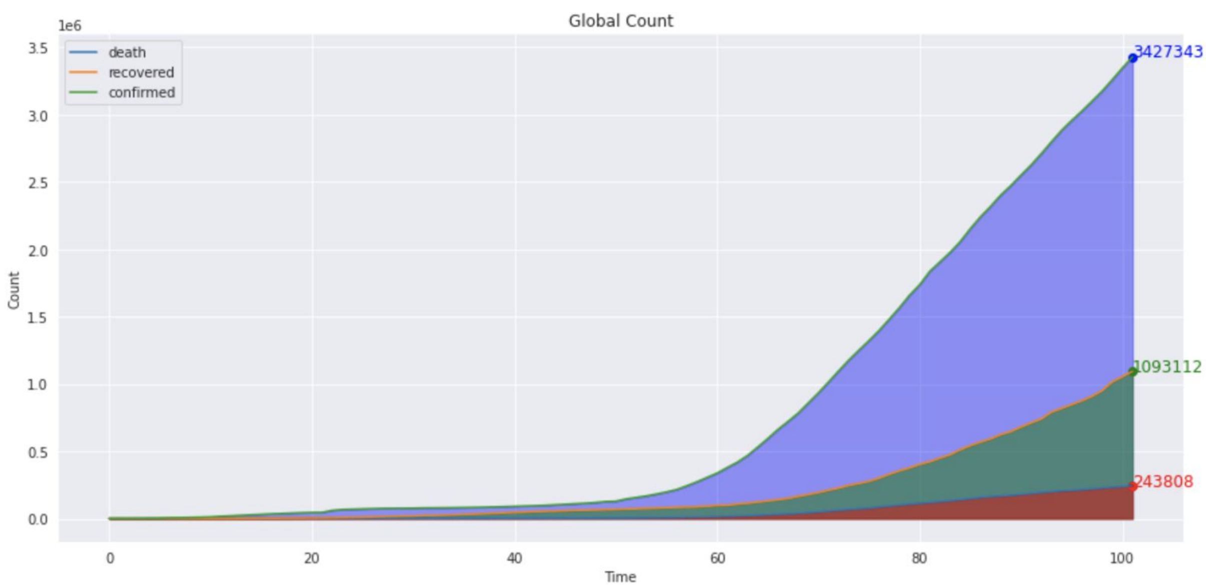
Coronavirus (COVID-19) Prediction Using Modified SIR Model

Tony Nguyen

Survey

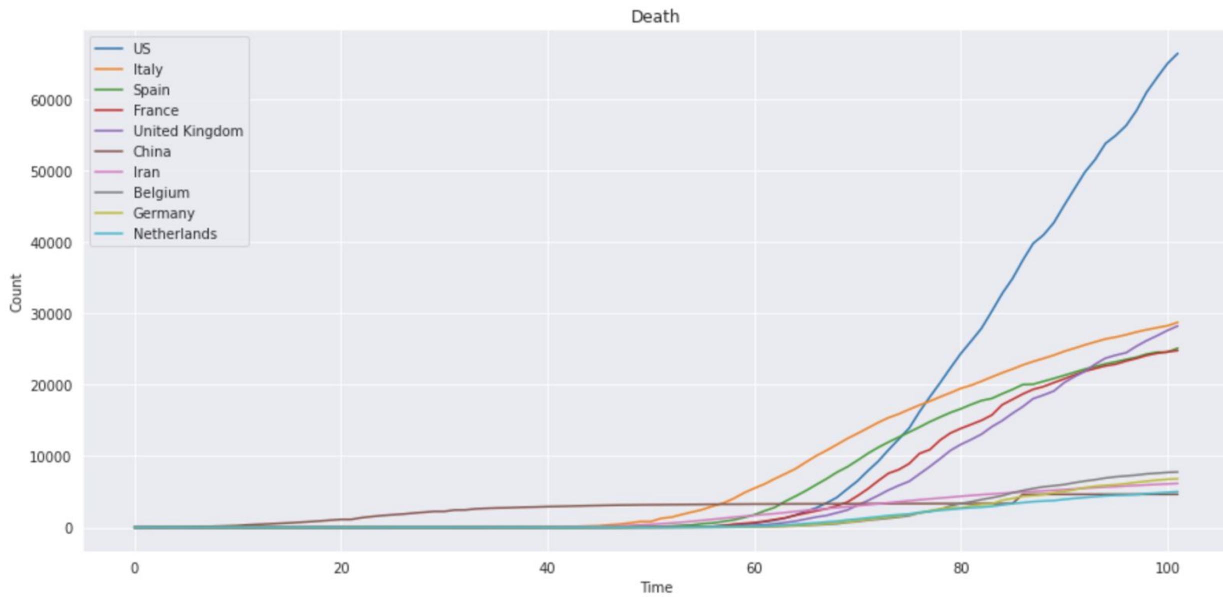
Recently, the COVID-19 pandemic, also known as the coronavirus pandemic, draws extensive attention from people everywhere in the world because of its negative impacts on everyone's life and the economy. Initially, coronaviruses are a large family of viruses which may cause illness in animals or humans. In humans, several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases, such as Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), [1] and the most recently discovered coronavirus causes coronavirus disease COVID-19. It is important to notice that the outbreak was identified in Wuhan, China - my second home country, in December 2019. [2] In this report, we will discover a model that can explain well the COVID-19 disease situation in China, which might potentially be applied to other countries and worldwide.
As the first step, we will analyze the Novel Coronavirus dataset [3] to provide a better overview of the current status of the COVID-19 pandemic. The dataset contains multiple data subsets of the COVID-19 confirmed cases, death cases and recovered cases with the time series. Firstly, we take a look into the Global Count figure below demonstrated the number of COVID-19 death, recovered and confirmed cases with x-axis is Time and y-axis is Count:

