

# Post-Modeling for Epidemiological Parameters of COVID-19

LI CHENGYANG

May, 2020

---

## 1. Introduction

COVID-19, short for Coronavirus disease 2019 (COVID-19), is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). As of 30 May 2020, more than 5.91 million cases have been reported across 188 countries and territories, resulting in more than 364,000 deaths. More than 2.49 million people have recovered. [1]

SERID model, developed for the 2014 Ebola epidemic occurring in the West African nations of Guinea, Liberia, and Sierra Leone, is a popular way to model infectious diseases by using compartmental models. This compartmental model divides population into compartments, assuming that every individual in the same compartment shares the same characteristics. The model describes the dynamical interaction of susceptible and infected populations, while accounting for the effects of hospitalization and the spread of disease through interactions with deceased, but infectious, individuals.[2]

In this report, we are about to simulate the SEIRD epidemic model and estimate its parameters from the data provided by New York Times[3] and the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.[4] The parameters due for estimation include the case fatality ratio denoted as  $\alpha$  and the basic reproduction number denoted as  $R_0$ , which are the most concerned and practical numeric indicators. In

epidemiology. The simulation consists of two stages. First is to utilize optimization method to estimate the wondered epidemic parameters from the existing data and the second is to apply these indicators to simulate future curve of death cases.

## 2. SEIRD Modeling

### 2.1 Model Explanation

A SEIRD model divide population into five classes:

- Susceptible (S)
- Exposed (E)
- Infected (I)
- Recovered, or Resistant (R)
- Dead (D)

Total population is constant and equal to N and the evolution of the number of people in each group is as follows:

$$\begin{cases} dS(t) = -\beta S(t)I(t)/N \\ dE(t) = -\beta S(t)I(t)/N - \sigma E(t) \\ dI(t) = \sigma E(t) - \gamma I(t) \\ dR(t) = (1 - \alpha)\gamma I(t) \\ dD(t) = \alpha\gamma I(t) \end{cases} \quad (1)$$

Below are the meaning of Latin alphabet parameters:

Parameter	Description
$\beta$	Rate at which infected people interact with others. The parameter controlling how often a susceptible-infected contact results in a new exposure. Often written as $R_0\gamma$ , where $R_0$ is the basic reproduction number
$\gamma$	The inverse of the time to recovery. The rate an infected recovers and moves into the resistant phase.
$\sigma$	The rate at which an exposed person becomes infective.
$\alpha$	The case fatality ratio.
$t$	Independent variable representing days here that controls how long the model will run.

It can be seen from the above introduction that SEIRD model is essentially a system of partial differential equations with the solution consisting of five classes of time series. As long as the initial parameters are given, the change trend of the population in different states in the total population over a period of time can be obtained. By comparing the estimated number of death cases with the actual number of death cases and making optimization adjustments, the required epidemiological parameters can be calculated.

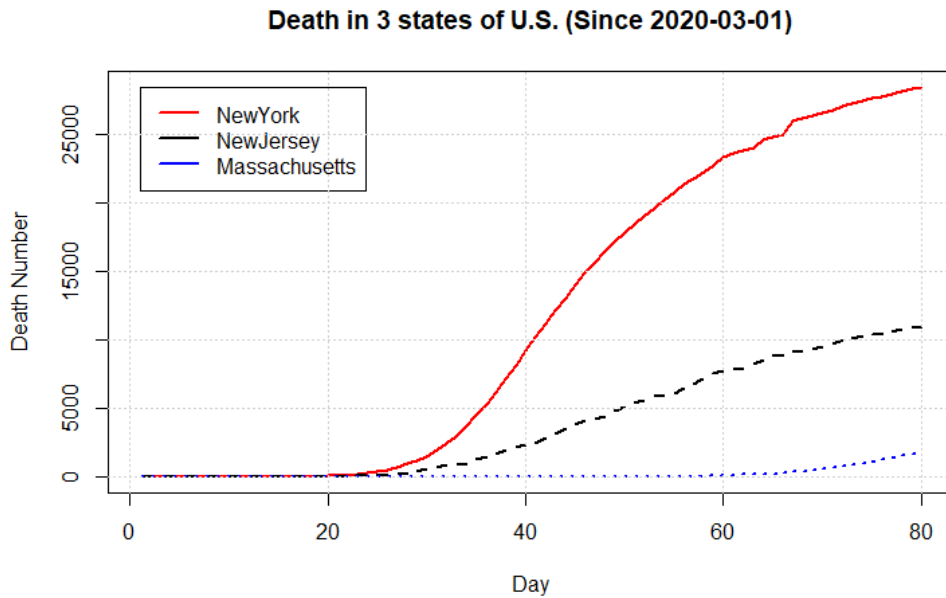
### 2.2 Data Acquisition and Common description

