

MoC-SEIR: Analysis of the Epidemic Situation of 2019-nCoV with Various Factors

Summary

The coronavirus ignites a fast-growing outbreak in China. To date, the death toll in China has risen to 811, surpassing the death toll from the SARS epidemic of 2002-3, according to official data released early on the morning of Feb. 9.

To assist the Chinese Center for Disease Control and Prevention (China CDC) better in decision making, we build a **modified compartmental SEIR model (MoC-SEIR)** in search for potential demographic trend of infection. Infectious disease model is frequently used to simulate the spread of disease. Given some characteristics of the coronavirus including an averaged 7-day incubation period and relative low fatality rate compared to SARS, we choose **SEIR** as the basis of our model among all mathematical models of epidemic disease.

Since the virus originated in Wuhan, We divide the regions into three groups, Wuhan, Hubei without Wuhan and Mainland China ignoring the Hubei, with different parameters in model. The raw SEIR does not explain well about the data we collected. So we take population flow and travel limitations into our model, which yields more satisfactory results. **The basic reproductive number of the virus is 3.15** calculated based on our parameters. We also make some predictions about how the trend will develop. A decrease in number of infections is expected around Feb. 18. We also build a model on a global scale. **Grey Verhulst model** is introduced to forecast oversea infections as little information about the foreign cases is disclosed, suggesting the spread trend has been stemmed abroad. That is to say, the virus would not cause too much trouble for other countries.

Based on our model, we analyze and quantify the impacts of measures taken by the Chinese government. The imposed travel limitations has greatly blunted the outbreak's impact and prevented a broader outbreak. The extended holidays postponed the huge population reflow which could bring devastating damage to the city when the virus is highly contagious. But **the coin has two sides**. The abrupt lockdown in a way overstretched the hospitals in Wuhan, which borne most brunt of the epidemic. The disaster also put tertiary industry and industrial production under severe strain.

Besides, we introduce various factors into our model. **Secondary infection, emergence of effective treatment and returning after Spring Festival** are reasonable expectations in the development of the outbreak. We consider their influence on the development of the outbreak respectively. Given the prediction on the global trend, we have every reason to believe that World Health Organization(WHO) will terminate the PHEIC by **the end of March**. In the last analysis, we test the stability of model and present both strengths and weaknesses.

Measures, in either personal or national perspective, are fully discussed, which might reduce the chance of virus spread between individuals. In future works, we hope to dive into the subtle influences of expositors and population variation.

Keywords: 2019-nCoV; SEIR Model; Grey Verhulst; Population Flow; Control Measure

MoC-SEIR: Analysis of the Epidemic Situation of 2019-nCoV with Various Factors

February 10, 2020

Contents

1	Introduction	3
1.1	Background	3
1.2	Related work	3
1.3	Our Work	4
2	Assumptions and Justifications	4
3	Nomenclature	5
4	Modeling	5
4.1	Data Collection	5
4.2	MoC-SEIR: Modified Compartmental SEIR Model	5
4.2.1	Motivation	6
4.2.2	Iterative Equations	7
4.2.3	Parameter Determination	8
4.2.4	Basic Reproductive Number	8
4.3	Implementation and Result	9
4.3.1	Model Validation	9
4.3.2	Domestic and Global Predictions	10
4.3.3	Quantification of Measures' Effect	11
4.4	Emergence of Various Factors	12
4.4.1	Presence of Secondary Infection	12
4.4.2	Discovery of Effective Treatment Method	13
4.4.3	Returning after Spring Festival	13
5	Model Analysis	14
5.1	Sensitivity Analysis	14
5.1.1	The Sensitivity of b and γ	14

5.1.2	The Sensitivity of θ and k	15
5.2	Strengths and Weaknesses	16
5.2.1	Strengths	16
5.2.2	Weaknesses	16
6	Application	16
7	Conclusion	17
8	Blog	18
Appendices		20
Appendix A Grey Verhulst Model		20
Appendix B Data Preprocessing		20
Appendix C SEIR Model		21

1 Introduction

1.1 Background

An outbreak of a pneumonia-like illness that started in the Chinese city of Wuhan has put health authorities on high alert around the world. The novel coronavirus was officially named as 2019-nCoV, which was identified as the causative virus for the acute respiratory disease on 7 January 2020. It has led cities to lockdown, cancellation of flights to and from China and disruption in global business. It takes a toll on all aspects of people's life. The WHO declared a global emergency on Jan. 30 as the illness continued to spread.



