- **Introduction**

In December 2019, a novel coronavirus (2019-nCoV) is thought to have emerged into the human population in Wuhan, China. The number of identified cases in Wuhan has increased. With the increasing incidence of confirmed cases, corresponding spread control policies and emergency actions are taking place. The symptom onset date of the first 2019-nCoV patient was identified in early December 2019 and the outbreak started in late December with most of cases epidemiologically connected to a seafood market in the city of Wuhan, Hubei province. Following the cases reported in other Chinese cities and overseas, the National Health Commission (NHC) of People's Republic of China confirmed the evidence of human-to-human transmission of such viral pneumonia. The first known human infection occurred in December 8, 2019. An outbreak of 2019-nCoV was first detected in Wuhan, China, in mid-December 2019.The virus subsequently spread to all other provinces of China and to more than twenty other countries in Asia, Europe, North America, and Oceania. Human-to-human spread of the virus has been confirmed in China, Germany, Thailand, Taiwan, Japan, and the United States. As of 28th January 2020, authorities had reported more than 4,515 confirmed cases across thirteen provinces in mainland China and confirmed 106 deaths. First laboratory-confirmed case was also reported in Taiwan on 21st January, and on the following day, both Hong Kong and Macao reported their first confirmed cases. Globally, Thailand, Japan, South Korea, Vietnam, Singapore, Malaysia, Nepal, France, Canada, Australia and the United States had reported cases infected by 2019-nCoV. All confirmed cases so far are travelers from or ever been to Wuhan or other Chinese cities. As of 1 February 2020, there were 12,024 confirmed cases of infection, of which 11,860 were within mainland China. Cases outside China, to date, were people who have either traveled from Wuhan, or were in direct contact with someone who traveled from the area. The number of deaths was 259 as of 1 February 2020. To combat the 2019-nCoV outbreak, authorities in China have implemented several preventive measures. Starting from 10am, 23rd January, all public transport has been suspended following by the lock down on the city of Wuhan. Neighboring cities also announced a lock down in sequence. Residents were advised to remain at home and avoid gathering in order to contain the virus spread. Following the raise of protection standards instructed by NHC, prevention and control measures, such as disinfection for public facilities, have been strengthened and taking places in other cities. For example, Hong Kong, a special administrative region of China, has undergone temperature screening of passengers at broader entries following the Prevention and Control of Disease Ordinance and the International Health Regulation.

Current clinical and epidemiological data are insufficient to understand the full extent of the transmission potential of the epidemic. This comes at a time when there is a substantial increase in travel volume within as well as in and out of China around the Lunar New Year on 25 January 2019. Over 3 billion passenger journeys were predicted for the period between 10 January and 18 February (CGTN 2020), which could significantly increase the spread of the virus.

In order to help people better assess the situation, and to promote rational reactions, the epidemic data must be easily accessible to the general public. Many websites have been publishing some epidemic numbers in real time. In this report I use Johns Hopkins CSSE data set. These data set provide timely information to the general public, but they fall short of providing enough data for analysis. For example, answering any of the following simple questions will be very difficult or nearly impossible by using the websites mentioned above.

Mathematical epidemiology, the mathematical modeling of disease spread in a population, has a long history and many early developments in the mathematical modeling of diseases are due to public health physicians. Mathematical epidemiology can aid medical professionals who are trying to manage diseases. During a disease outbreak ethical consideration make it impossible for medical experiments

to be performed on a population which would compare the effects of different disease management strategies. Therefore, predictions based upon mathematical models may be essential in addressing the impact of communicable diseases and formulating strategies for fighting them.

- **The model**
  Model formulation :

The disease model we are going to investigate is constructed within a static population, i.e., a population in which it is assumed that the population does not change from travel or births or deaths except possibly from deaths of people who are infected with the disease. In our model we are going to assume that the epidemic process is deterministic, that is, that population behavior is completely determined by its history and by the rules that describe the model. . Our population is divided into four sets (compartments) of people: those who are susceptible to disease infection but not yet infected; those who have been exposed,i.e., infected but not yet infectious; those who are infectious; and those who have been removed from being susceptible or infected by recovering and acquiring immunity to the disease. Therefore,the population is made up of N people and $N = S + E + I + R$ R, where we define

$$S = S(t) = \text{number of individuals in the population susceptible at time } t,$$
$$E = E(t) = \text{number of individuals exposed (but not infectious) at time } t,$$
$$I = I(t) = \text{number of individuals infectious at time } t, \text{and}$$
$$R = R(t) = \text{number of individuals recovered at time } t$$

The numbers in each compartment are integers.but if N is sufficiently large then $S(t), E(t), I(t)$ and $R(t)$ can be treated in our model as continuous variables.

Denote the disease transmission rate (per infectious person) by β. This model assumes mass action incidence, that an average member of the population makes contacts sufficient to transmit infection with βN others per unit time. The numbers of people who are susceptible, exposed, infectious, or removed at time t are denoted by $S(t), E(t), I(t)$ and $R(t)$,respectively,and the total population $N = S + E + I + R$.This SEIR model assumes constant rates of mass-action incidence and disease recovery, and mean periods of infectiousness and latency.Denote by ϵ the rate of transfer from exposure or latency to infectiousness;note that $\frac{1}{\epsilon}$ is then the mean period of exposure or latency.Denote by γ the recovery rate (so that $\frac{1}{\gamma}$ is the mean period of infectiousness).

If we assume that initially the entire population is susceptible, this is the same as writing $S(0) = N$.Given a susceptible individual, the probability that that person would have contact with an infectious individual is $\frac{1}{N}$ since $\frac{1}{N}$ N is the fraction of the population that is infectious; so the rate of new infectious people per susceptible is

$$(\beta N)\frac{1}{N}S = \beta IS \tag{1}$$

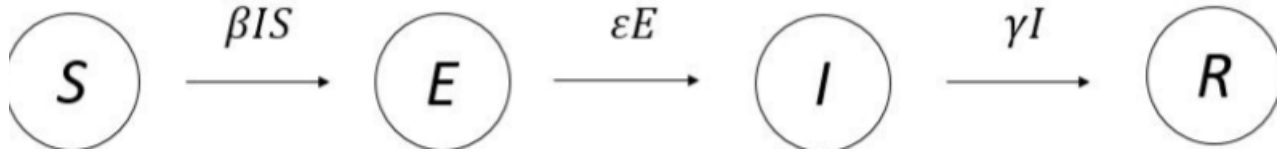The flowchart in Figure 1 shows the scheme of progression in the model from one compartment to the next.