The raw downloaded from [3] [4] contains detailed information in many levels and it is wise to extract the data needed by the model into time-series. A few lines of raw data are:

date	state	cases	deaths
2020/4/8	Minnesota	1154	39
2020/4/8	Mississippi	2003	67
2020/4/8	Missouri	3327	86
2020/4/8	Montana	332	6
2020/4/8	Nebraska	538	14

After data cleaning, in each state a time series of the cumulative number of deaths is obtained. The population in state levels are derived from the summation of population in county levels from the raw data.

Below is the plot of death cases of some states in U.S.



## 2.3 Parameter Estimation

### 2.3.1 Country Level

To illustrate the procedure of parameter estimation, we choose the level of total death cases in U.S and the optimization goal is:

$$(\hat{R}_0, \hat{\alpha}) = \arg \min \left[ \sum_i \left( D(i) - \hat{D}(i) \right)^2 \right] \quad (2)$$

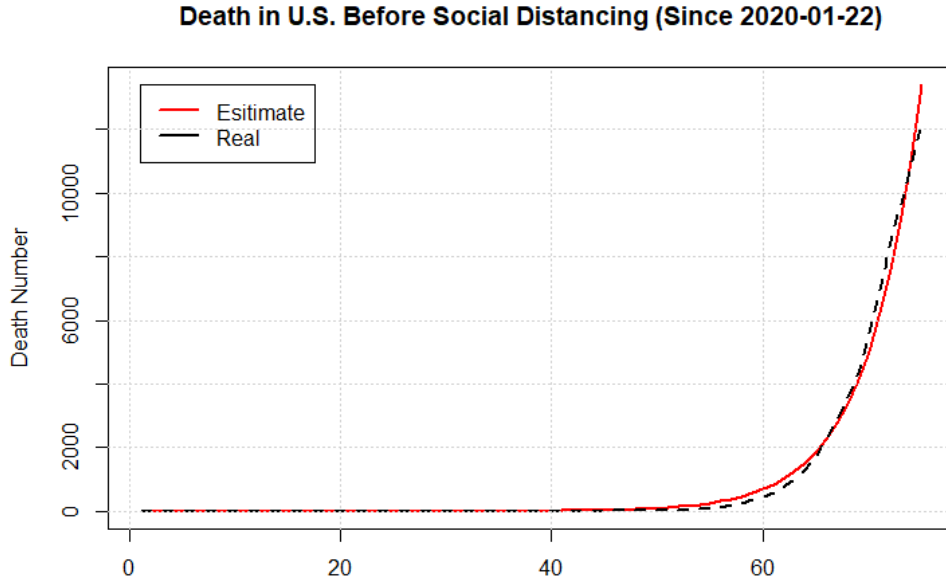
1. Takes random  $\alpha$  and  $R_0$  as inputs and simulates the SEIRD model. The output should include 5 time series:  $S(t)$ ,  $E(t)$ ,  $I(t)$ ,  $R(t)$ ,  $D(t)$ .
2. Compute the RSS (sum of squared differences between the number of deaths in the model  $D(t, R_0, \alpha)$  and in the data  $D(t)$ ) for the fixed values of  $R_0$  and  $\alpha$  and for time series data on deaths.
3. Find the parameters that minimize the RSS.

