VG101 Matlab final report
Confirmed case number prediction——based on SEIR model

1. Introduction

This program includes 5 subprograms: compara_Realdata.m, compara_Simulateddata.m, dofmincon.m, Getdata.m, and Task1.

Task1 related to the task1. It can draw the quantity change chart based on SIR model.

Other 4 programs are related to task2. By input available data, they can find the parameters of SEIR model and use these predicted parameters to predict the change of confirmed case.

Fmincon function and the solve ordinary differential equations are the main parts of these programs. During the process of programming, I encountered many problems because I was not so familiar with the grammar of matlab. As for the testing of the program, I try much different data to test the model to sure that its correctness.

More detailed Instructions will be attached in the appendix.

2. Methodology

2.1 SEIR model

$$dS/dt=-r*beta*I*S/N$$

$$dE/dt=r*beta*I*S/N-alpha*E$$

$$dI/dt=alpha*E-gamma*I$$

$$dR/dt=gamma*I$$

To simplify the model, in this program r is taken as '1'. So, r will not be discussed in this report.

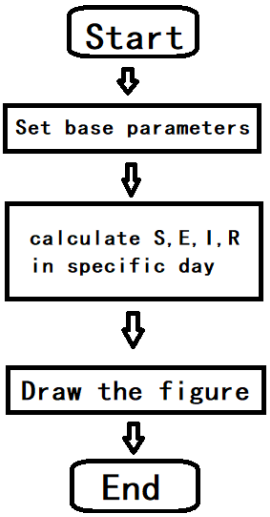
S means susceptible persons, E means exposed persons, I means infectious persons, R means recovered persons, t means time and N, which equals to S+E+I+R, means total population.

There are 3 parameters in this model, beta which means contact rate by an infected people at a point in time, gamma which means recovery rate, and alpha which means rate of conversion from exposed to infected.

2.2 The flow of execution of each function

2.2.1 Task1

To complete task1, this subprogram is based on **SIR model**.



Basic parameters like N, I(0), S(0) , R(0), beta, gamma and time should be set.

By using the model of SIR, the value of s,I,r in specific day can be calculated. Here are the codes of calculation part.

```
T = 1:200;    %time
for idx = 1:length(T)-1%%SIR model
    S(idx+1) = S(idx) - beta*S(idx)*I(idx)/N;
    I(idx+1) = I(idx) + beta*S(idx)*I(idx)/N - gamma*I(idx);
    R(idx+1) = R(idx) + gamma*I(idx);
End
```

The next 4 programs are related to task2.

2.2.2 Getdata

This subprogram is used to calculate the confirmed case number C=I+R.

The execution of this program is the same as Task1, but the calculation part is a little difference. Here are the codes of calculation part:

```
T = 1:100;   %day
for idx = 1:length(T)
    C(1)=1;
    S(idx+1) = S(idx) -beta*S(idx)*I(idx)/N;  %SEIR model to
calculate S E I R
    E(idx+1) = E(idx) + beta*S(idx)*I(idx)/N-alpha*E(idx);
    I(idx+1) = I(idx) + alpha*E(idx) - gamma*I(idx);
    R(idx+1) = R(idx) + gamma*I(idx);
    C(idx)=I(idx)+R(idx); % confirmed case number
end
```

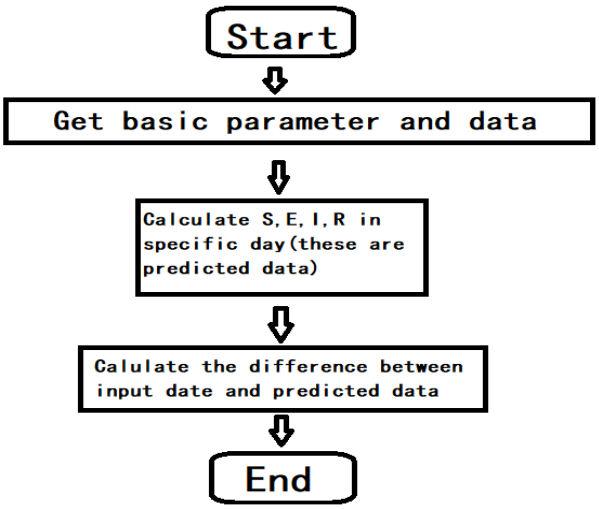
C is set to calculate the confirmed case.

2.2.3 Dofminon

This program is used to find the parameters of SEIR model based on the input of available data.