Figure 1: Flow chart for the SEIR model

$$S' = -\beta IS \tag{2}$$
$$E' = \beta IS - \epsilon E \tag{3}$$
$$I' = \epsilon E - \gamma I \tag{4}$$
$$R' = \gamma I \tag{5}$$

with measurement equation $y = kI$. The model is determined by the system of ordinary differential equations (primes denote differentiation with respect to time t)
$S(t_0) = S_0 \geq 0$, $E(t_0) = E_0 \geq 0$, $I(t_0) = I_0 \geq 0$.
The term βIS corresponds to the process of infection of susceptible individuals. The term ϵE corresponds to the process of acquiring of infectivity by the exposed individuals. The term γI describes the recovery process of the infectives. We take y to indicate that we are measuring a proportion of the infected population (e.g. if not all cases are reported). k represents proportion of the infected population which is reported/observed. We consider the intensity coefficients β, ϵ, γ and k non-negative.

Model Analysis :
We assume mass action incidence which implies that an average member of the population makes contacts sufficient to transmit infection with βN others per unit time, so that in the mean infectiousness period $\frac{1}{\gamma}$ a newly introduced infectious person could theoretically infect a total of $\frac{\beta N}{\gamma}$ individuals. Since the duration of the epidemic process is relatively short, we assume the influence of birth and migration processes on the disease dynamics negligible and do not include these processes into the model. Further without loss of generality we assume $t_0 = 0$ Denote

$$R_0 = \frac{\beta N}{\gamma} \tag{6}$$

the basic reproduction number. If we add together the equations (3) and (4) we get

$$E' + I' = (E + I)' = \beta IS - \epsilon E + \epsilon E - \gamma I$$
$$= \beta IS - \gamma I$$

$$E' + I' = (\beta S - \gamma)I \tag{7}$$

$E(t) + I(t)$ can be increased only if $E' + I' > 0$. By equation (7) and because $I(t) > 0$, we conclude that $E' + I' > 0$ as long as
$$\beta S - \gamma > 0$$

or

$$\frac{\beta S}{\gamma} > 1 \tag{8}$$

With $S(0) = N$ this becomes

$$\frac{\beta N}{\gamma} > 1$$

or $R_o > 1$ by equation (6).Thus we have an epidemic if $R_0 > 1$ and if $R_0 < 1$ then $E(t) + I(t)$ decreases and the disease die out

It is desirable to be able to solve the system of equations (2)–(5) and obtain an expression for $I(t)$.An expression for $I(t)$ was found in the paper of Kermack and McKendrick on 1927,but it did not model the number of infectious well when t was not close to zero. An alternative is to add the equations for E' and I',then divide by the equation for S'.

$$\frac{d(E+I)}{dS} = \frac{\beta IS - \epsilon E + \epsilon E - \gamma I}{-\beta IS} = -1 + \frac{\gamma}{\beta S}$$

and multiplying through by dS gives

$$d(E+I) = -dS + \frac{\gamma dS}{\beta S} \tag{9}$$

We next integrate both sides of this equation from 0 to t . Then we have

$$E(t) - E(0) + I(t) - I(0) = -(S(t) - S(0)) + \frac{\gamma}{\beta}(\ln S(t) - \ln S(0))$$

and since $E(0) = 0$ and $I(0) \approxeq 0$ this is equivalent to

$$E(t) + I(t) = N - S(t) + \frac{\gamma}{\beta} \ln \frac{S(t)}{N} \tag{10}$$

Remark 1
Denote $\lim_{t\to\infty} S(t)$ by $S(\infty)$, similarly for $\lim_{t\to\infty} I(t)$ by $I(\infty)$ and $\lim_{t\to\infty} E(t)$ by $E(\infty)$.

Moreover, we can prove that the quantity $E(\infty) = 0$ and $I(\infty) = 0$.
proof:
We have already seen in equation (7) that $E + I$ is increasing when $S > \frac{\gamma}{\beta}$. By equation (7) when $S < \frac{\gamma}{\beta}$ then $E' + I' < 0$. Thus when $S = \frac{\gamma}{\beta}z, E' + I' = 0$ and by the first derivative test $E + I$ has a maximum value there $E + I$ is positive valued and thus is bounded below by zero, so $E(\infty) + I(\infty) = 0$ which implies $E(\infty) = 0$ and $I(\infty) = 0$

Remark 2
Let R_0 be defined as in equation (6) Then $ln\frac{S(\infty)}{N} = -R_0(1 - \frac{S(\infty)}{N})$

proof:
In equation (10) as $t \to \infty$:

$$E(\infty) + I(\infty) = N - S(\infty) + \frac{\gamma}{\beta} \ln \frac{S(\infty)}{N}$$

where the left side is zero by the remark (1), so that

$$-\frac{\gamma}{\beta} \ln \frac{S(\infty)}{N} = N - S(\infty)$$
$$= N\left(1 - \frac{S(\infty)}{N}\right)$$

Multiply both sides by $\frac{-\beta}{\gamma}$, then apply equation (6) to get

$$\ln \frac{S(\infty)}{N} = -\frac{\beta}{\gamma} N \left(1 - \frac{S(\infty)}{N}\right)$$
$$= -R_0 \left(1 - \frac{S(\infty)}{N}\right)$$

Then

$$\ln \frac{S(\infty)}{N} = -R_0 \left(1 - \frac{S(\infty)}{N}\right) \tag{11}$$

- **Outbreak incidence data**

The data set I used is provided on this website:
`https://www.kaggle.com/brendaso/2019-coronavirus-dataset-01212020-01262020`

structure of the data set is as in the next table :

