days for critical cases. The average recovery time for symptomatic patients by weighted average was 16.66, $\gamma_2$=1/16.66. [6]

According to the World Health Organization's research report on the Wuhan epidemic, and we get patients with COVID-19 typically develop physical symptoms, including mild respiratory symptoms and fever, an average of 5-6 days after infection. [6]

According to U.S. statistics on people infected with COVID-19, the median incubation period from susceptible to infected is estimated at 5.1 days, but 97.5% of people develop symptoms within 11.5 days of infection. These estimates imply that, under conservative assumptions, most susceptible individuals will develop symptoms within 14 days. we can assume that $\delta$ is 1/14. [8]

The probability of asymptomatic infected person showing symptoms is between 0 and 1, we assume 0.2. According to a press conference by the Chinese Health Commission on April 15, 2020, the rate of asymptomatic infections is about 8%. [22]
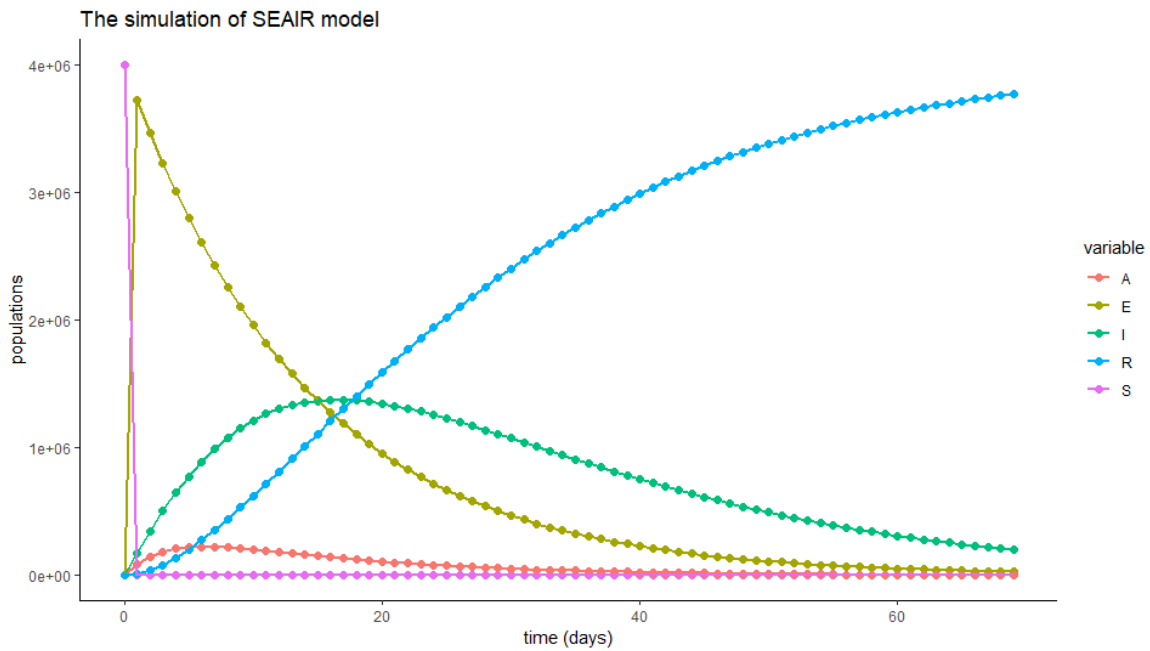
Figure 3.3:

### 3.2.5 Optimization result of fitting curve

The purpose of this article is primarily to analyze the outbreak by collecting data on confirmed cases and deaths in the United States from January 22, 2020 to February 28, 2022. Considering the emergence of new mutant strains, we choose the time series data from December 20, 2021 to February 28, 2022 to construct the model. These data are derived from official data, but there may be some deviations from the actual situation. This paper mainly focuses on building epidemic models using current datasets, focusing on the analysis of epidemic trends.

In the Quantile-Quantile plot below, the x-axis and y-axis are the different datasets, where the x-axis is the actual data, and the y-axis is the dataset simulated by the SEAIR model. Compare the two sets of data side by side to see if they are similar. The two distributions are similar if the scatters almost fall on a line similar to x=y. Most of the scatter points obtained from these two sets of data are concentrated on the straight line of x=y. In general, the number of infected people iterated by the SEAIR model is slightly larger than the actual data, but basically they are relatively close values. This means that the fitting results of this model are quite consistent.

