

READING AND WRITING MATHEMATICS

SIMULATION OF SEIR INFECTIOUS DISEASE MODEL FOR COVID-19 BREAKOUT IN CHINA

Yiran Jing

Feb, 2020

Abstract

Background: The deterministic SEIR (susceptible-exposed-infectious-recovered) model is one of mathematical models of infectious disease. The global Novel Coronavirus outbreak (named COVID-19) that was first detected in Wuhan, China is spreading quickly. My aim is to study and discuss the properties of the SEIR model and how to use it to forecast COVID-19 breakout in China based on the daily observation.

Result: Using the data between 2019-12-08 and 2020-02-13 to train dynamic SEIR model, model prediction for the next 50 days is shown in figure 5, in which the COVID-19 breakout will be close to an end by April (the estimated number of infected case will be less than 4000 in early April). By comparing the observation and predictions from 2020-02-14 to 2020-02-22 (figure 6), SEIR model predicts well for the peaking time and the general trend.

Code resource: on my github **model 3:** <https://github.com/YiranJing/Coronavirus-Epidemic-2019-nCov>.

Contents

Abstract

1	Overview of SEIR model for infectious disease	1
2	Formulation of the basic SEIR model	2
2.1	Understanding SEIR ordinary differential equations	2
2.2	Assumptions of SEIR model	3
3	Example of simulating COVID-19 breakout in China	3
3.1	Data source and COVID-19 statistics for SEIR model	3
3.2	Assumptions in COVID-19 simulation	3
3.2.1	SEIR model assumptions	3
3.2.2	Other assumptions	4
3.3	Simulate contact rate β as a function of time t	4
3.4	SEIR Prediction	6
3.4.1	Model performance	6
3.5	Challenges of using SEIR model in COVID-19 breakout simulation	7
4	Reference List	7

1 Overview of SEIR model for infectious disease

In this section, we briefly discuss the properties of the basic Susceptible-Exposed-Infected-Removed (SEIR) system that will be used to describe the recent outbreak of COVID-19 in China.

We considered a simple Susceptible-Exposed-Infected-Removed (SEIR) epidemic model for the simulation of the infectious-disease spread. Individuals were each assigned to one of the following disease states: **Susceptible (S)**: segment not yet infected, disease-free; **Exposed (E)**: individuals are experiencing incubation duration; **Infectious (I)**: the confirmed (isolated) cases, or **Removed (R)**: Recovered individuals.

The SEIR diagram Figure 1 below shows how individuals move through each compartment in the model.

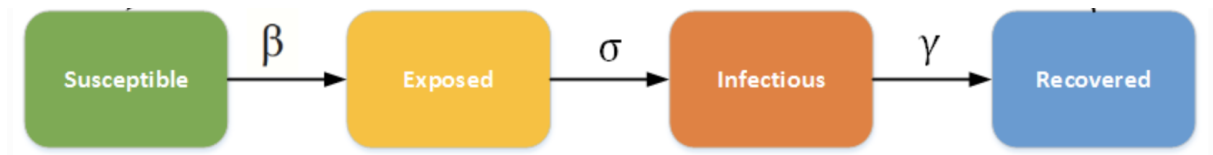


Figure 1: SEIR model with 4 states

Parameters within this model are:

1. **Contact rate** β , controls the rate of spread, which represents the probability of transmitting disease between a susceptible and an infectious individual.
2. **Incubation rate** σ , is the rate of latent individuals becoming infectious. Given the known *average duration of incubation* Y , $\sigma = 1/Y$.
3. **Recovery rate** $\gamma = 1/D$, is determined by *the average duration of recovery* D , of infection. After this period, they enter the removed phase.

Figure 2 shows the diagrammatic representation of virus progress in an individual, where infectious occurs at t_L , during the latent period, infected individual is not infectious, and at t_{sy} , symptoms appear. The first transmission to left healthy individuals is at t_{tr} . After t_R , the removed(recovered) people are no longer infectious.