2.2.3.1 function y=findmin(x,sdata)

This function is used to calculate the difference between available data and predicted date.



According to task2, in this program, beta, gamma, alpha and S(0) are **unknown**. They are the parameters we should find by available data.

Here are the codes in the part of calculating the difference.

```
minnumber=0;
for i=1:L
    minnumber=minnumber+(sdata(i)-pdata(i))^2;
end
y=minnumber;;
```

Sdata means the available data, pdata means the predicted data.

2.2.3.2 function y = do_SEIR(sdata)

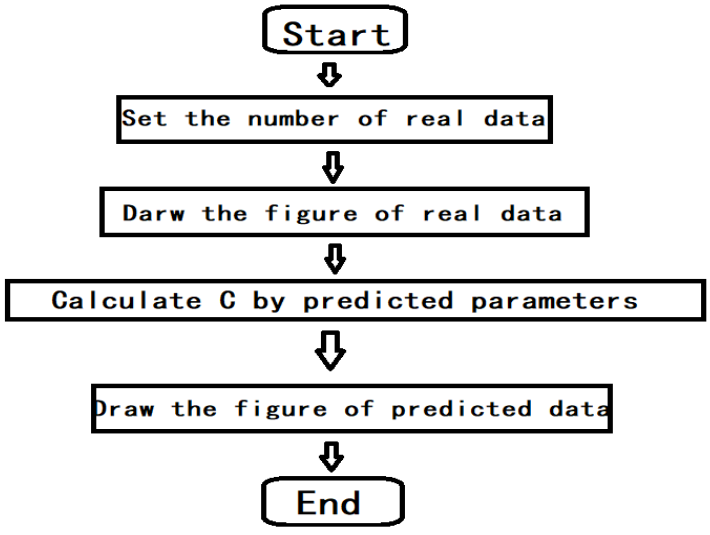
"fmincon" function is used in function. It can find the value of beta, gamma, alpha and S(0) when the minnumber(it is mentioned in the **findmin** function) is the smallest.

Fmincon function is chosen but not fminsearch because parameter range can be set in fmincon. It can prevent the occurring of some strange solutions.

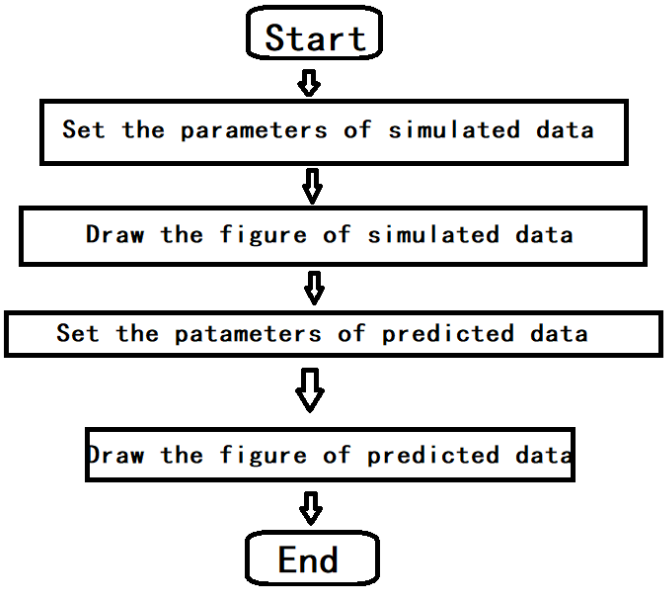
The output of this function are beta, gamma, alpha and S(0), which are predicted based on the available data.

2.2.4 compare_Realdata

This function is used to compare the difference between predicted confirmed case number and real confirmed case number



2.2.5 compare_Simulateddata



To confirm the correctness of the model, the user can try to set parameters in Getdata program. Then input half of the date in Dofmincon program and get the predicted parameters. Finally put the predicted data into these programs to see whether the model can predict correct parameters. If it can, the C from predicted parameters will be similar to the C the user calculated in the Getdata program.

More detail of this process will be attached in the appendix.

2.3 Technical problems

The first problem I encountered is the use of and fmincon, it took me a long time to understand how to use this function. There were many professional nouns I didn't know. For instance, function handle. By reading the help of matlab, I understood them finally.