By observing the curve comparison between the actual data and the number of infected people simulated by the SEAIR model, we can observe whether the model is in line with reality. Obviously, the actual data is not smooth, and the data has great fluctuations. However, the overall trend of the simulated data is basically consistent with the actual data. It is feasible to use the SEAIR model to simulate the outbreak of COVID-19, but our model still needs further training and improvement.
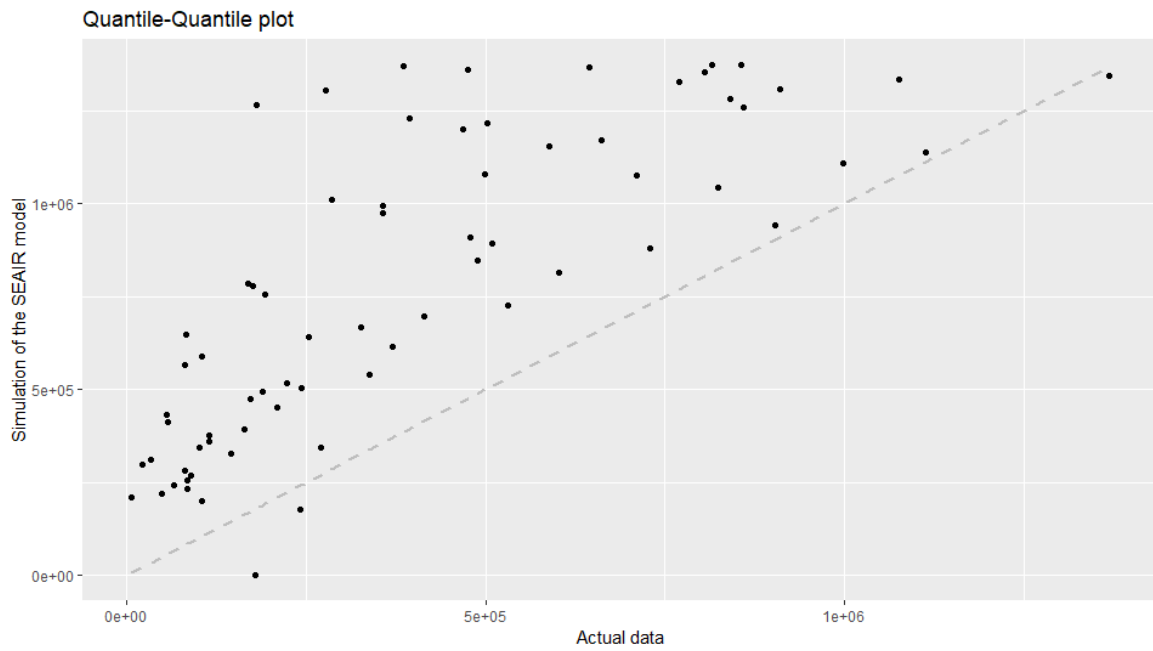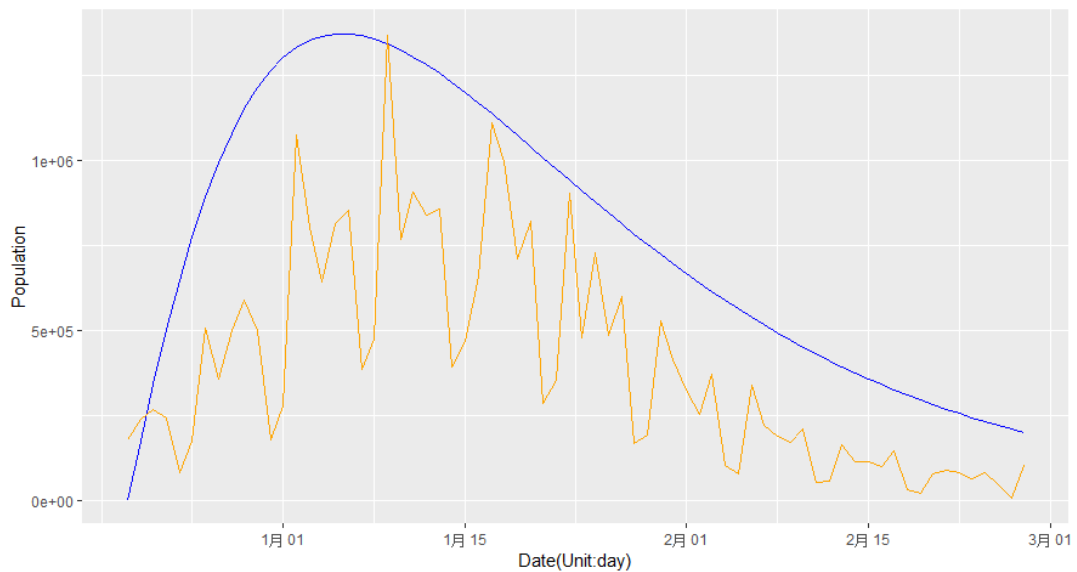
Figure 3.4:



Figure 3.5:

## 3.3  Main method

The classical epidemiological models are similar to SEIR models, or models that add more complex equations. From the perspective of traditional operations research models, these models are essentially a Markov process, and we can make predictions by solving Markov chains. Nevertheless, the usual problem with such classical models is that they are only a statistical feature, yet the convenience of transportation in modern society makes it difficult for these models to fully match the real situation. Therefore, we need to re-improve the SEAIR model so that the

infectious disease model can be mapped to a huge social network in real life. [9]

### 3.3.1 Adapt SEAIRD model

The previous SEAIRD model did not fit some clinical features of COVID-19. Coughing, fever, tiredness, and difficulty breathing are common clinical features of people with COVID-19, although some people have more severe symptoms and require medical attention. Therefore, we divide the infected patients into mild patients and severe patients. The difference between mildly ill patients and asymptomatic patients is that mildly ill patients know that they are infected with COVID-19. People with mild symptoms will be quarantined and their ability to spread will be reduced. Therefore, we need to improve the new model to adapt to the infectious characteristics of COVID-19.
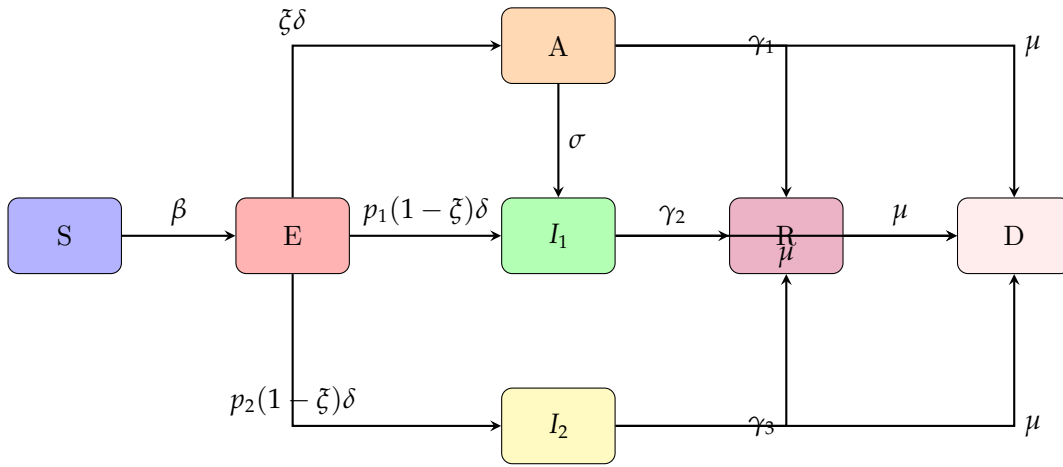


The transmission process of the virus can be represented by ordinary differential equations of population changes of 7 main groups. These seven ordinary differential equations are functions that change with time, and the following are the derivation formulas:

$$\frac{dS}{dt} = -\beta SE \tag{3.7}$$

$$\frac{dE}{dt} = \beta SE - \delta E \tag{3.8}$$

$$\frac{dA}{dt} = \xi \delta E - (\gamma_1 + \sigma + \mu)A \tag{3.9}$$

$$\frac{dI_1}{dt} = p_1(1-\xi)\delta E + \sigma A - (\gamma_2 + \mu)I_1 \tag{3.10}$$

$$\frac{dI_2}{dt} = p_2(1-\xi)\delta E - (\gamma_3 + \mu)I_2 \tag{3.11}$$

$$\frac{dR}{dt} = \gamma_1 A + \gamma_2 I_1 + \gamma_3 I_2 - \mu R \tag{3.12}$$

$$\frac{dD}{dt} = \mu(A + R + I_1 + I_2) \tag{3.13}$$

$$N(t) = S(t) + E(t) + A(t) + I_1(t) + I_2(t) + R(t) + D(t) = Constant \tag{3.14}$$

### 3.3.2 Parameters

The contact rate $\beta$ of susceptible controls the probability of susceptible being in contact with infected during the incubation period. The actual parameter settings of the model are in line with the actual situation.