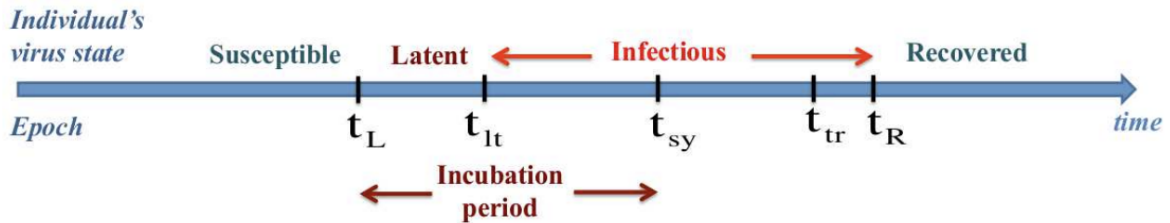


Figure 2: Virus progress in an individual by using the SEIR model

2 Formulation of the basic SEIR model

The transmission of the virus is then described by the following system of nonlinear ordinary differential equations:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \quad (1)$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E \quad (2)$$

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

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

where $S + E + I + R = N$ is the total population, and contact rate $\beta > 0$, incubation rate $\sigma > 0$, and recovery rate $\gamma > 0$. The $\frac{dS}{dt}$ represents the rate of change of S with respect to time t. Same as $\frac{dE}{dt}$, $\frac{dI}{dt}$ and $\frac{dR}{dt}$.

2.1 Understanding SEIR ordinary differential equations

The contact rate β is the rate of infection from an infected individual to one of their susceptible contacts on the unitary time step dt . For example, given two people A (infectious) and B (Susceptible), the probability of B becoming infected after contacting A during the unitary time step is β .

The term Δt is the difference between two observation points. Thus, The number of individuals transferred from Susceptible state to Exposed state is $\frac{\beta SI}{N} \Delta t$. $\beta * S/N$ is the force of infection (an official term in epidemiology) in the SEIR model which is the number of new infections divided by population N.

Similarly, on the unitary time step, there are $\sigma E \Delta t$ number of cases transferred from Exposed state to Infectious, and $\gamma I(t) \Delta t$ number of cases transferred from Infectious to Removed.

Let $S(t)$, $E(t)$, $I(t)$ and $R(t)$ be the number of susceptible, exposed, infectious and removed individuals at time t, then

$$S(t + \Delta t) = S(t) - \frac{\beta S(t) I(t)}{N} \Delta t \quad (5)$$

$$E(t + \Delta t) = E(t) + \frac{\beta S(t) I(t)}{N} \Delta t - \sigma E(t) \Delta t \quad (6)$$

$$I(t + \Delta t) = I(t) + \sigma E(t) \Delta t - \gamma I(t) \Delta t \quad (7)$$

$$R(t + \Delta t) = R(t) + \gamma I(t) \Delta t \quad (8)$$

Based the definition of first order derivative, $\frac{dX}{dt} = \frac{X(t+\Delta t) - X(t)}{\Delta t}$, as $\Delta t \rightarrow 0+$. Thus formula (5)-(8) can be rewritten as formula (1)-(4).

2.2 Assumptions of SEIR model

1. The SEIR model assumes a closed population, which means that the total number of population is fixed, no births, no death, or introduction of new individuals. From formula (1)-(4), we see that $\frac{d}{dt}[S(t) + E(t) + I(t) + R(t)] = 0$, that is, the population N is constant in any time t : $S(t) + E(t) + I(t) + R(t) = N$ for any $t \geq 0$.
2. The individuals in exposed state is is infected but not yet infectious.
3. Well-mixed population.
4. SEIR model assumes that the latent and infectious times of the pathogen are exponentially distributed.

3 Example of simulating COVID-19 breakout in China

3.1 Data source and COVID-19 statistics for SEIR model

The daily observation before 2020-02-07 are collected from the published paper and the official report of Central Commission for Discipline Inspection of the Communist Party of China (CCDI). The data after 2020-02-07 are real-time query from DingXiangYuan, authorized by Chinese government.

According to the information provided by World Health Organization (WHO), the mean time of COVID-19 incubation period is 7 days, thus Incubation rate $\sigma = 1/7$. Infectious(I) is the net confirmed cases (total confirmed case - heal - died). And the China population is 1.4 billion.

Training data: the observations between 2019-12-08 and 2020-02-13. **Test data:** observation from 2020-02-14 to 2020-02-22.