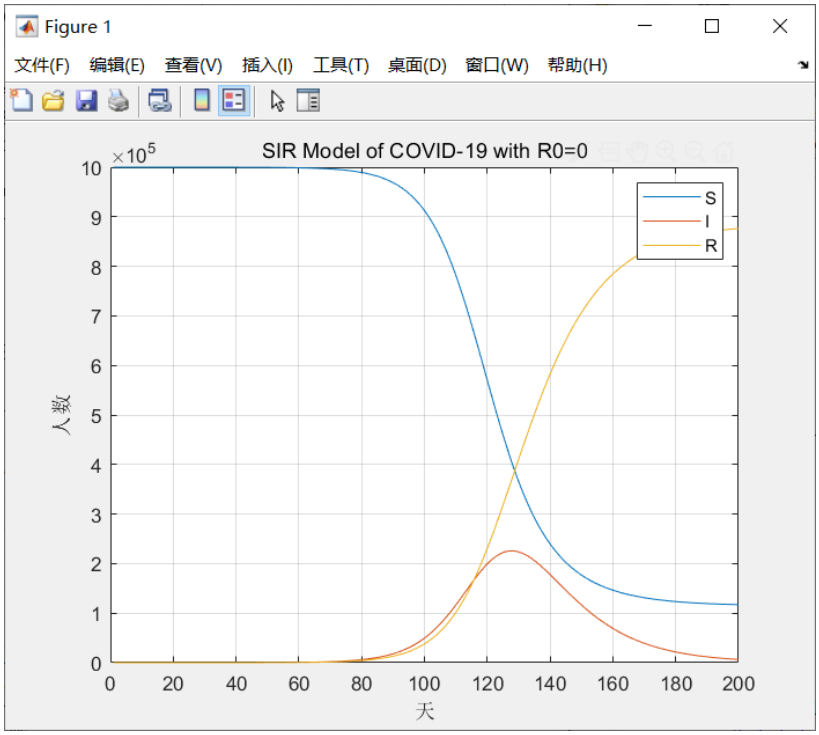
Another problem is also about fmincon. In the process of testing, I suddenly found that the output became a constant, no matter how I change the input. I tried to change the input again and again. Finally, I found that the first data of the input should not be 0. So, I tried to set it as 1 and solved the problem.

Last but not least, the solution of ordinary differential equation. In fact, we had not learned how to solve ordinary differential equation. So, it was a challenge for me to understand the SEIR model. After searching many data, I could understand it and program it in the matlab.