Figure 1: Confirmed Cases in Wuhan and China

However, as there exists a one-week incubation period for the infected on average, massive people have unconsciously exposed to virus carriers and then travelled to other places before the widespread outbreak. Although the Chinese government takes powerful measures in a short time, trying to prevent the epidemic situation from worsening, how to control the trend of outbreak development effectively still becomes an intractable problem.

1.2 Related work

Since the abrupt increase in the number of 2019-nCoV infections in China, much attention has been paid to the research on its trend and prevention. Many teams estimated the reproductive number of the virus based on the preliminary statistics, such as 3.11 by J. M. Read, *et al.* [1], 2.24 to 3.58 by Shi Zhao, *et al.* [2] and so on. Some efforts have been made looking into the taken and future control measures of the Chinese government, including travel restrictions, holiday extension etc. Tang Biao et al. worked on how travel restriction would limit the spread in Beijing [3] and Ai Siqi et al. explored the relation between the number of cases and the timing of city closure policy [4].

Up to the present, most articles build their models for a certain region, but few work constructs a systematic mathematical model to describe the interzone spread of 2019-nCoV, which is the main purpose of our attempt in this paper.

1.3 Our Work

In summary, our work addresses three primary problems:

1. According to current reports and statistics on the appraisal of the epidemic, we build a compartmental model based on the SEIR [5] proposed by Kermack et al. to predict the spread trend of 2019-nCoV. Our model fits the collected data well and gives a reasonable prediction of both domestic and global infections in the next period of time.
2. As traditional transportation brings a sudden and significant increase to the intensity of population flow, we quantify its influence based on epidemiology, either with or without the prevention and control measures by Chinese government, and thus explain the positive and negative effects of those measures.
3. Based on our prediction model, we list some factors closely related to the development of epidemic situation and analyze their impact in the next two months. We also show that WHO will terminate the PHEIC in the end of March.

In Section 2, we state the basic assumptions of our model. Section 3 contains the nomenclature used for model description. Section 4 presents our modeling process and result. The characteristics of our model and the sufficient condition of high parameter sensitivity are discussed in Section 5 and some trains of thought against the spread of virus are proposed in Section 6. Finally, we conclude our paper in Section 7.

2 Assumptions and Justifications

Given current evidence and reports, our model makes the following assumptions:

- **Virus carrier flows except those out from Wuhan are ignored.** According to our collected data, most victims of the virus got infected in Wuhan and the infection ratio in other regions is much lower. When considering the influence of population flows on the epidemic situation, only those from Wuhan to the rest of provinces make a significant difference.
- **The total population of any region is considered constant.** Birth and death are omitted when considering the total population. On one hand, we have few access to the detailed statistics of population flows and birth/death rate of a certain region. On the other hand, all of them make little difference to the spread of virus.
- **The incubation period is a random variable with exponential distribution with parameter θ .** The Exposed developed into infectious ones at a proportion of θ .
- **There is no outflow of population in Hubei once the extreme control measures are taken,** which means virus transfer among provinces is suppressed entirely. So we do not consider imported cases in this stage.
- **The infections brought by expositors are neglected.** Although it's reported that expositors are also able to spread the virus, their spread ability is relatively much weaker than the infected ones and thus does not make significant difference to the epidemic situation.

- **The mutation speed of virus is relatively slow.** The cured individuals are immune to the virus. As for task 3, we will consider the scenario of secondary infection.

3 Nomenclature

Here are some notations to clarify before we introduce our work. The meaning of the superscript is also shown in table 1.

Table 1: Nomenclature

Notation	Definition
D	Death toll
R	Individuals having recovered from infection
I	Infected individuals
C	Confirmed cases
E	Exposed individuals
S	Susceptible individuals
P	The total population of a region
F	The flowing population of a region
θ	The reciprocal of incubation period
$f_{t,o}$	The ratio of outflow population from a region at t
$f_{t,i}$	The ratio of inflow population to a regions at t
$f^{(0)}$	f_{out} or f_{in} of Wuhan
$f^{(1)}$	f_{out} or f_{in} of other regions in Hubei
$f^{(2)}$	f_{out} or f_{in} of other regions outside Hubei in China

4 Modeling

4.1 Data Collection

To have a better insight into this outbreak and simulate our model, we first collected the number of confirmed cases from a GitHub repository [6]. We organize the data in a temporal-spatial order to facilitate our training. Most provinces declared its first confirmed case around Jan. 22. Therefore, our data for these regions ranges from Jan. 22 to Feb. 4. However, since the virus originate from Wuhan in Hubei province, more data can be acquired from Hubei government official sites compared to other provinces. So more previous data has been collected for Hubei.