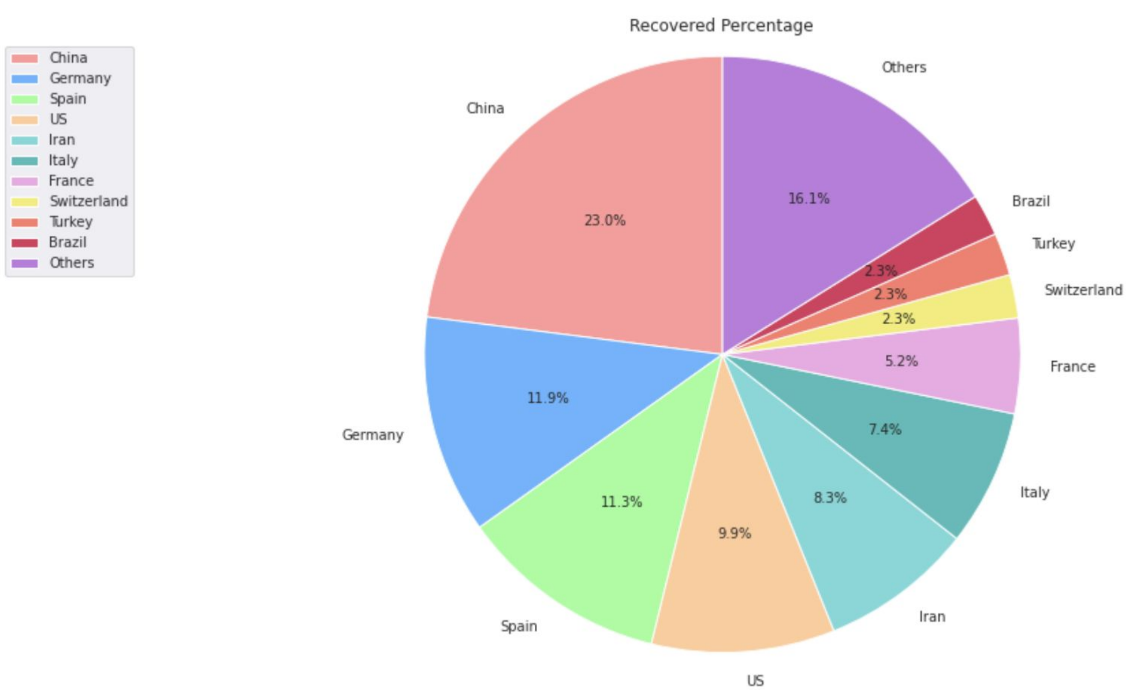


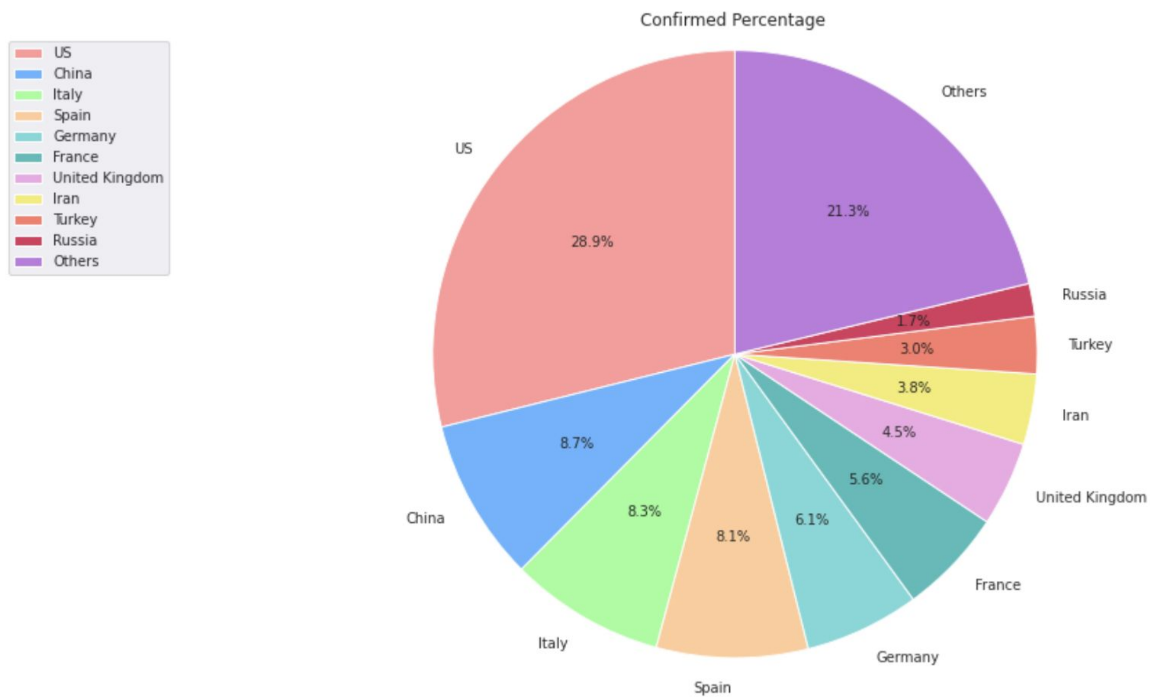

```
TOTAL DEATHS:  5217688.0
TOTAL RECOVERED PATIENTS:  21555429.0
TOTAL CONFIRMED CASES:  80420084.0
```

As an observation, the Confirmed has the steepest rise if compared to the Recovered and Death, in which the COVID-19 virus spreads exponentially with the total number of confirmed cases skyrockets to its peak within