Sno	Date	Province/State	Country	Last Update	Confirmed	Deaths	Recovered
1	2020-01-22 12:00:00	Anhui	China	2020-01-22 12:00:00	1.00000	0.00000	0.00000
2	2020-01-22 12:00:00	Beijing	China	2020-01-22 12:00:00	14.00000	0.00000	0.00000
3	2020-01-22 12:00:00	Chongqing	China	2020-01-22 12:00:00	6.00000	0.00000	0.00000
4	2020-01-22 12:00:00	Fujian	China	2020-01-22 12:00:00	1.00000	0.00000	0.00000
5	2020-01-22 12:00:00	Gansu	China	2020-01-22 12:00:00	0.00000	0.00000	0.00000
6	2020-01-22 12:00:00	Guangdong	China	2020-01-22 12:00:00	26.00000	0.00000	0.00000
7	2020-01-22 12:00:00	Guangxi	China	2020-01-22 12:00:00	2.00000	0.00000	0.00000
8	2020-01-22 12:00:00	Guizhou	China	2020-01-22 12:00:00	1.00000	0.00000	0.00000
9	2020-01-22 12:00:00	Hainan	China	2020-01-22 12:00:00	4.00000	0.00000	0.00000
10	2020-01-22 12:00:00	Hebei	China	2020-01-22 12:00:00	1.00000	0.00000	0.00000
11	2020-01-22 12:00:00	Heilongjiang	China	2020-01-22 12:00:00	0.00000	0.00000	0.00000
12	2020-01-22 12:00:00	Henan	China	2020-01-22 12:00:00	5.00000	0.00000	0.00000
13	2020-01-22 12:00:00	Hong Kong	China	2020-01-22 12:00:00	0.00000	0.00000	0.00000
14	2020-01-22 12:00:00	Hubei	China	2020-01-22 12:00:00	444.00000	0.00000	0.00000
15	2020-01-22 12:00:00	Hunan	China	2020-01-22 12:00:00	4.00000	0.00000	0.00000
16	2020-01-22 12:00:00	Inner Mongolia	China	2020-01-22 12:00:00	0.00000	0.00000	0.00000
17	2020-01-22 12:00:00	Jiangsu	China	2020-01-22 12:00:00	1.00000	0.00000	0.00000
18	2020-01-22 12:00:00	Jiangxi	China	2020-01-22 12:00:00	2.00000	0.00000	0.00000

Figure 2: Structure of the csv file for 2019-nCorv

EDA Exploratory Data Analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. EDA is used for seeing what the data can tell us before the modeling task. It is not easy to look at a column of numbers or a whole spreadsheet and determine important characteristics of the data. It may be tedious, boring, and/or overwhelming to derive insights by looking at plain numbers. Exploratory data analysis techniques have been devised as an aid in this situation.
At the beginning ,we apply pre processing on the data above. The first I am going to clean up the data to since in some records it is recorded as China and in other rows as Mainland China , Just make organization of the data. Next, I removed some of unnecessary columns to me in my study of the data such as "Sno","Last Update" and add new column for the country name adjustment which I did.
Moreover, in the data set, a place may have reported data more than once per day. For more effective analysis, we convert the data into daily. If the data for the latest day is not available, we will fill it with previous available data.
After preprocessing the data will be structured as in figure 3

Date	Province/State	Confirmed	Deaths	Recovered	Country
2020-01-22 12:00:00	Anhui	1.00000	0.00000	0.00000	Mainland China
2020-01-22 12:00:00	Beijing	14.00000	0.00000	0.00000	Mainland China
2020-01-22 12:00:00	Chongqing	6.00000	0.00000	0.00000	Mainland China
2020-01-22 12:00:00	Fujian	1.00000	0.00000	0.00000	Mainland China
2020-01-22 12:00:00	Gansu	0.00000	0.00000	0.00000	Mainland China

Figure 3: Structure of the data frame file after data preprocessing

Regarding to the confirmed cases
At the moment of writing, a vast majority of cases is in Hubei Province in China, and overseas cases

are just at the beginning. Therefore, in the coming analysis, we separate the cases into three regions, 'Hubei PRC', 'Others PRC' and 'World'on daily segment. In table we can see the number of confirmed cases of each of these three regions on the last 10 days

	Hubei PRC	Others PRC	World
2020-02-08T00:00:00.000000000	27100.00000	10098.00000	373.00000
2020-02-09T00:00:00.000000000	29631.00000	10529.00000	459.00000
2020-02-10T00:00:00.000000000	31728.00000	10582.00000	540.00000
2020-02-11T00:00:00.000000000	33366.00000	11275.00000	559.00000
2020-02-12T00:00:00.000000000	48206.00000	11599.00000	606.00000
2020-02-13T00:00:00.000000000	51986.00000	11855.00000	664.00000
2020-02-14T00:00:00.000000000	54406.00000	11886.00000	678.00000
2020-02-15T00:00:00.000000000	56249.00000	12098.00000	768.00000
2020-02-16T00:00:00.000000000	58182.00000	12264.00000	863.00000
2020-02-17T00:00:00.000000000	59989.00000	12375.00000	979.00000

Table1: Confirmed cases in the last 10 days

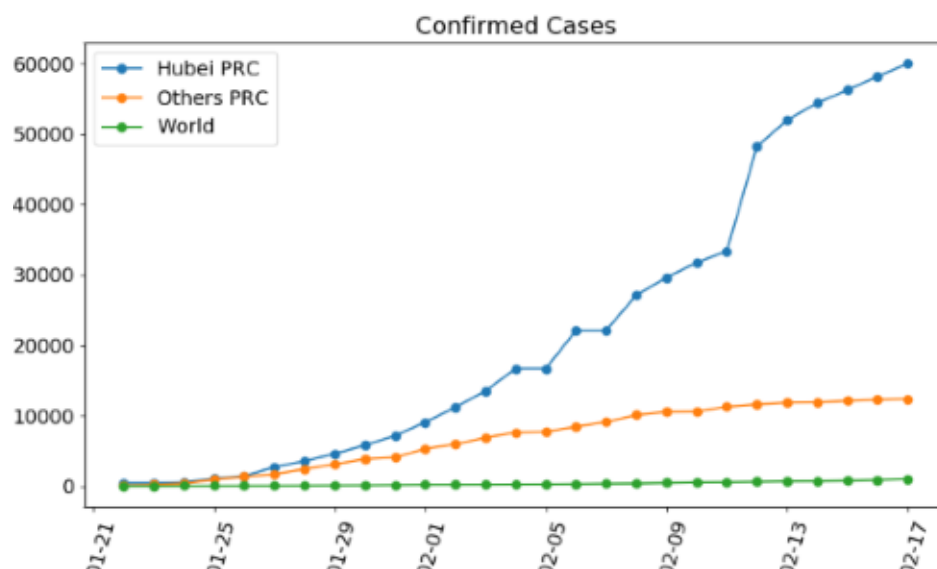In the next figure , we have a plot of growth of the confirmed cases over time :