To simplify the calculation process in this case, let  $\gamma = 1/10$ ,  $\sigma = 1/4$  and the initial people distribution is:

$$\begin{aligned} E(0) &= 1, I(0) = D(0) = R(0) = 0 \\ S(0) &= N - I(0) - E(0) - R(0) - D(0) \end{aligned} \quad (3)$$

It turns out that  $R_0 = 1, \alpha = 0$ , apparently wrong answer based on the actual situation. A powerful explanation is that in the actual development of epidemiology, various policy measurements including limiting social distance have changed these parameters. If we divide the time series of death cases into two stages, where the first stage considering there is no social distance limitation, and the second part is the opposite. In this case, let stage-1 be the former 70 days of the whole time series, then the parameters turn out to be  $R_0 = 5.321, \alpha = 0.036$ , consistent with the result of medical reports[5].

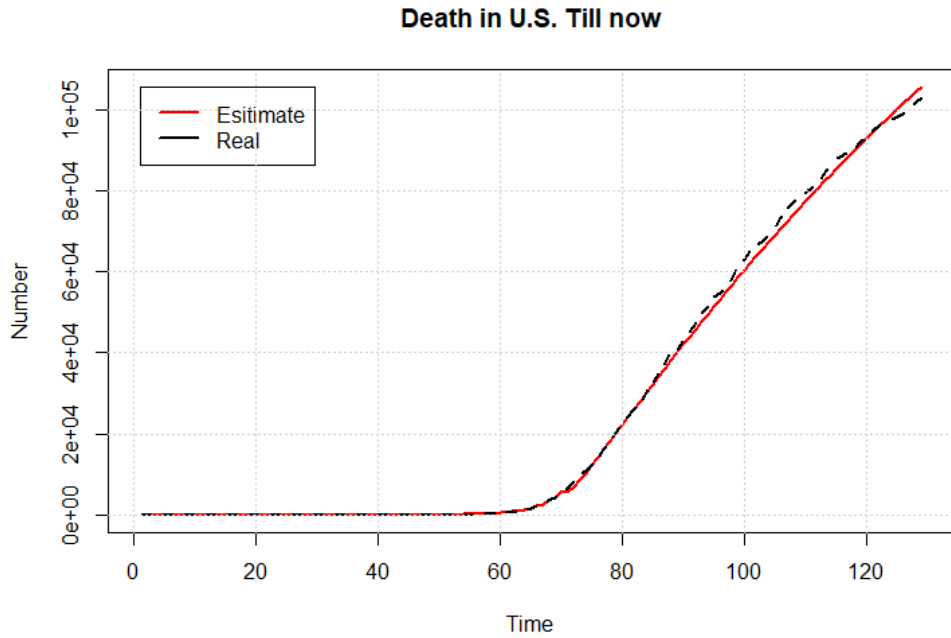
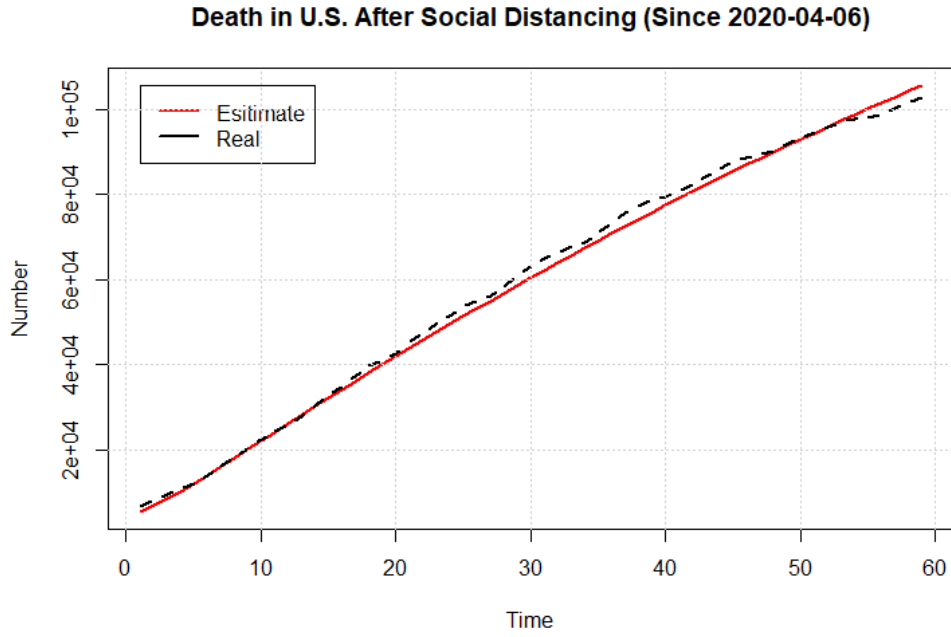
Moreover, we take these parameters into SERID model and compare the output of death cases time series with the actual data and plot them.