a few weeks. Within about 100 days, there are uncontrollable numbers of 80420084 confirmed cases and 5217688 death cases, which really shows us about the dangerous impact of the COVID-19.

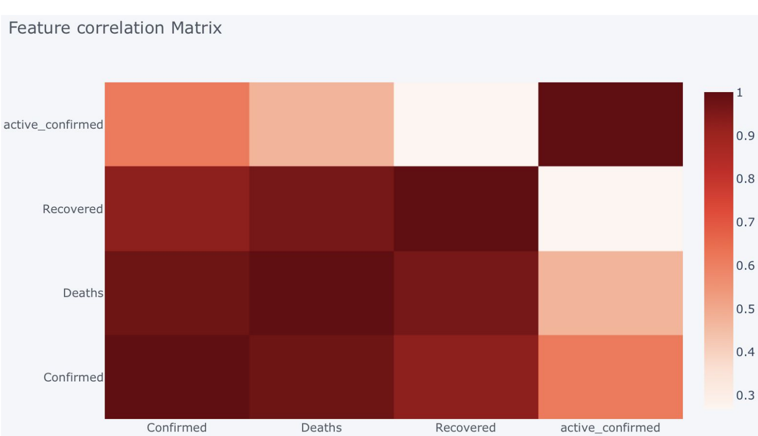
Moreover, we will analyze the effect of COVID-19 on different aspects, each with top 10 countries having the highest death, recovered and confirmed cases using the figures below:

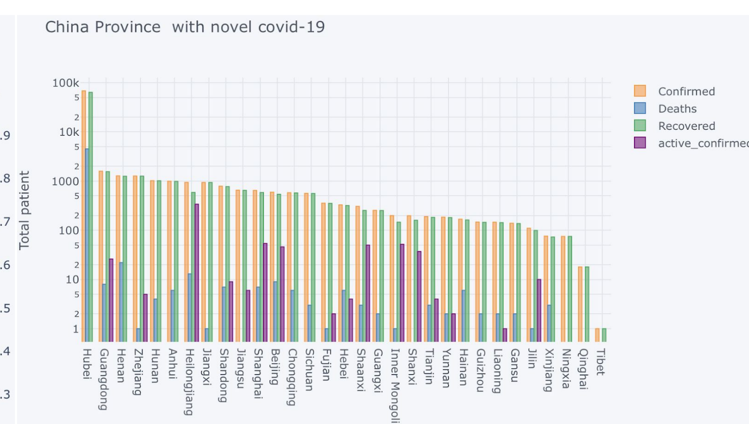




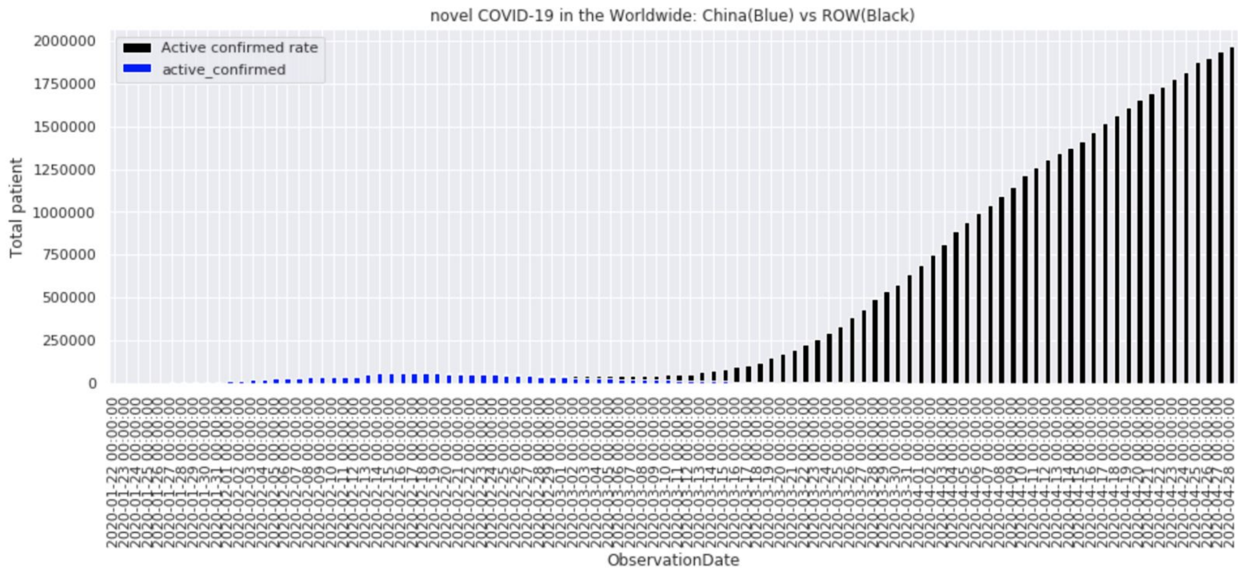


From the 3 figures, we can see that besides the US has the most increasing death rate, China is the country with the 2nd rank contributing to the number of confirmed cases and 1st rank in the number of recovered cases. Since China was the initial infected country, the COVID-19 behavior is different from the rest of the world. The medical system was not prepared for the pandemic, in fact no one was aware of the virus until several cases were reported. Moreover, the Chinese government took strong contention measures in a considerable short period of time and, while the virus is widely spread, they have been able to control the increase of the infections, which can also be proved from the below figures:





This heatmap result of China tells us that if we take for example deaths and recovered and fix all other features, recovered and deaths are more associated between them. And recovered is the opposite direction on deaths. In all provinces in China, the Chinese had the pandemic under control as we can see the recovered cases having almost equal rate if compared to the confirmed cases.



While the rest of the world (ROW, black color) in the figure above tremendously increases in the active_confirmed cases rate for non-stop, China instead decreases its rate of active_confirmed cases after reaching its peak. However, there are very limited resources that have models to potentially explain the COVID-19 disease control in those countries have successfully done it. Therefore, we will go ahead and find out a modified model from the original SIR model to explain in the case of China.