Figure 4: Growth of confirmed cases for the three regions

I can observe that :

• Hubei province has more confirmed cases than other provinces in China combined. Due to change in

criteria of 'Confirmed' case, Hubei saw a huge increase of $\sim 15,000$ cases on Feb 12.

• Cases outside China are much less.

Regarding to the death cases we can see the number of cases on the last 10 days as

	Hubei PRC	Others PRC	World
2020-02-08T00:00:00.000000000	780.00000	31.00000	2.00000
2020-02-09T00:00:00.000000000	871.00000	37.00000	2.00000
2020-02-10T00:00:00.000000000	974.00000	37.00000	2.00000
2020-02-11T00:00:00.000000000	1068.00000	45.00000	2.00000
2020-02-12T00:00:00.000000000	1310.00000	56.00000	2.00000
2020-02-13T00:00:00.000000000	1426.00000	62.00000	3.00000
2020-02-14T00:00:00.000000000	1457.00000	63.00000	3.00000
2020-02-15T00:00:00.000000000	1596.00000	66.00000	4.00000
2020-02-16T00:00:00.000000000	1696.00000	69.00000	5.00000
2020-02-17T00:00:00.000000000	1789.00000	74.00000	5.00000

Table2: Death cases in the last 10 days

The next figure shows the growth of death cases over time starting from Jan 21,2020 to Feb 17,2020.
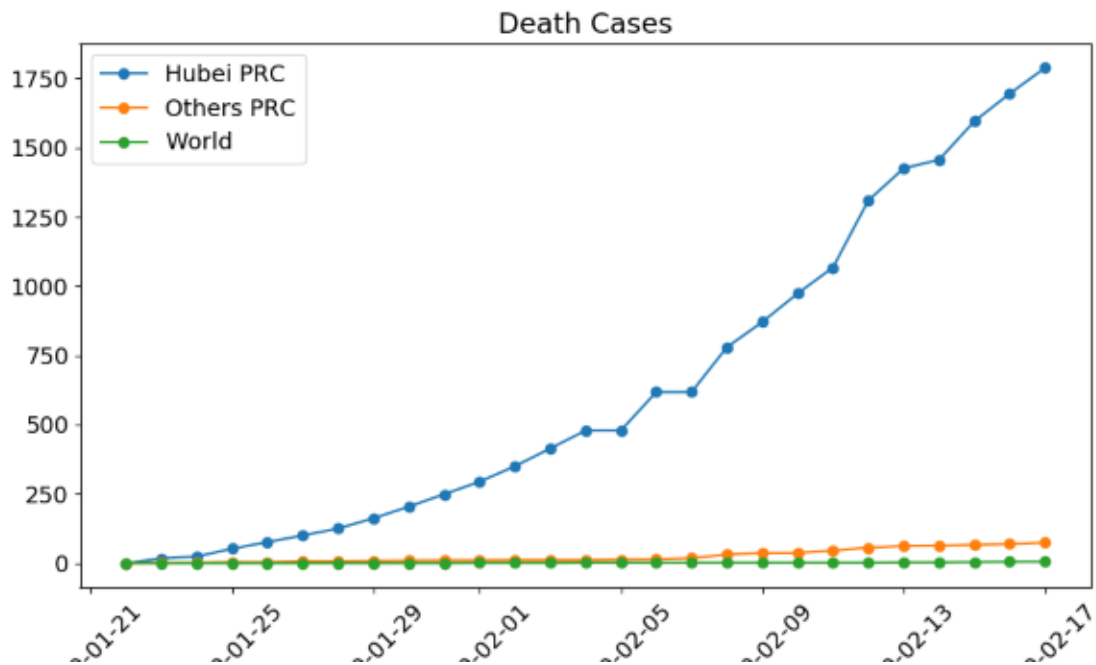

Figure 5: Growth of Death cases for the three regions

- Majority of death comes from Hubei province.

  Regarding to the Recovered cases we can see the number of cases on the last 10 days as

	Hubei PRC	Others PRC	World
2020-01-22T00:00:00.000000000	0.00000	0.00000	0.00000
2020-01-23T00:00:00.000000000	17.00000	1.00000	0.00000
2020-01-24T00:00:00.000000000	24.00000	2.00000	0.00000
2020-01-25T00:00:00.000000000	52.00000	4.00000	0.00000
2020-01-26T00:00:00.000000000	76.00000	4.00000	0.00000
2020-01-27T00:00:00.000000000	100.00000	7.00000	0.00000
2020-01-28T00:00:00.000000000	125.00000	7.00000	0.00000
2020-01-29T00:00:00.000000000	162.00000	8.00000	0.00000
2020-01-30T00:00:00.000000000	204.00000	9.00000	0.00000
2020-01-31T00:00:00.000000000	249.00000	10.00000	0.00000
2020-02-01T00:00:00.000000000	294.00000	10.00000	1.00000
2020-02-02T00:00:00.000000000	350.00000	11.00000	1.00000
2020-02-03T00:00:00.000000000	414.00000	11.00000	1.00000
2020-02-04T00:00:00.000000000	479.00000	11.00000	2.00000
2020-02-05T00:00:00.000000000	479.00000	13.00000	2.00000
2020-02-06T00:00:00.000000000	618.00000	14.00000	2.00000
2020-02-07T00:00:00.000000000	618.00000	18.00000	2.00000
2020-02-08T00:00:00.000000000	780.00000	31.00000	2.00000