$\xi$ represents the proportion of asymptomatic infected persons among infected persons. $\delta$ is the probability of exposure to infection. The proportion of untreated SARS-CoV-2 patients is $p_1$, and such patients represent a heavy proportion of symptomatic patients. Most people with mild COVID-19 do not seek medical support due to mild symptoms. Some studies show that about 30%-60% of people with mild symptoms are infected. [9] The percentage of infected people receiving treatment, expressed as p2, is an important ratio that helps reduce mortality. In this model, $\gamma_1$, $\gamma_2$ and $\gamma_3$ represent the recovery rate of asymptomatic infected patients, the recovery rate of mildly infected patients and the recovery rate of severely infected patients, respectively. The death rate is represented by $\mu$.

| Parameter | Meaning |
| --- | --- |
| S | person who is not Susceptible to the disease, but lacks immunity and is Susceptible to infection after coming into contact with an infected person. |
| E | Exposed the person, refers to a person who have been in contact with an infected person but cannot transmit for the time being, usually appear in infectious diseases with a long incubation period. |
| A | Asymptomatic infections have no obvious symptoms, but they have the ability to spread the disease，so they are not isolated. |
| $I_1$ | Mild infected persons,most people with mild infections do not need medical attention. |
| $I_2$ | Severely infected persons, infected persons requiring medical assistance. |
| R | Recovered, refers to people who have been isolated or Recovered from illness and are immune. If the immunity period is limited, members of class R may rever t to class S. |
| D | The dead, whether it is an asymptomatic infection, a mild infection, a severe infection or a cured person, it is possible to become a dead person. |
| $\beta$ | The rate of transmission for the susceptible to exposed. |
| $\xi$ | Rate of asymptomatic infections. |
| $\delta$ | Probability of asymptomatic infected person showing symptoms. |
| $p_1$ | The proportion of untreated patients, which can help us study how many infected people have mild symptoms of infection. |
| $p_2$ | Proportion of patients treated. |
| $\sigma$ | The incubation rate. |
| $\gamma_1$ | Asymptomatic infection recovery rate. |
| $\gamma_2$ | Recovery rate for symptomatic infected persons. |
| $\gamma_3$ | Recovery rate for severely infected persons. |
| $\mu$ | The death rate. |

### 3.3.3 Assumptions

We follow the assumptions of the SEAIR model, and on the basis of these assumptions, add some assumptions that conform to the adaptive SEAIRD model:

- Most mildly infected people do not need to seek help from the medical system, and severe or critically ill patients need to seek help from the medical system.

- If an asymptomatic infection is transformed into a symptomatic infection, it will be a mild infection.

- Whether it is asymptomatic infection, mild infection, severe infection, and recovery, the probability of death is $\mu$.

### 3.3.4 Simulating

The simulation process is similar to the SEAIR model, using the R language to build the model. Import the deSolve package to solve ordinary differential equations. And use the tidyverse package to visualize data. Some of the parameters in the adaptive SEAIRD model are the same as those in the SAEIR model, and the numerical value of the parameters in the SEAIR model are used. The new parameter values need to be added.

We learned from the Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) on the Wuhan epidemic that about 80% of the patients had mild to moderate symptoms, 13.8% had severe symptoms, and 6.1% had critical symptoms. The median recovery time for mild patients was 14 days, assuming a value of $1/14$ for $\gamma_2$. Recovery time for severely symptomatic and critically ill patients is approximately 21 to 42 days. It can be assumed that the recovery time is 21 days for severely symptomatic patients and 42 days for critically ill patients. By calculating the weight, the weighted recovery time of severe patients and critical patients is 27.3 days. From this, $\gamma_3$ is $1/27.3$. [6]