3.2 Assumptions in COVID-19 simulation

Let's discussion the SEIR model assumptions in COVID-19 case

3.2.1 SEIR model assumptions

1. Constant (closed) population size: Due to the international travel ban, strict home quarantine rules in China and the low death rate of COVID-19 (less than 2%), we can assume the China population is constant.
2. In SEIR models, the exposed individuals is infected but not yet infectious, and the first transmission can only happen after symptoms appear (see figure 2). However, In COVID-19 case, we know that individuals are infectious during the whole incubation period.
3. The well-mixed population of SEIR model is not suitable here, because most people in China are forced to stay at home.

3.2.2 Other assumptions

1. Suppose the average duration of recovery is 14 days, which is similar with SARS, and thus Recovery rate $\gamma = 1/14$.
2. Suppose the total number of individuals within incubation period is 4 time of susceptible case reported by CCDI.
3. Assume the dead people on proposition of diseased is around 2%, belonging to the removed individuals (R).
4. Assume latent period is the same as incubation.

3.3 Simulate contact rate β as a function of time t

In the traditional SEIR model, the infectious rate, β the per capita rate of infection given contact) is constant, while in COVID-19 case, the contact rate is changing over time: increase in the early stage due to the unawareness of the public and virus variation, then decrease under the strict control policy. So we suppose β follows the slope of logistic function, by which we will get use an estimate of β per day, which would be used in the dynamic SEIR model.:

$$\beta(t) = \frac{Ce^{-\alpha(t+b)}}{(1 + e^{-\alpha(t+b)})^2} \quad (9)$$

where t is the number of days after 2019-12-08 (time point the first infected case found). The α term is a regularization parameter, C is a scaling constant, and b is bias.

A Loss Functions tells us how good our model is at making predictions for a given set of parameters (C , α and b in our case). To get the optimal β over time, we define mean square loss function (10) in this case, and parameters C , α and b were estimated by solving:

$$\begin{aligned} \text{minimize} \quad & L = \frac{1}{n} \sum_{j=1}^n (I_{real,j} - I_{siml,j})^2 \\ \text{subject to} \quad & \text{equations (1)-(4) of SEIR model} \end{aligned} \quad (10)$$

The goodness of fit is measured by computing the value of the objective function Loss(L) of (10). where $I_{real,j}$ corresponds to real observation of net confirmed cases provided by CCDI and $I_{siml,j}$ is obtained by SEIR model(equations (1)-(4)). We use Gradient Descent, an optimization algorithm, to update the parameters of C , α and b . The figure 3 shows the fitted value result using the optimal C , α and b .