Table3: Recovered cases in the last 10 days

The next figure shows the growth of death cases over time starting from Jan 21,2020 to Feb 17,2020.
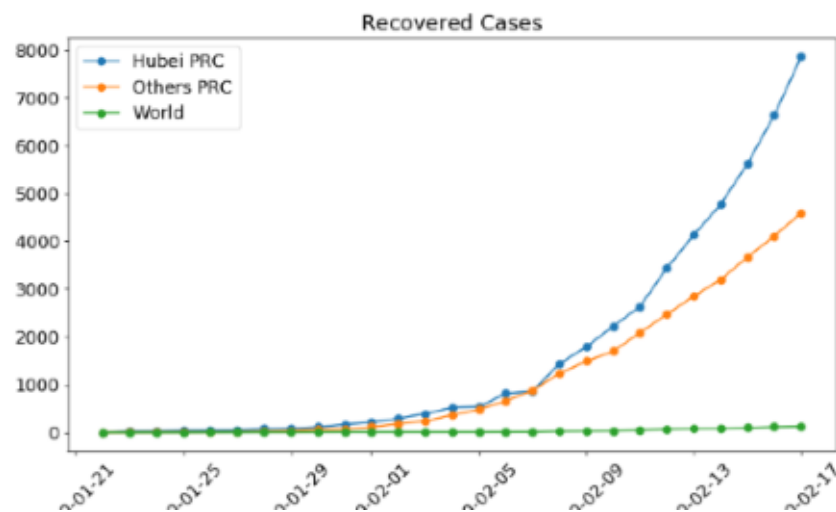

Figure 6: Growth of recovered cases for the three regions

We can see an exponential increase in recovered cases in China.

Rate of Death and Recovery
The next figure show us the rate of death and recovery over those three regions for time segments starting on Jan 21,2020 to Feb 17,2020.
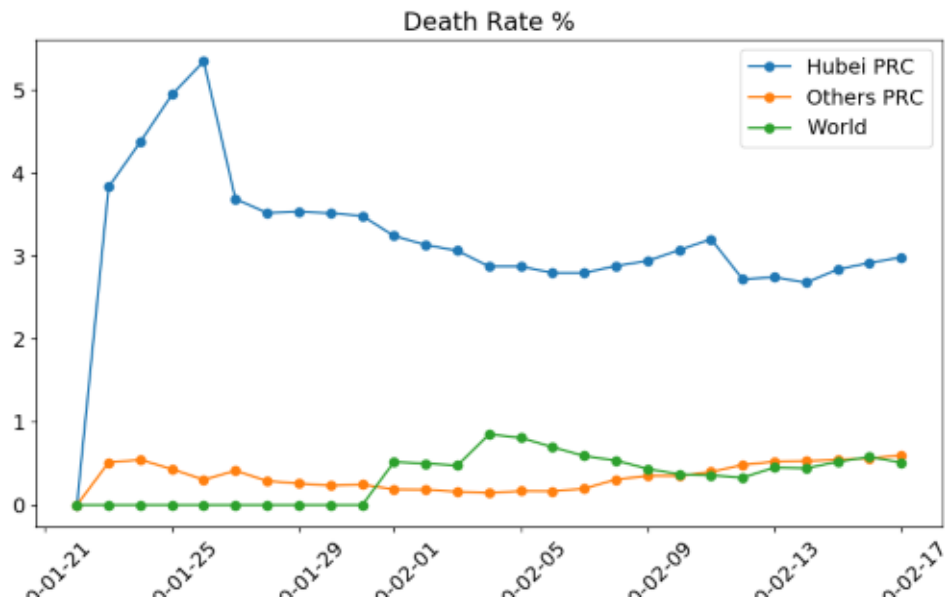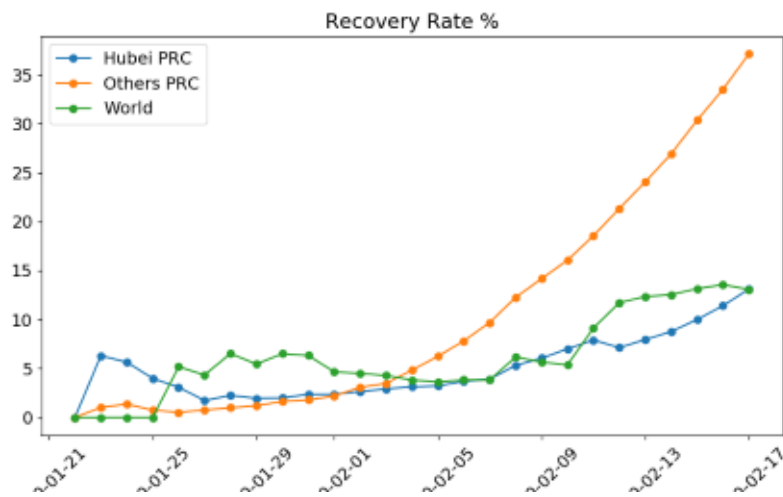

Figure 7: Death rates

Figure 8: Recovery rates

The average rate for these regions are listed in the following table :

Hubei PRC_death_rate	2.98221
Others PRC_death_rate	0.59798
World_death_rate	0.51073
Hubei PRC_recover_rate	13.10574
Others PRC_recover_rate	37.11515
World_recover_rate	13.07457

Table 4:Average recovery and death rates in the three regions

- Hubei has a higher death rate of 3%, Others PRC and the world are lower at $0.5 - 0.6\%$

- Recovery rate increased fast for PRC excluding Hubei, reaching 37% at the latest. Hubei and the world were similar at 13%. Both shows a rising trend.

  Extracting the data on 2019-nCov outbreak from the whole data, and start to a fitting algorithm for this data.The algorithm aimed at fitting the SEIR model to the incidence data for Hubei province.

