SIR Modelling for COVID-19 Pandemic Based on MATLAB

Yuchen Song Student No.201830360498

I. Introduction

The project is aimed at building an SIR model for COVID-19 based on confirmed case number data and make a prediction over twice the time span of available data. To achieve this goal, we need MATLAB or OCTAVE as development environment, as well as the accurate data of confirmed case number from a region, a city, a province or a state. I chose to use the data of China during the time period from January 11th to February 29th (50 days in total) and made a prediction between January 11th and April 19th. To estimate the parameter of SIR model, I applied least square method.

II. Methodology

A. Flow of execution

I.There are two functions in task 1, which are model and main. function model is the characteristic function of SIR model (a set of three ordinary differential equations), and the purpose of main function is to find the integral of model function.

II.There are three functions for task 2, they are: prediction(β, γ, S_0), obj(β, γ, S_0), and main(). Just like task 1, the obj function will first find the integral of prediction function(the SIR model ODEs), then it calculates and returns the sum squared residual(SSE) of the predicted confirmed case. When the main function is called, the MATLAB will check the obj function and find the minimum value with corresponding parameters β, γ and S_0. After that. the program will plot the predicted case number and real case number with corresponding time span so we can compare them directly.

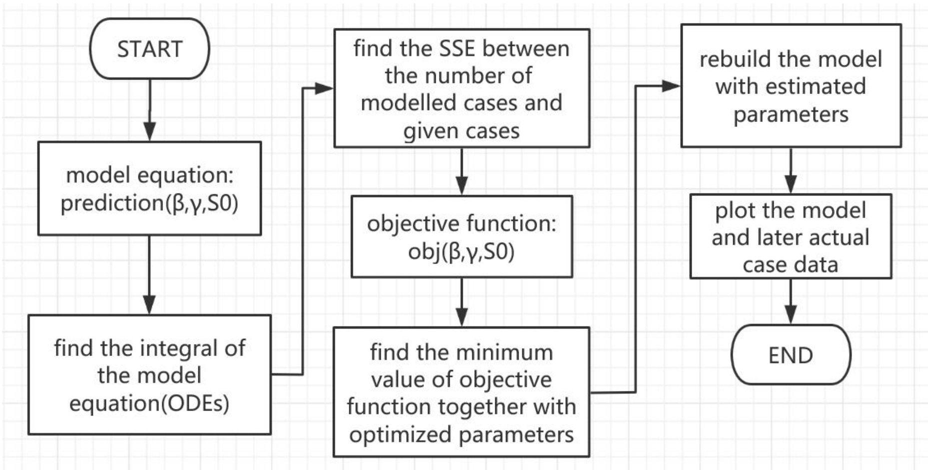

Figure 1 Execution of Each Function

B. Technical Problems

I encountered several technical problems on my way to accomplishing the tasks. The first one was how to use the build-in function ode45 to solve a set of ODEs. After a few trials, I finally regarded y as an array instead of a single variable and solved the problem.

The second obstacle was not easy to access the confirmed cases data. Due to the limitation on linking to the outer net. I had to browse the official website of China National Health Commission to type down the daily confirmed case data manually.

The last problem I met was how to pass parameters to the objective function. Referring to the help document of ode45 and fminsearch, I tackled it by rewriting the integral part like this: ode45(@(t,y)fun(t,y,para),tspan,y0);. But when I replaced S_0 with an unknown parameter, the program kept running and no result was returned. Now I am still working on it and set the value of S_0 manually as a provisional measure.

C. Testing of Program

The program went well and the predicted case number are so close to the real ones that the two chart almost coincide with each other.

III. Results

In task 1 I used the given parameters and did a simple modelling. The chart is as follows. It is clear from the chart chat susceptible population sees a notable downward trend from day 80 to 140, while the recovered cases shows an opposite change. The infected cases, however, rises moderately and at day 120 it reaches a peak, then goes down to 0.

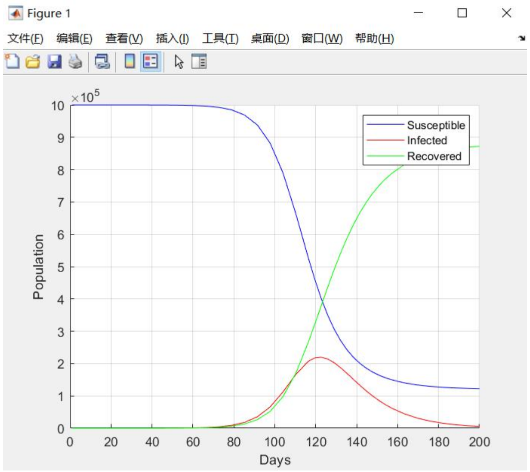

Figure 2 A Simple Modelling

In task 2, I used the number data of 50 days and built a model, the graph is shown like this. The changing tendency of each category is similar to the chart in task 1, the newly added item, confirmed cases are actually the sum of infected cases and recovered cases(no death cases concluded). The curve also illustrates that in day 33, the infected cases reach a peak, which coincide with the real world news. Besides, the confirmed cases slows its growing speed in day 60, when the National Health Commission reported the peak of epidemic has gone.