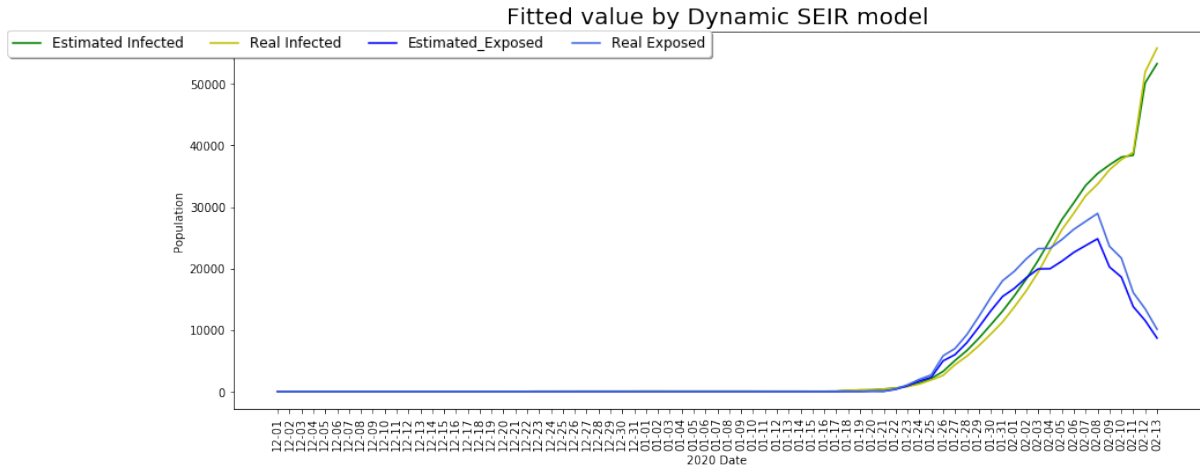


Figure 3: Fitted value by Gradient Descent

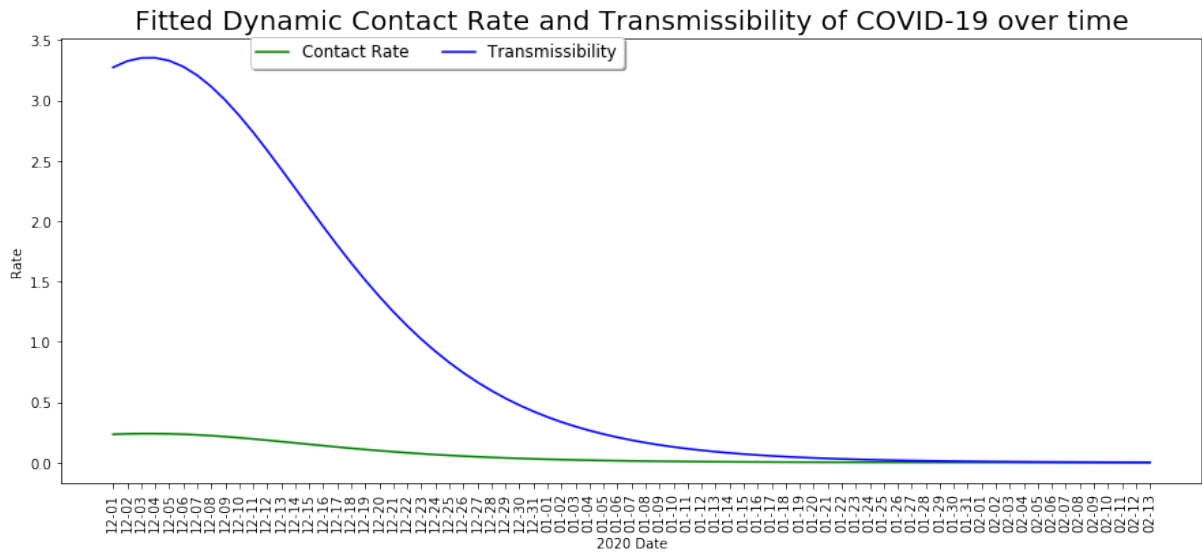


Figure 4: Fitted Dynamic Contact Rate and Transmissibility of COVID-19 over time

The transmissibility is β/γ , which measure the transmission ability over time. In figure 4, the first point of blue line above is R_0 (the basic reproduction number) of COVID-19 (around 3.2). (*The R_0 (R naught) represents how many people an infected individual will transmit the virus to others.*)