The fitting algorithm

- **Description of fitting parameters**
  The list of parameters involved in the fitting procedure (see Table 5) includes four epidemiological parameters, β , ϵ , γ, k and I_0 from model equations (2) to (5),and two auxiliary parameters, Δ, r_{inc} corresponding to horizontal and vertical positioning of the modeled incidence curve relatively to the epidemic data points.
  The fitting procedure is based on several simplifying assumptions:

- I assume that the epidemic starts with the appearance of a small number of infected individuals in the population $I(0) = data(0)/k$ (where $data(0)$ is the first data value),, whereas $E(0) = 0$ Thus $S(0) = 1 - I(0)$

Definition	Description	Value	Value
	Epidemiological parameters		
β	Intensity of infection	Estimated	1/(person.day)
ϵ	Intensity of transition to infection form of the disease	Varied	1/day
γ	Intensity of recovery	Varied	1/day
k	the proportion of the infected population which is reported/observed	Varied	—*
I_0	Initial ratio of infected	$data(0)/k$	—*
	Curve positioning parameters		
r_{inc}	Relative vertical bias of the modeled incidence curve position	$[0.8; 1.0]$	—*
Δ	Absolute horizontal bias of the modeled incidence curve position	$1, 2, \ldots$	day

Table 1: Parameters for model fitting

*dimensionless

- Since we do not have any a *priori* information on the distribution of bias for the incidence data, we assume that the bias is independent in each point and normally distributed, which makes it possible to apply the least squares method to fit the model curve to data

Algorithm structure

Let $Y^{(dat)}$ be the set of incidence data points loaded from the input file and corresponding to one particular outbreak. Assume that the number of points is T, which equals the observed duration of the outbreak.
The limited-memory Nelder-Mead optimization method is used to find the best fit.For each value of Δ the algorithm varies the values of parameters β , ϵ,γ ,r_{inc} get the model output, which minimizes the distance between the modeled and real incidence points:

$$F\big(Y_{(mod)},Y_{(dat)}\big) = \sum_{i=1}^{n} \big(y_i^{(mod)} - y_i^{(dat)}\big)^2 \tag{12}$$

Here $y_i^{(mod)}$ and $y_i^{(dat)}$ represent the absolute 2019-nCoV incidence on the ith day taken from the input data-set and derived from the model, respectively.
Since the existence of several local minima is possible, the algorithm has to be started several times with different initial values of input variables. The best fit is chosen as a minimum of distances from all the algorithm runs.
The algorithm operations are performed in the following order. For each $Delta \in 1, 2, \ldots$:

For each fixed combination of values β , ϵ , γ, r_{inc} and I_0 generated by Nelder-Mead optimization procedure:

1.Find the numerical solution of model with the initial conditions mentioned above.

2-Calculate the modeling nCoV incidence in relative numbers:$y^{(mod,rel)}(t) = N_{E \to I}(t)$.
Since from SEIR model equations

$$E(t) = E(t-1) + N_{S \to E}(t) - N_{E \to I}(t)$$

and

$$N_{S \to E}(t) = S(t-1) - S(t)$$

We obtain:

$$y^{(mod,rel)}(t) = -\Delta S(t) - \Delta E(t), t = 1, 2, \ldots$$
$$\Delta S(t) = S(t) - S(t-1)$$
$$\Delta E(t) = E(t) - E(t-1)$$

3.We assume that the data incidence points from the data-set are shifted by Δ from the model curve start.Thus,we are to compare the distance between the following data sets :

$$Y^{(dat)} = \{y_0^{(dat)}, y_1^{(dat)} \ldots, y_{T-1}^{(dat)}\}$$

$$Y^{(mod)} = \{y^{(mod)}(\Delta), y^{(mod)}(\Delta+1) \ldots, y^{(mod)}(\Delta+T-1)\}$$

4.Convert the relative model incidence values to absolute values:

$$y_i^{(mod)} = y_i^{(mod,rel)}.N_L(m) \tag{13}$$

where $N_L(m)$ is the total population of the city L in the year m equal to the starting year of the considered epidemic.

5.Calculate the value of the fit function $F\left(Y_{(mod),Y_{(dat)}}\right)$, $F = F(\Delta)$

In the described manner for each value of Δ the Nelder-Mead algorithm finds the least distance F_Δ. We define $\Delta_{min} : F(\Delta_{min}, \ldots) = minF(\Delta, \ldots)$, and the parameter set $\{\beta$, ϵ,γ ,$r_{inc}\}$, corresponding to Δ_{min}. These values are the final result of our optimization procedure.

After the optimization algorithm has established the best fitting model parameter values, the model can be used to estimate the dynamics of population groups $S(t)$, $E(t)$, $I(t)$ and $R(t)$ over time

Simulation and Estimation of the parameters
I am going to simulate the SEIR model and plot both the data set and the measurement equation$y = kI$. Using the following parameter values: $\beta = 0.4$,$\epsilon = 0 : 25$,$\gamma = 0.071$ and $k = 80000$. For initial conditions, let $I(0) = data(0)/k$ (where $data(0)$ is the
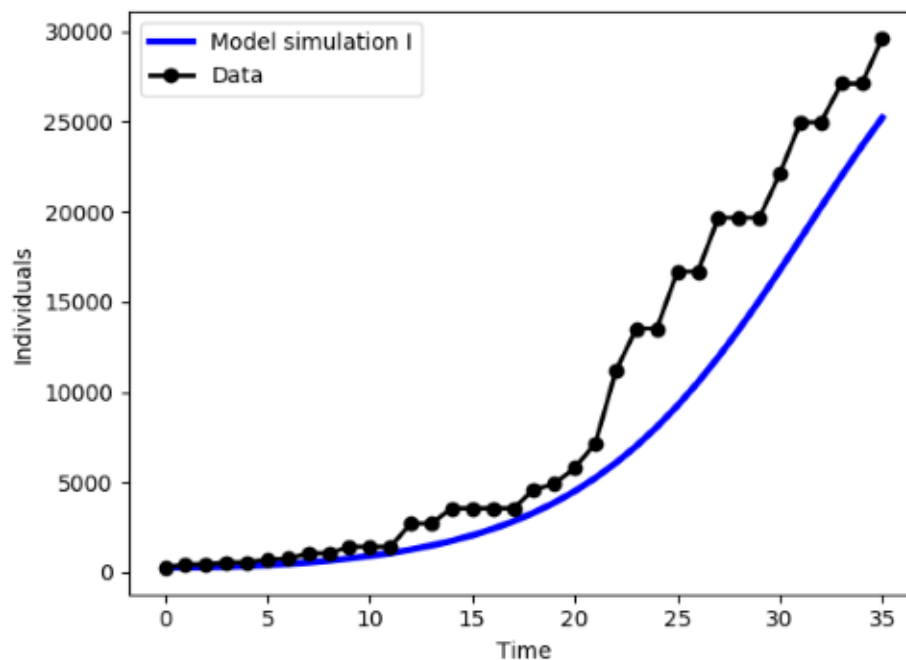first data value), $S(0) = 1 - I(0)$, and $R(0) = 0$.