The tiny difference between estimation and reality indicates a satisfactory modeling result and we could in some way certain that these two epidemic parameters have certain practical reference significance.

Then for the stage-2, when the related policy considered to start work, the SERID model is simulated into the future future with changing  $R_0$ . As we set the resulting values  $S(T_0), E(T_0), I(T_0), R(T_0), D(T_0)$  as initial values from stage-1's model into the new model and simulate it with different parameter values. The reason for only changing  $R_0$  is that in this epidemic, the mortality rate is mainly related to the medical level, and the basic reproduction number is mainly related to the social distance, so in the case of U.S. whose medical system has not collapsed we should mainly consider the change of  $R_0$ .

Below are some simulated plots of stage-2 with changing  $R_0$  (stage-2 and the whole periods are given):



In the stage-2, the  $R_0$  is only 0.145 times that of stage-1(5.321), saying that a series of measures such as social isolation have indeed played a role in limiting the spread of the virus.

### 2.3.2 State Level

Next consider New York State, where the epidemic is more than severe. When used the same parameters as the previous example for simulation, we found a large deviation in the model simulation. After the analysis, the most likely reason is that the start date of New York State's data is March 1, 2020, when the epidemic has developed for a while, so the values of  $E(0)$  and  $I(0)$  need to be adjusted accordingly. The opination goal becomes into:

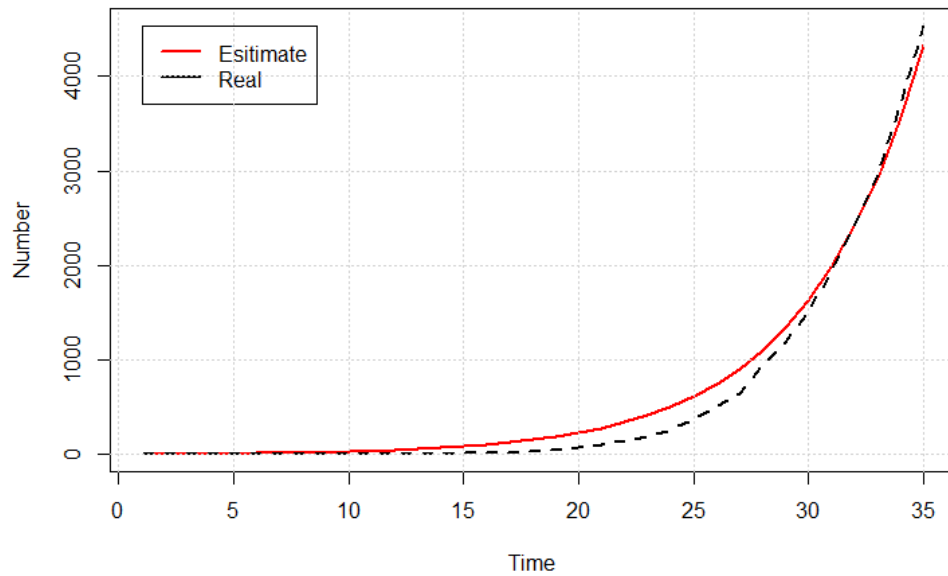
$$(\hat{R}_0, \hat{\alpha}, \hat{E}(0), \hat{I}(0)) = \arg \min \left[ \sum_i \left( D(i) - \hat{D}(i) \right)^2 \right] \quad (4)$$