3.4 SEIR Prediction

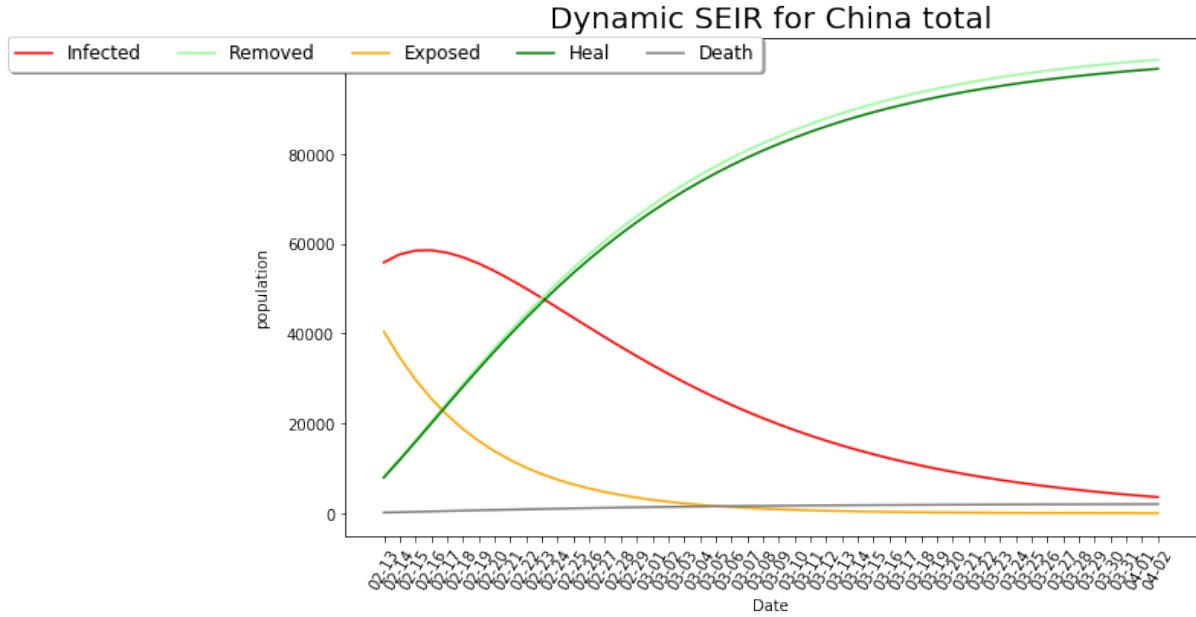


Figure 5: The prediction for next 50 days after 13 Feb 2020

The Figure 5 above shows SEIR prediction in the next 50 days since 14 Feb. Main conclusions are:

1. The number of net confirmed cases will be above 58500, and the peak can be reach before 20 Feb (16 Feb according to red line in the plot).
2. The transmissibility has been controlled and decrease from initial 3.2(R_0) to less than 0.5.
3. The COVID-19 breakout is expected to close to an end in early April. That is, the estimated number of infected case will be less than 4000 in early April, and continue to decrease.

3.4.1 Model performance

The mean absolute percentage error (MAPE) is a measure of prediction accuracy of a forecasting method in statistics. The MAPE of confirmed cases using data between 2020-2-14 to 2020-02-22 is 0.0066.

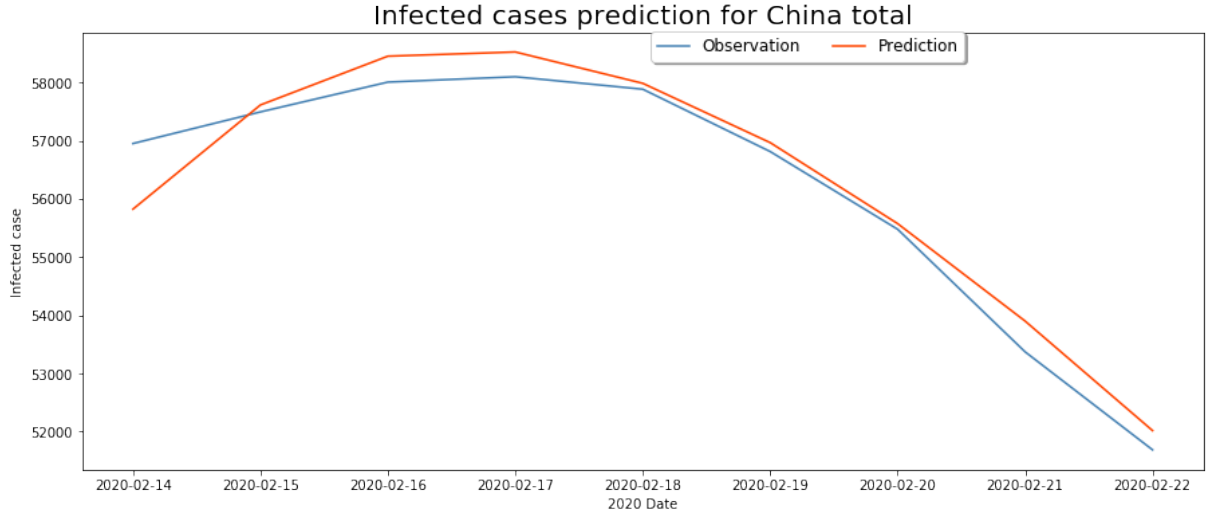


Figure 6: Real observations and predictions from 2020-02-14 to 2020-02-22

The Figure 6 above visualizes the real observation and the SEIR model predictions for the next 7 days. Overall, SEIR model predicts well for the peaking time and the general trend.

3.5 Challenges of using SEIR model in COVID-19 breakout simulation

Firstly, We have limited understanding of this new disease. For example, we did not test all people with 2019-nCoV correctly. (unclear symptoms: whether people with 2019-nCoV who do not have symptoms can transmit an infection). Secondly, It is hard to get the real-time correct information. (official Chinese data can be under-reported), especially for Wuhan. Thirdly, the prediction is highly sensitive to the policy (travel restriction, force people stay in home, Wuhan build 3 new hospital for 2019-nCoV etc.), all of these policies influenced a lot on the time-line. When we do prediction, the key assumption is no new policy in the future.

4 Reference List