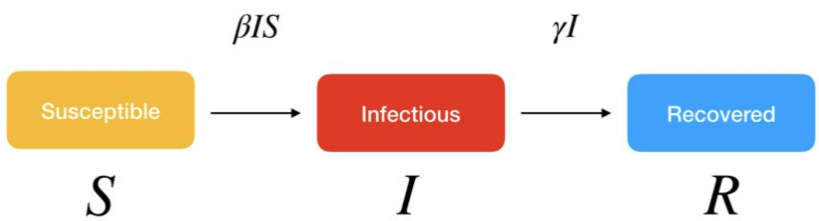
Experimental analysis

The original SIR model is a classic model in epidemiology, which contains three subpopulations, the susceptibles **S**, the infectious **I** and removed individuals **R**: [4]

1. Susceptible **(S)**: the individual hasn't contracted the disease, but she can be infected due to transmission from infected people.
2. Infectious **(I)**: this person has contracted the disease.

3. Recovered/Deceased **(R)**: The disease may lead to one of two destinies: either the person survives, hence developing immunity to the disease, or the person is deceased.

Diagrammatically: [5]



β: Effective contact rate [1/min]

γ: Recovery(+Mortality) rate [1/min]

In this context of COVID-19:

Susceptible is people containment

Infectious is currentConfirmedCases

Recovered is recovered and deaths

The susceptible can become infectious, and the infectious can become removed, but no other transitions are considered. In fact, there are many versions of this model, considering birth and death (SIRD with demography), with intermediate states, etc. However, since we are in the early stages of the COVID-19 expansion and our interest is focused in the short term, we will consider that people develops immunity (in the long term, immunity may be lost and the COVID-19 may come back within a certain seasonality like the common flu) and there is no transition from recovered to the remaining two states. The population $N = S + I + R$ remains constant. The model describes the movement between the classes by the system of differential equations:

$$dS/d\tau = -\beta IS$$

$$dI/d\tau = \beta IS - \gamma I$$

$$dR/d\tau = \gamma I$$

In terms of fraction $x = S/N,\ y = I/N,\ z = R/N,$ and the rescaled time variable $t = \gamma\tau$, the equations become:

$$dx/dt = -R_0 xy$$

$$dy/dt = R_0 xy - y$$

$$dz/dt = y$$

where $R_0 = \beta N/\gamma$ which is called reproduction ratio

We can use only two equations to define a system:

$dx/dt = -R_0 xy,\ dy/dt = R_0 xy - y$ and find $z = 1 - x - y$

Mark region $D = \{(x, y) \in R^2 \mid 0 \le x, y \le 1,\ x + y \le 1\}$ in the (x,y)-plane.

By this model, we will define a modified SIR model which can explain the behavior of confirmed cases, current confirmed cases and (recovered + death) in China.

x is a susceptible fraction

y is a infectives fraction

z is a recovered fraction

We acknowledge that $z = 1 - x - y$. So, the confirmed case fraction is $x^c = 1 - x$

Having $z = x^c - y$ that is similar to recovered + deaths = confirmed case - current confirmed case.

We have modified a SIR Model which now becomes a new model called Modified SIR Model for COVID-19 in China.

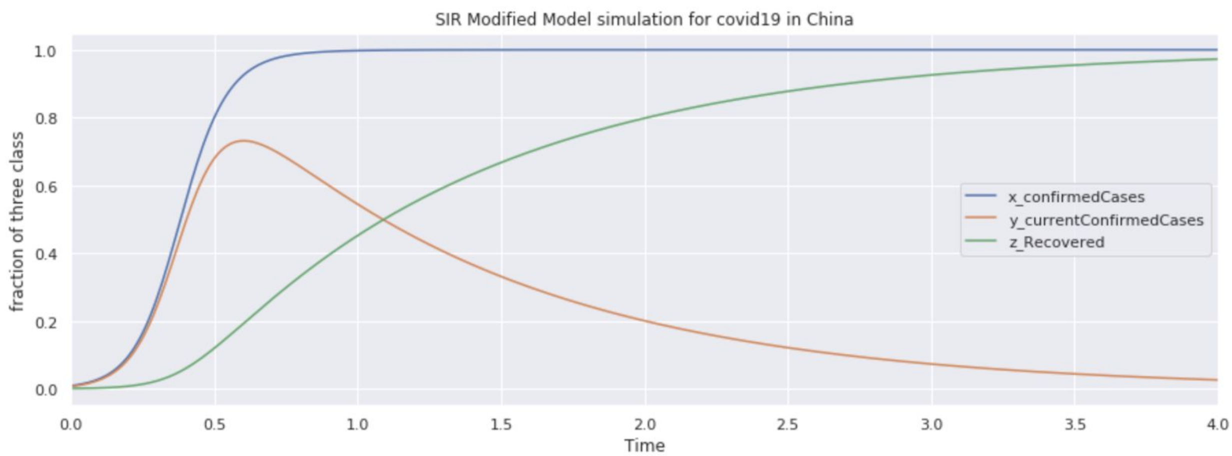
With $x = 1 - x^c$ we have:

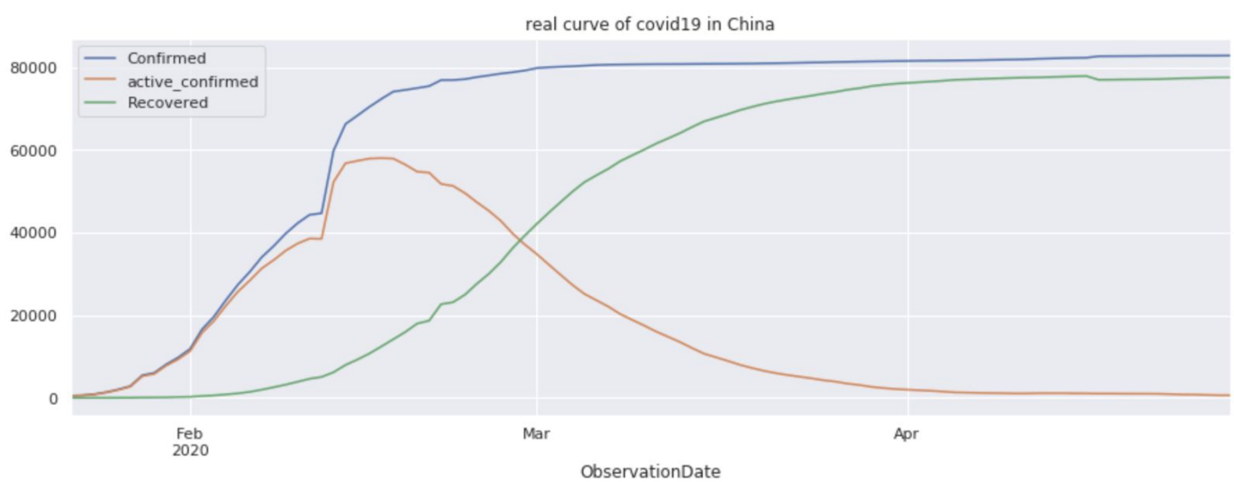
$$dx^c/dt = R_0(1-x^c)y$$

$$dy/dt = R_0 y(1-xc)-y$$

$$z = x^c-y$$

For testing, we are going to do some simulation of this model to see if it matches well, following with the below figures:

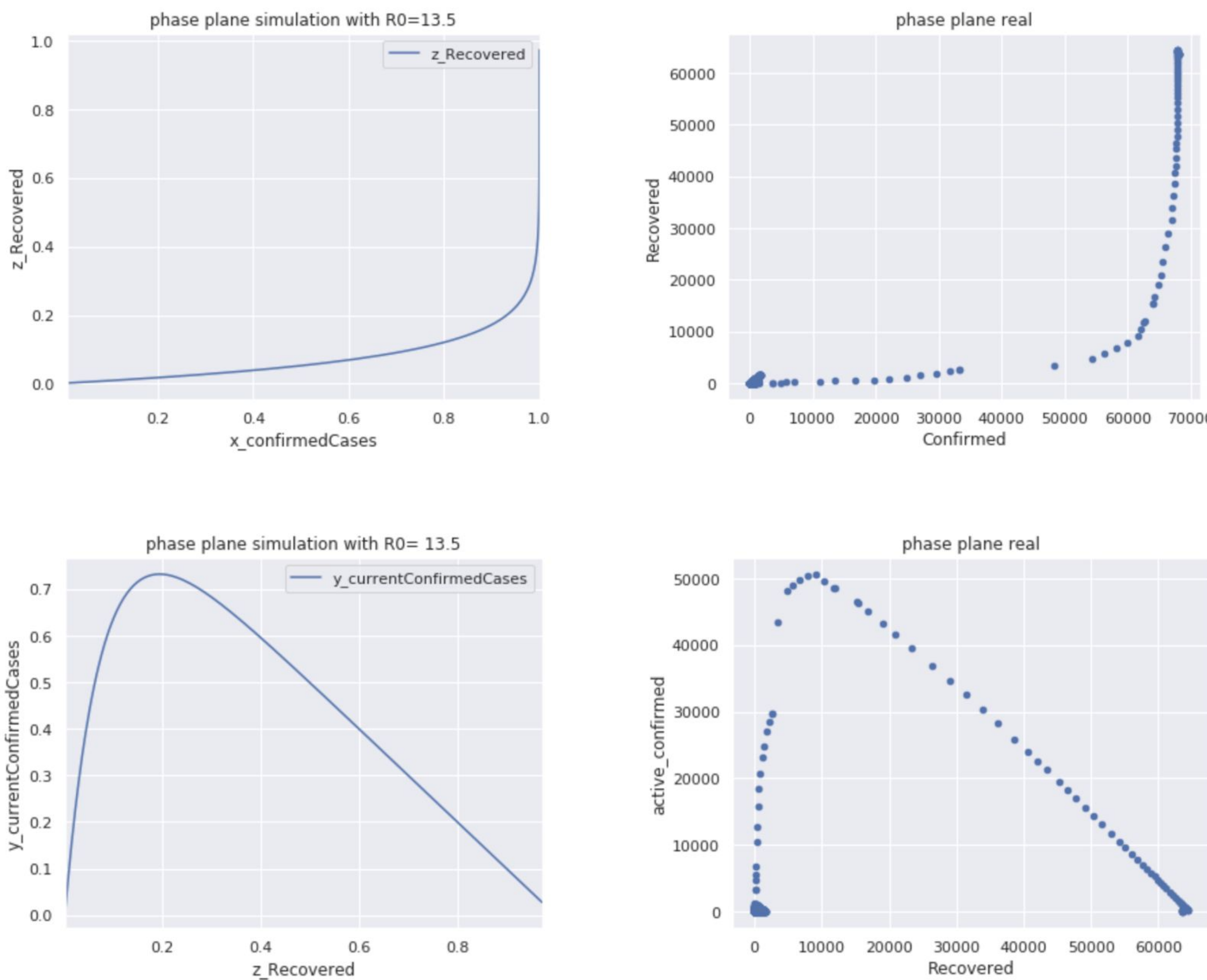




As we observe, the curves of our Modified SIR Model mostly match the shape of the real curves of COVID-19 in China. However, there are still rooms for improvement since the curves in our

Modified SIR Model are smoother, which would not demonstrate some unusual behaviors of the data.

Specifically, we will look closer into how our Modified SIR Model behaves with each categories as shown below:



As the 1st scenario with x-axis is confirmed cases, y-axis is recovered cases, and the 2nd scenario with x-axis is recovered cases, y-axis is active-confirmed cases (both with $R_0 = 13.5$), if we connect all the dots in the real phase plane (extracted from real data of COVID-19 in China), we will obtain the similar curves if compared to the curves from our Modified SIR Model.

Conclusions

As a result, we can conclude that our Modified SIR Model might explain well the COVID-19 control in China with minimal differences if compared to the real results extracted from the Novel Coronavirus dataset. However, our model still needs improvement to reflect the real scenarios with some unusual behaviors of the data. Moreover, we also can potentially apply our Modified SIR Model to other countries and the worldwide to predict some results in the future. Yet, it will be a challenge for us to test the accuracy of our Modified SIR Model with other countries and the worldwide because China is just among the first countries to put COVID-19 under control until now while we are still fighting with it as the rest of the world. Therefore, it will be nearly impossible for us to obtain the real data after the COVID-19 peak, for example, from a country which has not reached its COVID-19 peak. In the case of China, which all policies that Chinese authorities are implemented allow to control COVID-19 and lead to the end of the pandemic in China soon (if another disease does not come), we should all learn from it to rescue the world from the COVID-19 pandemic.

References

[1] https://www.niaid.nih.gov/diseases-conditions/covid-19

[2] https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/

[3] https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset

[4]

https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model

[5] https://epubs.siam.org/doi/10.1137/S0036144500371907