After adjusting the parameter optimization method, the results obtained are as follows:

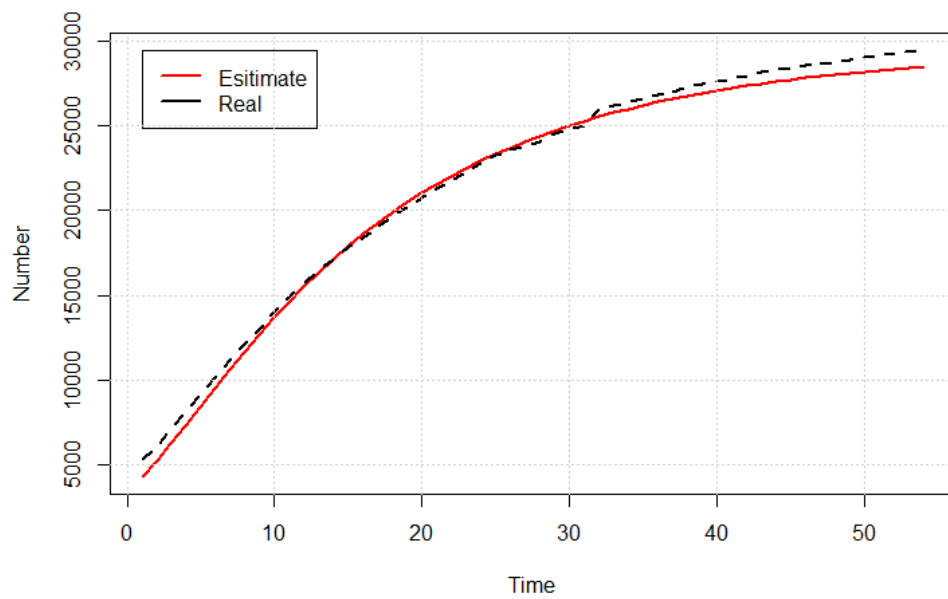
Divide Boundary=35

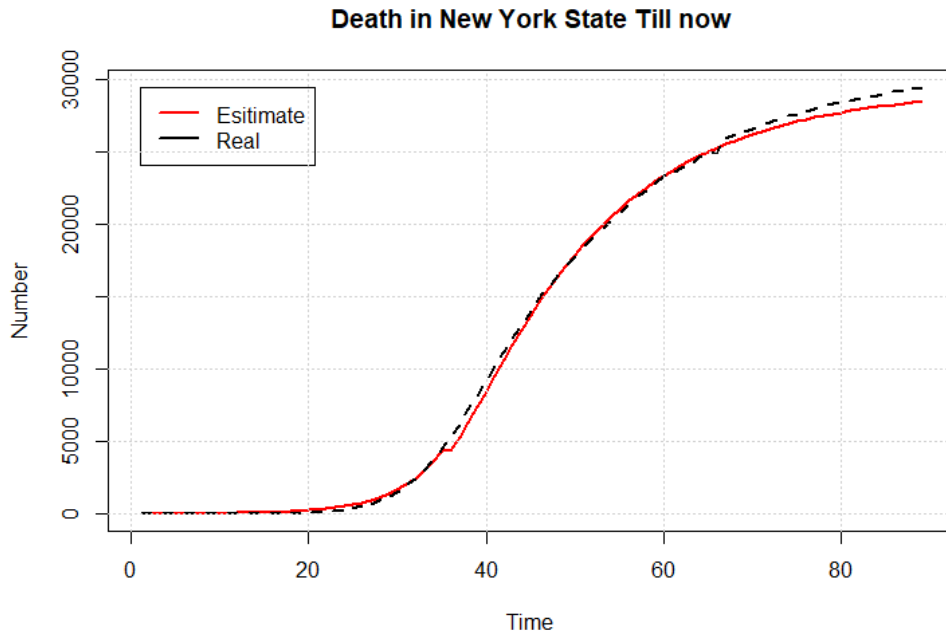
Stage	R_0	alpha	Initial Parameters
1	5.263	0.030	E = 1000, I = 50
2	0.05*R_0	0.030	Based on stage-1