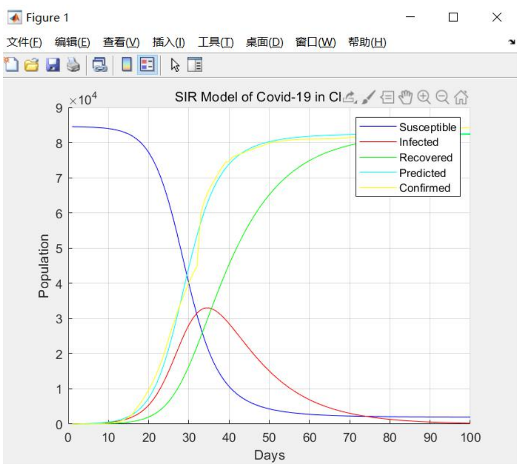

Figure 3 SIR Model Based on Real Data

In addition, I made a few comparisons on how the size of available data influences modelling and find some interesting things.

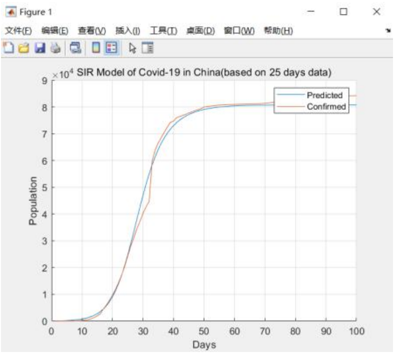

Figure 4-1 Based on 25 days

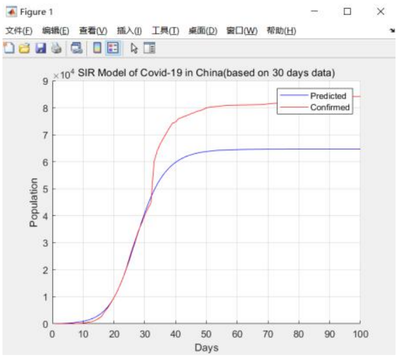

Figure 4-2 Based on 30 days

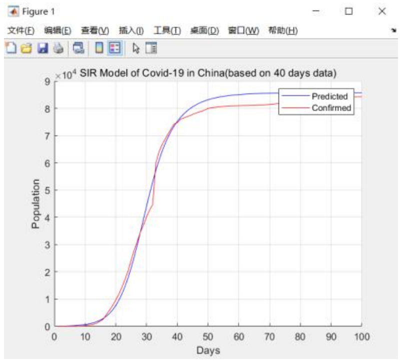

Figure 4-3 Based on 40days

It can be seen from the three graphs that the model based on 30 days data has a relatively large error. Considering the data, we can figure it out that the newly added confirmed case rocketed on February 12th, which is exactly the 33rd day of the data span. It has been claimed that on the very day, National Health Commission added clinical diagnosis cases to the confirmed cases, also making a calibration on the data forward. That is to say, some data before February 12th are not accurate enough, which causes the error to our chart.

Another finding is that when the available dates are less than 22 days, we cannot build an SIR model successfully.

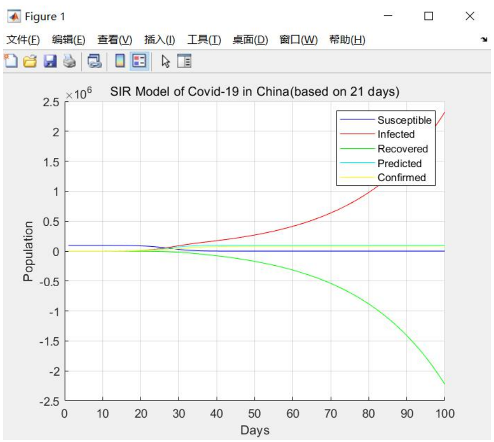

Figure 5-1 Based on 21 days

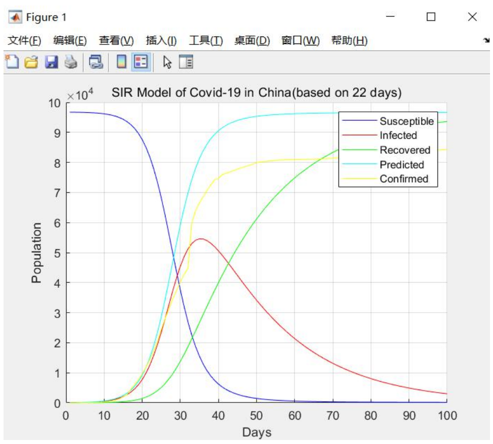

Figure 5-2 Based on 22 days

We can guess that the least dates of data needed has some relationship with the inflection point of the curves(which seems close to day 22).

IV. Conclusion and future development

In this project, I built an SIR model based on the COVID-19 confirmed case data of China, using the build-in function ode45 and fminsearch with MATLAB and made a prediction for 100 days. The modeled data fits the actual data very well.

Future works include comparing SIR model with another famous epidemiological model SEIR, finding the least data required to estimate a model with high accuracy, and trying to use more data from different countries to find the influence from government implementation such as asking civilian to stay at home or keeping tracking and curing the infected ones.

Appendix

A. User Manual

I. Read the "readme.txt" file before running the code.

II. Copy the data file "China_culmulative_confirmed_case.txt" to the same folder of "main.m".

III. Run the main.m to see the task output.

IV. If you want to modify the size of input data, change it in obj.m at line 14.

V. Always close the figure window before the second running.

B. Data

The data used in the program comes from the website https://news.sina.cn/zt_d/yiqing0121, more information can be found in file "China_culmulative_confirmed_case.txt" in the fold named "task 2"