2.4 Testing

3. Results

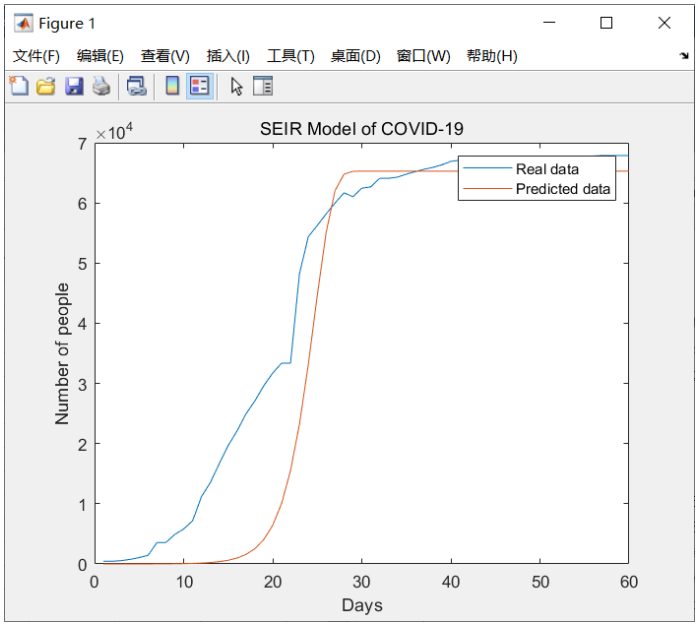
3.1 Task1



$\beta = 0.2; \gamma = 0.083; N = 1000000; I0 = 1; R0 = 0;$

3.2 Task2

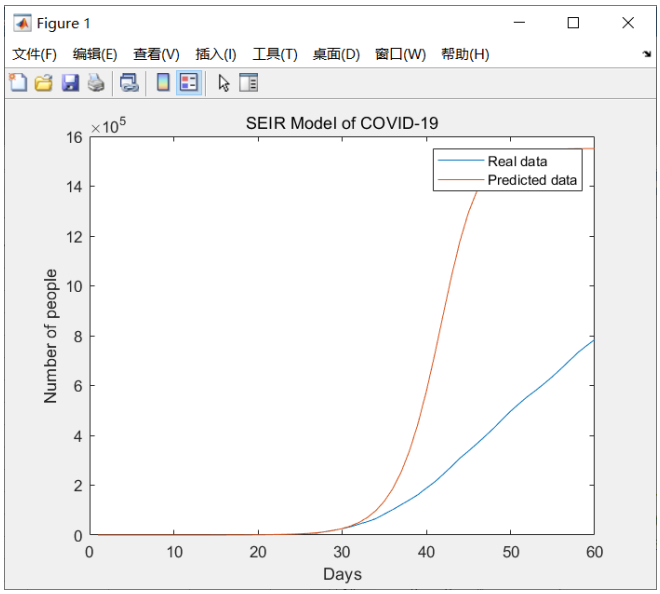
3.2.1 The data from Wuhan



These data comes from https://github.com/datasets/covid-19/blob/master/data. Here are the detailed numbers (1.22-3.21):

444 444 549 761 1058 1423 3554 3554 4903 5806 7153 11177 13522 16678 19665 22112 24953 27100 29631 31728 33366 33366 48206 54406 56249 58182 59989 61682 61031 62442 62662 64084 64084 64287 64786 65187 65596 65914 66337 66907 67103 67216 67332 67466 67592 67666 67707 67743 67760 67773 67781 67786 67790 67794 67798 67799 67880 67880 67880 67880

3.2.2 The data from America

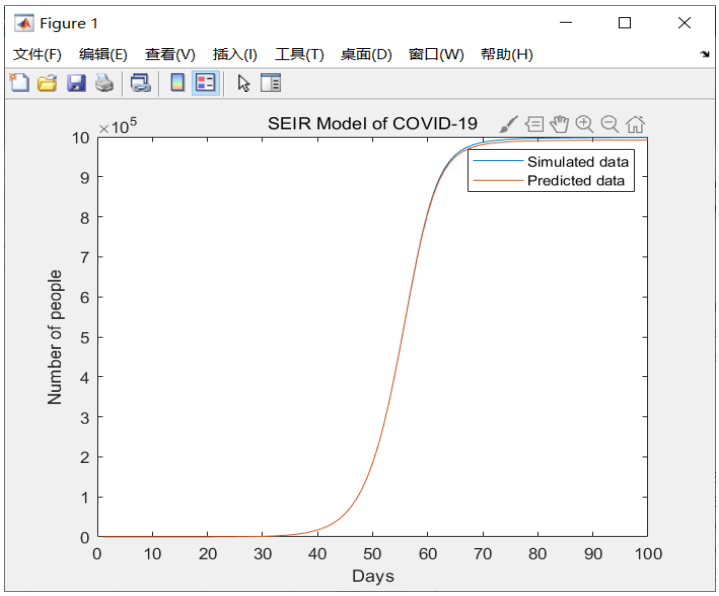


These data also come from https://github.com/datasets/covid-19/blob/master/data. Here are the detailed numbers(2.21-4.20):

15 15 15 51 51 57 58 60 68 74 98 118 149 217 262 402 518 583 959 1281 1663 2179 2727 3499 4632 6421 7783 13747 19273 25600 33276 43843 53736 65778 83836 101657 121465 140909 161831 188172 213242 243622 275367 308650 336802 366317 397121 428654 462780 496535 526396 555313 580619 607670 636350 667592 699706 732197 758809 784326

It seems that there is difference between theory and practice.

3.2.3 The simulated data



Beta=0.5, gamma=0.083, alpha=0.8, S0=999999, day=100. Here is the simulated data calculated by these parameters:

1 1 1.39999960000000 1.84679896980018 2.43251331164868

3.18353223168473 4.14928484140181 5.39069912313295

6.98653742919847 9.03797340161828 11.6750764236322

15.0650457256202 19.4228117726078 25.0246604622387

32.2257458638662 41.4826000563864 53.3820650955391

68.6784777519343 88.3414584727443 113.617324294681

146.108002514046 187.872420402780 241.556752927187

310.561709728798 399.257340826677 513.258770406219

659.779994150501 848.087599001030 1090.08222667609

1401.04308661702 1800.58014839038 2313.85014239445

2973.10647945845 3819.66986378407 4906.42565686137

6300.97535995907 8089.59134378618 10382.1428612792

13318.1712088837 17074.2816697730 21872.9711118904

27992.8930711258 35780.3309707605 45661.2378377185

58152.5167028908 73870.1510017589 93530.2484571862

117937.012739915 147949.303984298 184415.461664757

228065.921843179 279357.433426034 338274.777984915

404118.467435584 475337.937919666 549497.147823142

623458.625083945 693816.147605999 757494.734898111

812322.804707169 857355.375723146 892836.106152244

919864.979663552 939957.200352648 954668.599029796

965369.003686728 973155.024484627 978852.717955803

983061.268693276 986206.067293454 988586.146617295

990411.333442490 991829.399006528 992945.165615343

993833.693169360 994549.306162833 995131.767592085

995610.517208653 996007.600320472 996339.708798240

996619.616888367 996857.201399089 997060.173973279

997234.612054361 997385.347740633 997516.255340216

997630.466021406 997730.529497541 997818.536877801

997896.214791189 997964.998077583 998026.086355492

998080.488364416 998129.056966731 998172.516961127

998211.487325117 998246.499111218 998278.009930481

998306.415739972 998332.060487801

4. Conclusion and future development

4.1 Summary

During the process of programming, I gained lots of experience.

Make a good use of the internet. I met a lot of problems I don't know, and I searched the internet for help. Despite the fact that I could not find the answer every time, it really helped me a lot.

Be patience and calm. I would feel angry and upset when I did not know how to solve some problems. It reduced the efficiency of my work and drove me in a bad mood. When I calmed down and thought it once again, the solution would be clearer.

Reasonable time allocation. I had both C++ work and Matlab work to do, and time was limited. So, how to allocate time was a problem for me. After this project, I know how to allocate my time

when I need to do many works in limited time.

Communication with friends. It's significant to communicate with friends to change our ideas. It will spark more ideas and make all of us go farther.

Every cloud has a silver lining. (In Chinese, we can say "**山重水复疑无路，柳暗花明又一村**")

4.2 Extension

If I have more time, I can do much improvement.

Perfect figure. I will try to make more labels for the figure if I have more time.

More accurate model. SEIRS model may be a more interesting model to try. Or I can try to set r as a parameter in the model.

More detailed notes.

5. Appendix: User Manual

5.1 Introduction

5.1.1 The purpose of this user manual

This user manual was written to explain to the user how to use these programs.

5.1.2 The function of this program

By setting parameters, this program can predict the change of S(Susecptible persons),E(exposed persons),I(the number of infected persons) and R(recovered persons) by using SEIR model.

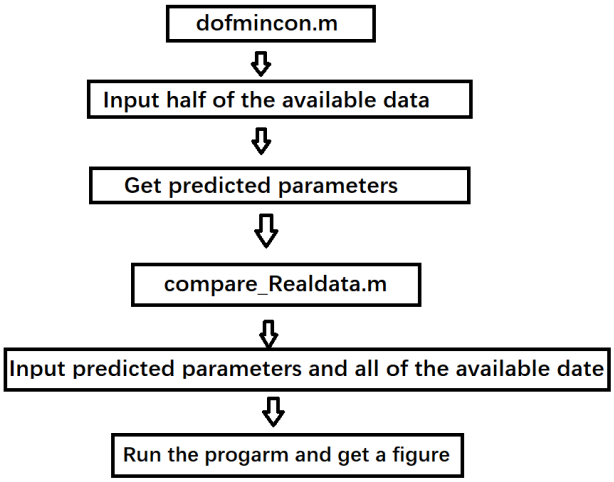
By entering the data of confirmed cases from one area in a period of time, this program can predict the change of S(Susecptible persons),I(the number of infected persons),E(exposed persons),R(recovered persons) by using SEIR model.

5.2 The use of this program

5.2.1 Draw the figure of S,I,R by default parameters.

Open the "Task1.m" program and run it by matlab.

5.2.2 Predict the confirmed case by available data and compare them



1.Open dofmincon.m and compare_Realdata by matlab. First, go to dofmincon. Input half the available data into array “data” and run the program.

```
%These data are used to predict, you should reset it
data=[444 444 549 761 1058 1423 3554 3554 4903 5806 715
```

2.Then you can get predicted parameters as output in y. y(1)=beta, y(2)=gamma, y(3)=alpha, y(4)=S0

y

1x4 double

1	2	3	4
1.0000	1.1744e-15	1.0000	6.5277e+..