**Death in New York State Before Social Distancing (Since 2020-03-01)**



**Death in New York State After Social Distancing (Since 2020-04-05)**



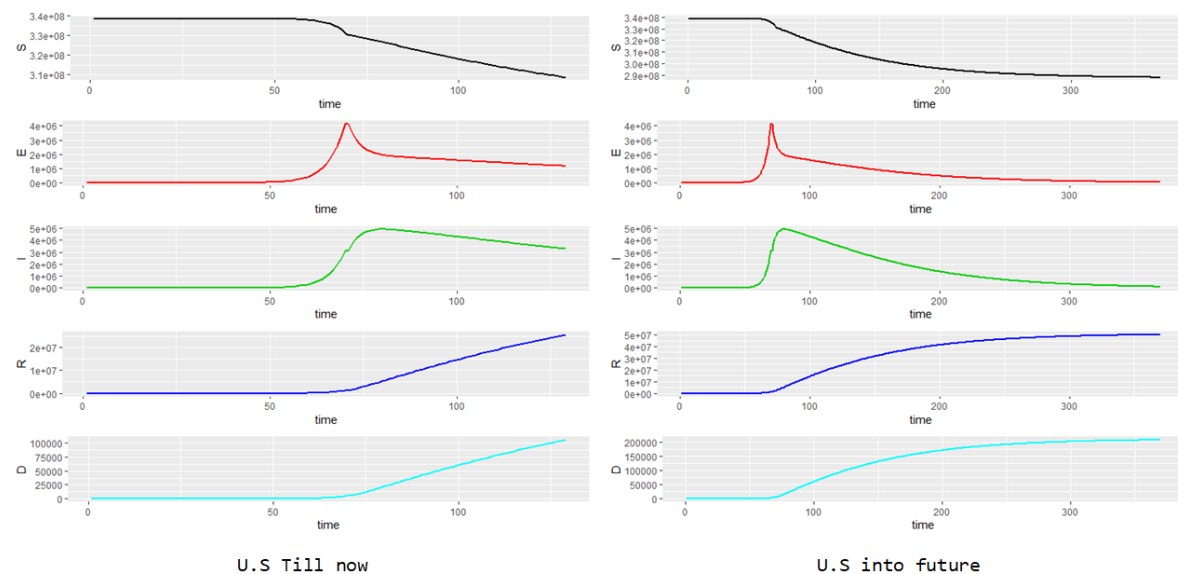


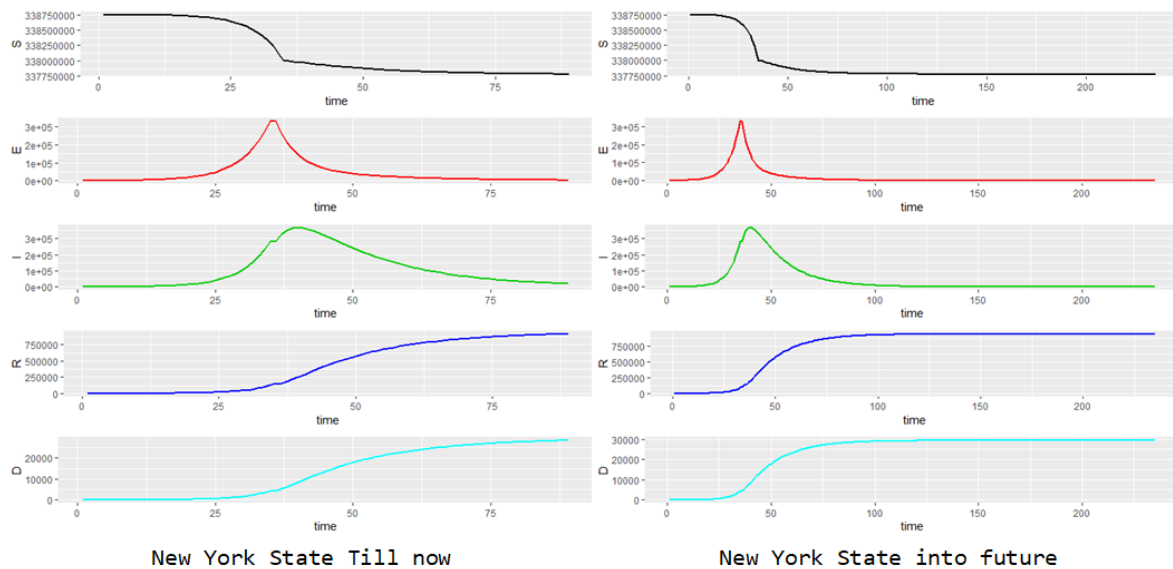
### 3. Discussions

As the main purpose of this report is to simulate the SEIRD epidemic model and estimate its parameters from the existing data, observe the similarity between simulated number and actual number in the plots and it is confident to say that the parameters we obtain reflect real ones. According to the above, based on the existing death data, the stage differentiation and the appropriate adjustment of basic reproduction number can obtain the epidemiological parameters with good verification results.

The first conclusion is that social distance, preventive measure adopted by various countries, does limit the value of basic reproduction number with a large extent, but there is a certain lag in its effective period.

Moreover, based on the existing parameters, we can estimate more about the trend of population change in different epidemiological categories of a country or region. By expanding the model from the number of death cases to the number of five population types in the SEIRD model and the plot is as below (also predict the model into future):





Unfortunately, judging from the predicted results, the number of death cases in the United States and New York state has not its top, even though  $R_0$  has fallen a lot under control measures. Theoretically speaking, if the current level of epidemic prevention is maintained, the number of deaths in the United States will eventually double to 200,000, which will be a humanitarian disaster. Therefore, it is suggested that more effective administrative and medical measures be implemented to save more people's lives.