According to the US CDC (Center for Disease Control and Prevention), about 4.64 million people have received medical treatment, but the total cumulative number of patients in the United States is about 81.17 million. Suppose $p_2$ is 5.7% and $p_2$ is 94.3%. [5] As of April 2020, the total number of deaths due to COVID-19 in the United States reached about 990,000, accounting for about 1.2% of the total number of confirmed cases.
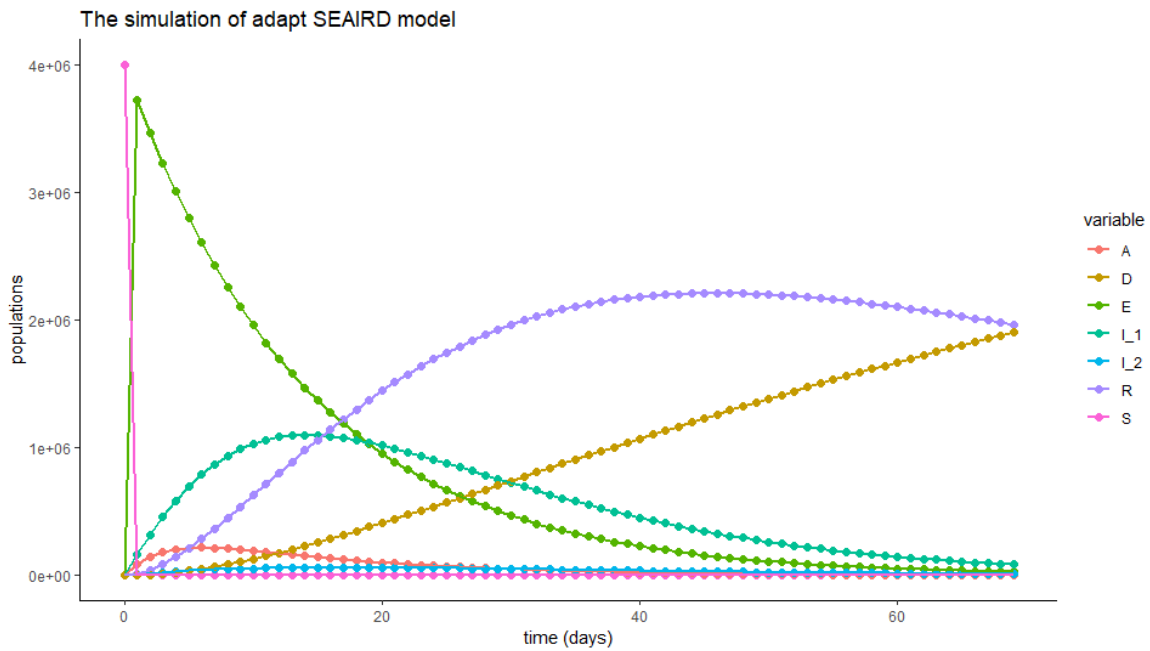
Figure 3.6:

### 3.3.5 Optimization result of fitting curve

Using the same time interval as the SEIAIR model, 69 days from December 20, 2021 to February 27, 2022. Choosing the same time interval allows for a better comparison of the strengths and weaknesses of the models. Similarly, plot the comparison graph and Quantile-Quantile graph of the number of infected people between the actual data and the adaptive SEAIRD model iteration.

By observing the comparison chart, we can find that the overall trend of the actual data of the number of infected people iterated by the adaptive SEAIRD model is consistent, but the actual data will fluctuate to a certain extent. The comparison map results of the adaptive SEAIRD model are not very different from the comparison map results obtained by the SEAIR model.