Figure 9: Simulation of SEIR model using $\beta = 0.4$,$\epsilon = 0.25$,$\gamma = 0.071$

The next step is making parameter estimation.I make assumption here that the values of those paramters are not changed over the 35 days.Using python I write a code to estimate the parameters β , ϵ,γ and k using Poisson maximum likelihood. I used the parameters in the last simulation as initial values for the parameters using the Nelder-Mead optimizer.
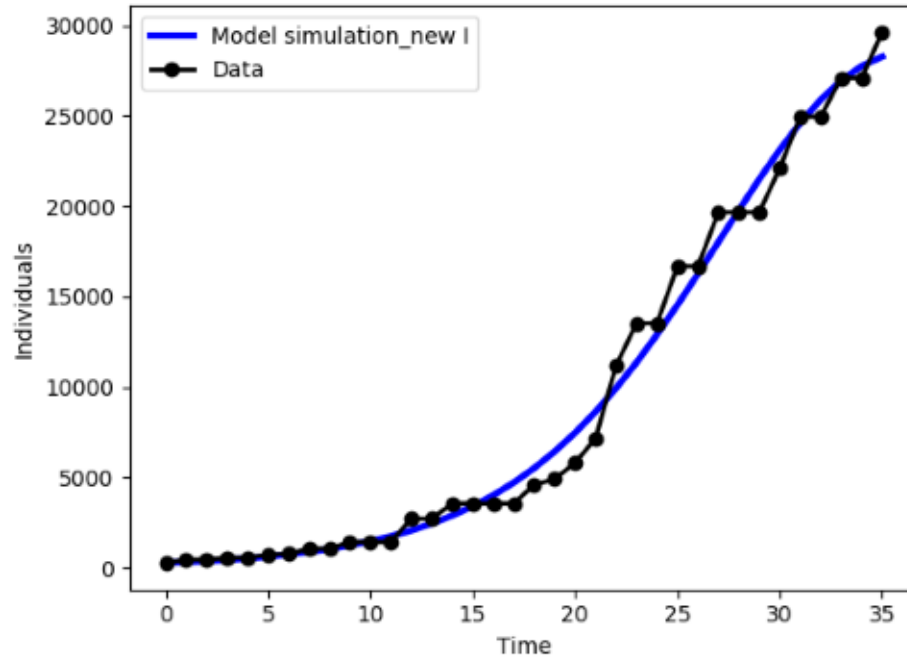
Figure 10: fitted SEIR model using Nelder-Mead optimizer

The parameters for the fitted model is :

$$\beta = 0.342$$
$$\epsilon = 5.52$$
$$\gamma = 0.155$$
$$k = 154842$$

Taking these values of the parameters I can make simulation after solving SEIR model equations. The dynamics $S(t)$, $E(t)$, $I(t)$ and $R(t)$ are as in the next figure
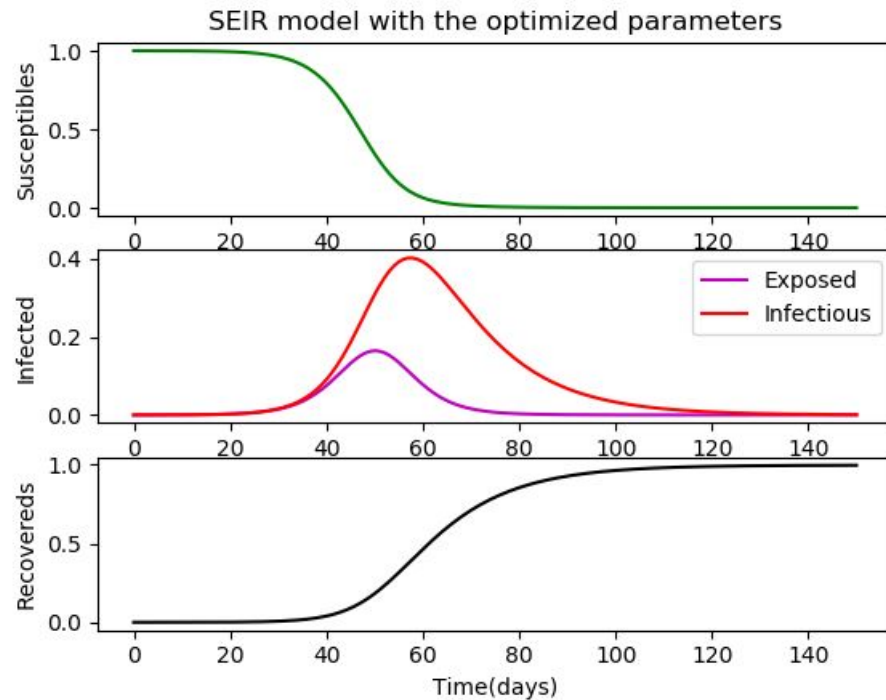
Figure 11: Dynamics of SEIR model using estimated parameters

These dynamics according to the estimated values show that the rate of infliction does not reach its maximum value yet. and after one month from now we will have approximately 320,000 infection cases The value of $R_0 = 2.20$ which is greater than 1 so the disease will spread out more.

Parameter Uncertainty: Profile Likelihoods

Now let's examine the structural and practical identifiability of the model parameters and generate confidence intervals using the profile likelihood. Generate profile likelihood for each of your model parameters β , ϵ,γ and k.