In this report, we apply the SEIRD model to estimate the epidemiological parameters of COVID-19, compare with the real data and perform some predictions about the future development of death cases. In terms of historical data fitting, the model shows satisfactory accuracy, but it is essential for the segmentation of the epidemic development stage. In terms of future data prediction, the results of this model could somehow be referred as an estimate of future development based on current conditions.

For future work, possible directions include: whether  $R_0$  is a function of time (involving knowledge of virology), whether there are more efficiently investigation methods to detect the values of E and I rather than these estimated by optimization methods, and how to obtain more convincing predictions of future epidemic development.

## 4. Reference

- [1] COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)". ArcGIS. Johns Hopkins University. Retrieved 30 May 2020.
- [2] Paul Diaz, Paul Constantine, Kelsey Kalmbach, Eric Jones, Stephen Pankavich, A modified SEIR model for the spread of Ebola in Western Africa and metrics for resource allocation, Applied Mathematics and Computation, Volume 324, 2018, Pages 141-155, ISSN 0096-3003
- [3] Data from The New York Times, based on reports from state and local health agencies.<https://github.com/nytimes/covid-19-data>
- [4] CSSEGISandData [https://github.com/CSSEGISandData/COVID-19/tree/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series](https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series)
- [5] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. (January 2020). "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia". The New England Journal of Medicine. 382 (13): 1199–1207. doi:10.1056/NEJMoa2001316. PMC 7121484. PMID 31995857.