The distribution of the quantile map is also mainly concentrated on the curve x=y. However, unlike the SEAIR model, the iterative values of the adaptive SEAIRD model and the scatter points drawn by the actual data are concentrated in the lower left corner of the picture. This means that the model is more in line with the real data in the initial time. However, the fit of the two models is basically similar.
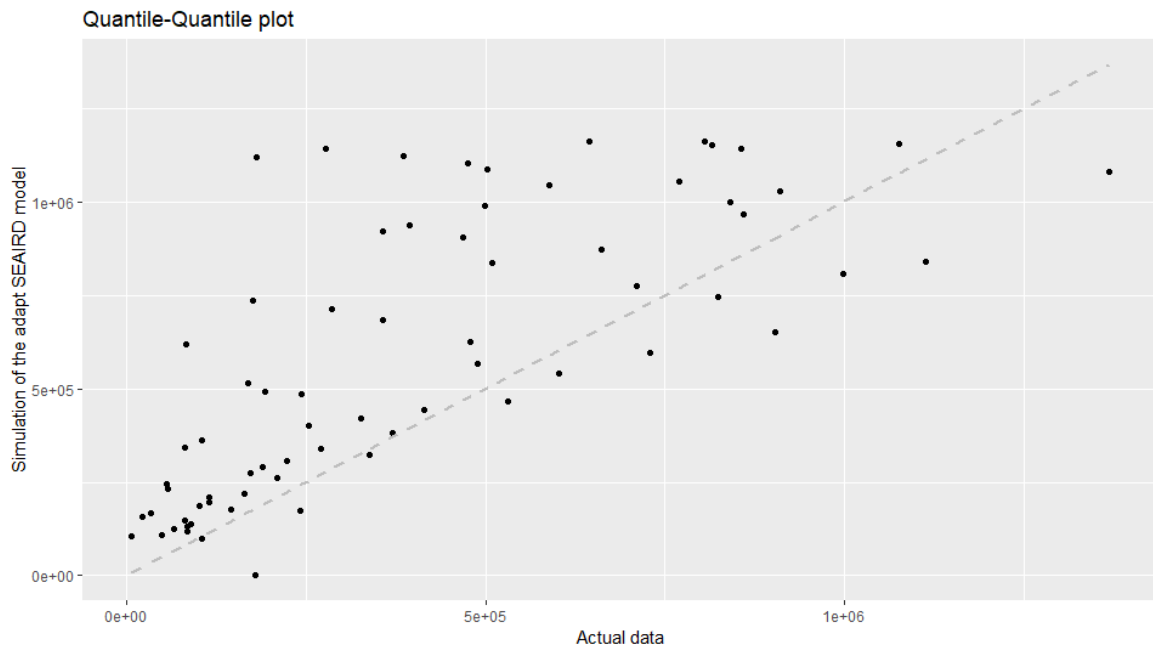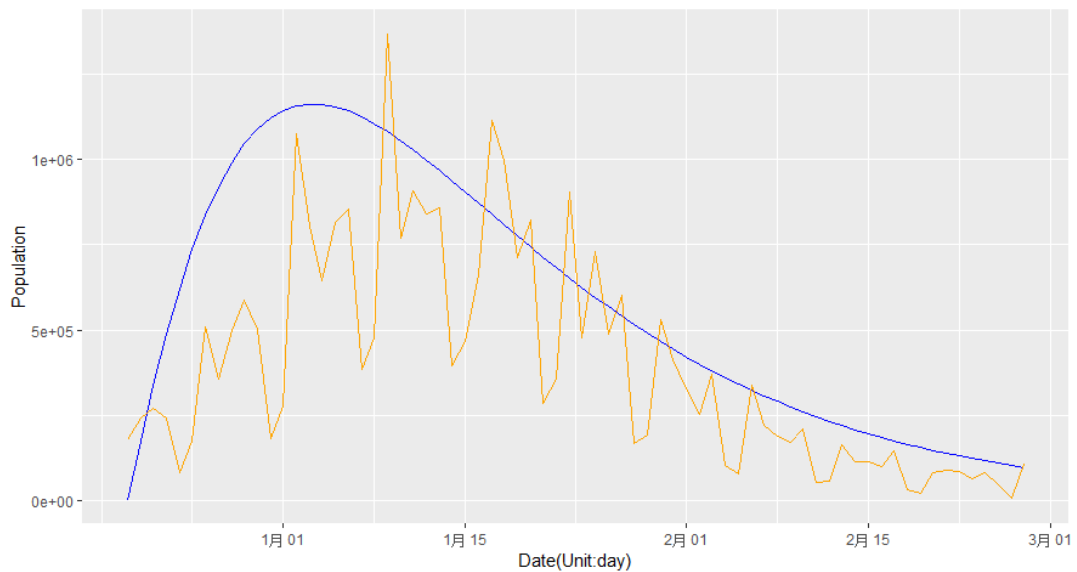
Figure 3.7:



Figure 3.8:

## 3.4 Model comparison simulation experiment

To test the adaptive SEAIRD model proposed in this paper, the root mean square error for judging the fit of the model is introduced. The fit of the model is compared as the root mean square error (RMSE) between the simulated and actual values of the adaptive model. RMSE is often used to measure the error between the real data and the fitted value. The closer the mean square error value is to 0, the better the fitting. The formula for RMSE is as follows, where $\hat{y}_i$ represents the fitted value and $y_i$ represents the true value:

$$RMSE = \sqrt{\frac{1}{n}\sum_{n}^{i=1}[\hat{y}_i - y_i]^2} \qquad (3.15)$$

SEAIR is a commonly used epidemiological model. Compared with the traditional SIR model, the epidemics studied by the SEAIR model have a certain incubation period, and there will be a certain proportion of asymptomatic infections. Closer to the contagion pattern of COVID-19 than the SIR model. According to the root mean square error, it can be seen that the goodness of fit of the adaptive SEAIRD model is smaller than that of the SEAIR model and smaller than the SIR model (the result is the smallest and the best). From the comparison results, the adaptive SEAIRD model has better goodness of fit, and the adaptive SEAIRD model is more in line with the actual spread of COVID-19. The adaptive SEAIRD model outperforms the first two models. The root mean square error values of the three infectious disease models SIR, SEAIR and adaptive SEAIRD model in this paper are as follows:

|  | SIR | SEAIR | adapt SEAIRD |
|---|---|---|---|
| RMSE | 709874.023418092 | 457576.005800768 | 314317.425382512 |

## 3.5   Conclusion and future directions

Using basic models such as SIR and SEIR will be insufficient in prediction after the outbreak of COVID-19. The SARS-CoV-2 pathogen is very different from the pathogens of other epidemic diseases. The main reasons for the inability to use epidemiological models include asymptomatic infections and mild infections not choosing to seek medical care, and higher mortality. Therefore, this paper proposes an improved epidemic model based on the traditional model, the adaptive SEAIRD model.

The complex model proposed in this paper improves the basic model and optimizes the model by analyzing the actual situation, revealing future research directions and potential applications of epidemiological models. The improved model has the following advantages:

1  Detailed parameter, can simulate a variety of interactive infection sources. [9]

2  The adaptive SEAIRD model avoids the drawback of treating each person affected by another person and can aggregate each population into a model based on multiple ordinary differential equations, and the calculation of the model is simple and fast. [9]

3  In models based on ordinary differential equations, individuals are divided into groups according to their characteristics. The model is divided into more detailed and representative, and the model becomes more complex and more realistic.

The compartment model is divided into 7 categories, and a new asymptomatic infection module is added based on the SEIR model. The symptomatic infected persons were refined into mildly infected persons ($I_1$) and severely infected persons ($I_2$), and the removed patients were divided into recovered and dead groups. Natural birth and death rates are not considered in our

model. Simulate ordinary differential equations for different models to predict the development of epidemics. Observe whether the model is suitable by drawing Quantile-Quantile plots and comparison curves.Mainly by comparing the root mean square error, the overall fitting effect of the SIR model and the SEAIR model is inferior to the adaptive SEAIRD model, especially at the early selected times. [9]

There are still many problems in this article that need to be solved urgently, and there are some limitations of the model: The model presented in this paper does not address reinfection in recovered individuals. Some people with COVID-19 will be re-infected. After a while, the people who have recovered may be infected with the disease a second time due to the decline of their resistance. The model only considers the temporal dynamic factors but not the spatial dynamic factors. The independent variable of the adaptive SEAIRD model is time, which means that the model focuses on the change of the epidemic situation over time, and the change in space is not considered. [9]

Further improvements and refinements of the adaptive SEAIRD model in this report. Added some cases that the traditional model did not consider, such as regional lockdowns imposed by countries, different population densities in different regions, imported cases from abroad, some people who are naturally resistant to the COVID-19 virus, as well as moving in and out of people. Therefore, the model needs to be further refined for more accurate classification. It is important to consider the secondary infection of recovered persons, although the number of secondary infections is a very small percentage of the number of infections, it is an important feature. In addition, some actual data was not collected, and many data reflecting the real situation of the epidemic were lost. Therefore, some parameters cannot be determined, and more accurate parameters are needed to build the model. Adaptive models can only reflect the time-varying process of COVID-19, but cannot reflect the spatial spreading process. Therefore, the model will explore the spread of time and space in the future to realize the visualization of data.

# References

[1] Roy M Anderson, Hans Heesterbeek, Don Klinkenberg, and T Déirdre Hollingsworth. How will country-based mitigation measures influence the course of the covid-19 epidemic? *The lancet*, 395(10228):931–934, 2020.