For the threshold to use in determining your confidence intervals, we note that $(NLL(p) - NLL(\hat{p}))$ (where NLL is the negative log likelihood) is approximately χ^2 distributed with degrees of freedom equal to the number of parameters fitted (including the profiled parameter). Then an approximate 95% (for example) confidence interval for p can be made by taking all values of p that lie within the 95th percentile range of the χ^2 distribution for the given degrees of freedom. In this case, for a 95% confidence interval, we have three total parameters we are estimating (β , ϵ,γ and k) ,so the χ^2 value for the 95th percentile is 7.8147. Then the confidence interval is any p such that:

$$NLL(p) \leq NLL(\hat{p}) + 7.8147/2 = NLL(\hat{p}) + 3.9074$$

Where $NLL(\hat{p})$ is the cost function value at our parameter estimates. The next figures show the profile likelihood of each model parameter.
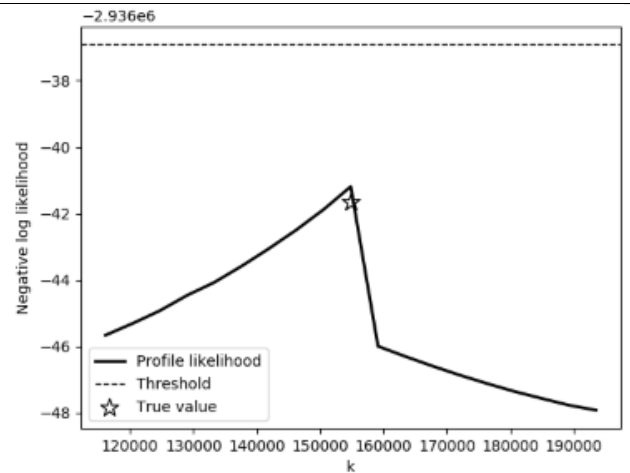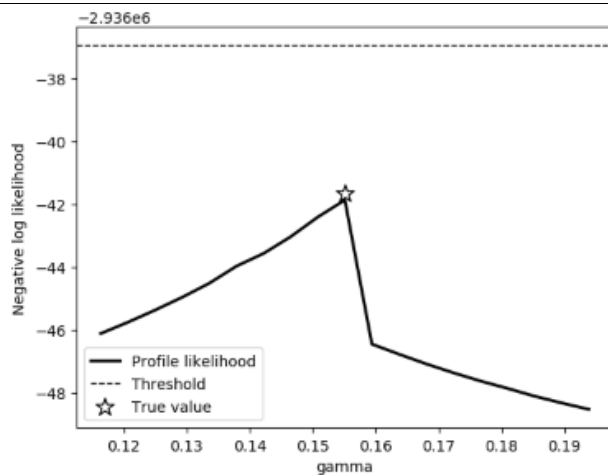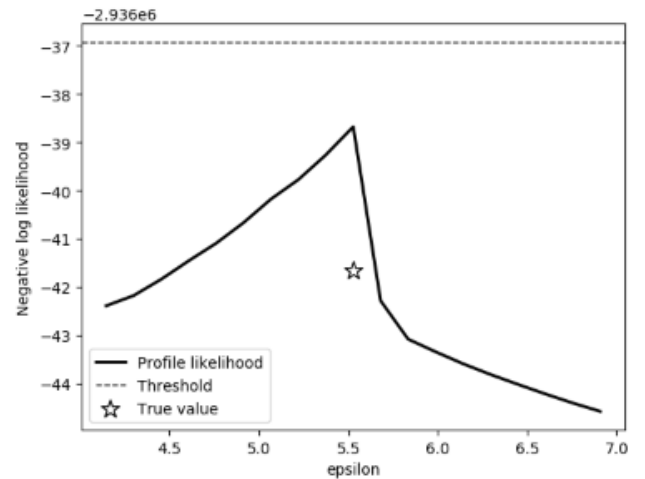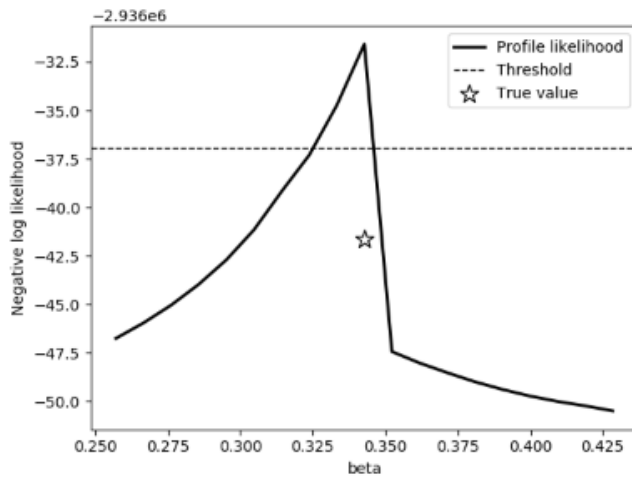
Conclusion and future work

It was found that in the compartmental model that assumes mass action incidence and homogeneous mixing, there are greater numbers of infectious individuals and the disease spread during an epidemic occurs more rapidly. It appears that disease spread is less acute when a more realistic assumption is made of heterogeneous mixing of individuals within the population.

Moreover, One limitation of this model that it does not consider the delay in time t since for that virus the incubation period is longer than many other disease.

There are several possible avenues of future study. One focuses on developing and analyzing variations of the SEIR model. For example, there could be more than one infectious stage in the model, so the model could be SEIIR similar. With more stages the challenge increases as to how one could define the model so that meaningful and useful mathematical analysis can be employed without paying too high a penalty in

complexity. An alternative avenue of study would be to take the SEIR model discussed here and try applying it to a particular disease whose behavior seems similar. Ideally the model could shed light on the disease behavior in a way that would assist health care workers in the containment or treatment of the disease. It is possible that a variation of the SEIR model would be most helpful in this regard.

In more detail, we can investigate a general SEIR-type epidemiological model, which incorporates appropriate compartments relevant to interventions such as quarantine, isolation and treatment. We stratify the populations as susceptible (S), exposed (E), infectious but not yet symptomatic (pre-symptomatic) (A), infectious with symptoms (I), hospitalized (H) and recovered (R) compartments, and further stratified the population to include quarantined susceptible (Sq), isolated exposed (Eq) and isolated infected (Iq) compartments.