3.Now go to compare_Realdata, input all of the available data into array ‘C’.

```
%C is the real data, please reset it.
C=[15 15 15 51 51 57 58 60 68 74 98 118 1
```

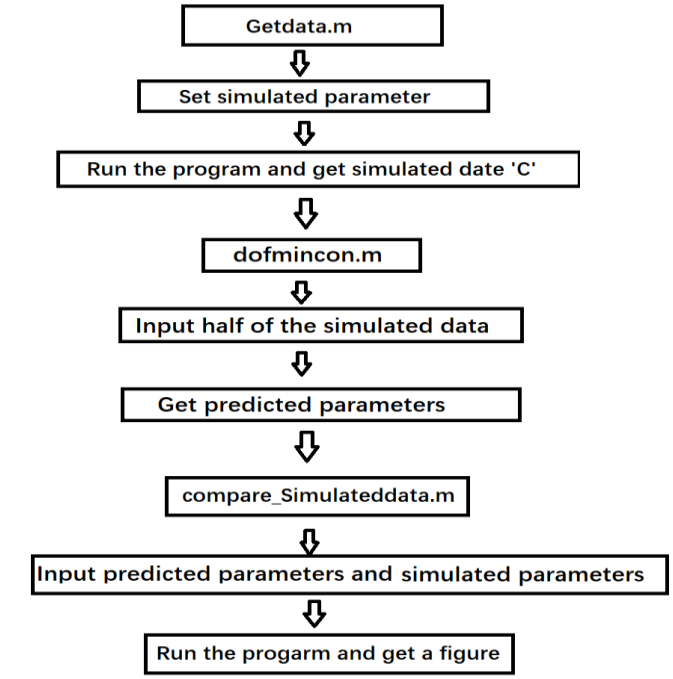
4.Then change the default parameters into predicted parameters and run the program.

```
E = 0;          %exposed persons
I = 1;    %the number of infected person
S = 1.680782364879681e+06;  %Susecptible persons
N = S+I; %Total population
R = 0;  %recovered persons

beta = 0.358104550138834;%contact rate by an infected people at a point in time
alpha = 0.999952070203294;  %rate of conversion from exposed to infected
gamma = 0.494767599018241;  %recovery rate
```

5.2.3 Predict the confirmed case by simulated data and compare them

This function is used to confirm the correctness of the SEIR model.



1. Open dofmincon.m, compare_Simulateddata.m, and Getdata.m. Go to Getdata first.
2. Reset the parameters as what you want and run the program, you can get 'C' as simulated data.

C

1x100 double

1	2	3	4	5	6	7	8	9	10	1
1	1	1.4000	1.8468	2.4325	3.1835	4.1493	5.3907	6.9865	9.0380	1

```
N = 1000000; %Total population
E = 0;         %exposed persons
I = 1;   %the number of infected person
S = N - I;  %Susecptible persons
R = 0;  %recovered persons

beta = 0.5;%contact rate by an infected people at a point in time
alpha = 0.8;  %rate of conversion from exposed to infected
gamma = 0.083;  %recovery rate

T = 1:100;   %day
```

3. Go to dofmincon.m, input half of the simulated data into array 'C'. Run the program and get the predicted patameters as 'y'. y(1)=beta, y(2)=gamma, y(3)=alpha, y(4)=S0.

```
%C is the real data, please reset it.
C=[15 15 15 51 51 57 58 60 68 74 98 118 1
```

y

1x4 double

1	2	3	4
1.0000	1.1744e-15	1.0000	6.5277e+..

4. Go to compare_Simulateddata.m. Input the parameters of simulated data you set in Getdata in the simulated data part, and input the parameter of predicted data in the predicted part, then run the program. (T should be set and should be same in this program)

```
%These are the parameters of simulated data, please reset them

E = 0;         %exposed persons
I = 1;   %the number of infected person
S = 999999;  %Susecptible persons
N = S+I; %Total population
R = 0;  %recovered persons

beta = 0.5;%contact rate by an infected people at a point in time
alpha = 0.8;  %rate of conversion from exposed to infected
gamma = 0.083;  %recovery rate

T = 1:100;   %day
```

```
%These are the parameters of predicted data data, please reset them

E = 0;          %exposed persons
I = 1;    %the number of infected person
S = 9.939704436822236e+05;  %Susecptible persons
N = S+I; %Total population
R = 0;  %recovered persons

beta = 0.531156078724007;%contact rate by an infected people at a point in time
alpha = 0.701989654571206;  %rate of conversion from exposed to infected
gamma = 0.091752660308921;  %recovery rate

T = 1:100;    %day
```

5.3 Attention

You have better to open all programs as required at the same time.