[2] H. Andersson and T. Britton. *Stochastic epidemic models and their statistical analysis.* Stochastic epidemic models and their statistical analysis /.

[3] Lasko Basnarkov. Seair epidemic spreading model of covid-19. *Chaos, Solitons & Fractals*, 142:110394, 2021.

[4] T. Berge, M. S. Lubuma, G. M. Moremedi, N. Morris, and R. Kondera-Shava. A simple mathematical model for ebola in africa. *Journal of Biological Dynamics*, 11(1):42, 2017.

[5] Centers for Disease Control and Prevention. New admissions of patients with confirmed covid-19, united states. `https://covid.cdc.gov/covid-data-tracker/#new-hospital-admissions`,.

[6] Carlos Gomes. Report of the who-china joint mission on coronavirus disease 2019 (covid-19). *Brazilian Journal of Implantology and Health Sciences*, 2(3), 2020.

[7] Zaineb El Kharrazi and Sahar Saoud. Simulation of covid-19 epidemic spread using stochastic differential equations with jump diffusion for sir model. In *2021 7th International Conference on Optimization and Applications (ICOA)*, 2021.

[8] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9):577–582, 2020.

[9] Xian-Xian Liu, Simon James Fong, Nilanjan Dey, Rubén González Crespo, and Enrique Herrera-Viedma. A new seaird pandemic prediction model with clinical and epidemiological data analysis on covid-19 outbreak. *Applied Intelligence*, pages 1–37, 2021.

[10] Saul C Mpeshe, Nkuba Nyerere, and Stephano Sanga. Modeling approach to investigate the dynamics of zika virus fever: A neglected disease in africa. *Int. J. Adv. Appl. Math. and Mech*, 4(3):14–21, 2017.

[11] World Health Organization. Who coronavirus (covid-19) dashboard. `https://covid19.who.int/`,.

[12] A. Rachah and Dfm Torres. Analysis, simulation and optimal control of a seir model for ebola virus with demographic effects. *Communications*, 67(1), 2017.

[13] B Raga, C Khl, D Yjh, E Sr, E Shha, E Sy, F Ghj, G Jl, H Jyl, and H Jwy. Estimation of global case fatality rate of coronavirus disease 2019 (covid-19) using meta-analyses:

Comparison between calendar date and days since the outbreak of the first confirmed case. *International Journal of Infectious Diseases*, 100:302–308, 2020.

[14] Sara Reardon. How the delta variant achieves its ultrafast spread. *Nature*, 21(3), 2021.

[15] Rahul Saxena, Mahipal Jadeja, and Vikrant Bhateja. Propagation analysis of covid-19: An sir model-based investigation of the pandemic. *Arabian Journal for Science and Engineering*, pages 1–13, 2021.

[16] The New York Times. Coronavirus in the u.s.: Latest map and case count. `https://www.nytimes.com/interactive/2021/us/covid-cases.html`.

[17] Johns Hopkins University. Covid-19 dashboard by the center for systems science and engineering (csse) at johns hopkins university (jhu). `https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6,`.

[18] wikipedia. 2019 冠状病毒病美国疫情. `https://zh.wikipedia.org/wiki/2019%E5%86%A0%E7%8B%80%E7%97%85%E6%AF%92%E7%97%85%E7%BE%8E%E5%9C%8B%E7%96%AB%E6%83%85`.

[19] WIKIPEDIA. John snow. `https://en.wikipedia.org/wiki/John_Snow,`.

[20] wikipedia. 美国人口. `https://zh.wikipedia.org/wiki/%E7%BE%8E%E5%9B%BD%E4%BA%BA%E5%8F%A3`.

[21] Worldometer. Covid-19 coronavirus pandemic. `https://www.worldometers.info/coronavirus/,`.

[22] 中华人民共和国国家卫生健康委员会. 2020 年 4 月 15 日新闻发布会文字实录. `http://www.nhc.gov.cn/xcs/fkdt/202004/f2b50e681e7042f8bef2abf2029ffa13.shtml`.

[23] 南京医科大学. 沈洪兵院士团队研究 24 例新冠肺炎无症状传染者提出中后期疫情防控建议. `https://www.njmu.edu.cn/2020/0303/c606a163809/page.htm`.