4.2 MoC-SEIR: Modified Compartmental SEIR Model

Compared with common infectious diseases, we find nCoV is a bit different in that carriers are likely to spread it even during incubation period when carriers might show similar symptoms with those suffering from serious flu. The trait makes it harder to distinguish between the positive and negative ones. It also translates into the soaring potentially infected population flowing out. Therefore, we modify the classic SEIR model so that it can fit current situation more precisely. In our modified SEIR model, each member of the population typically progresses from susceptible to exposed, then to infectious and finally to recovered

or dead. The whole process is illustrated in the flow diagram Figure 2, in which the boxes represent the different compartments and the arrows the transition between compartments.

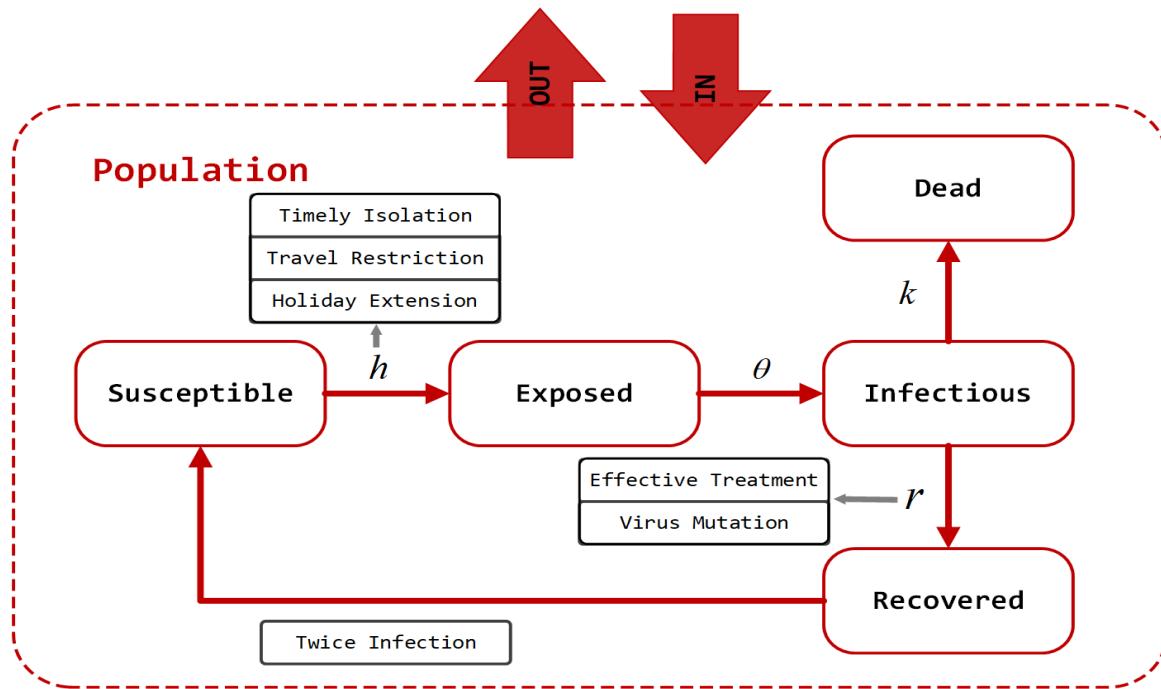


Figure 2: Modified Compartmental SEIR with Population Flow

4.2.1 Motivation

Classic SEIR seems a satisfying model for the recent outbreak. It gives a curve of infections that grows exponentially in early stage and then levels off. However, given the known spread intensity of nCoV, we find that the curve has great deviation from the reported data in Wuhan, as shown in Figure 3. At the same time, we can hardly explain why countless infections have appeared in other regions of China without the idea of interzone population flows. In addition, it is obvious in Figure 4 that the number of confirmed cases is highly correlated ($r \approx 0.85$) to the scale of inflow population from Wuhan. Therefore, we think it is necessary to introduce a variable to describe the effect of population flow.

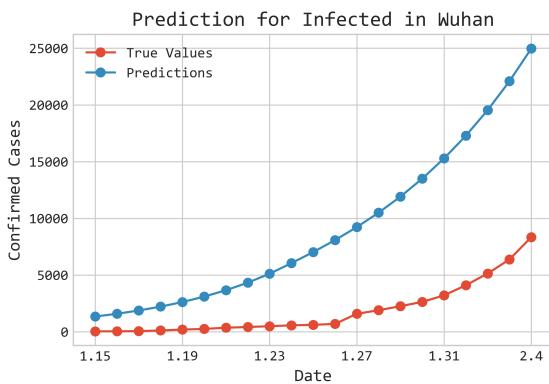


Figure 3: Curve Without Pop. Flow

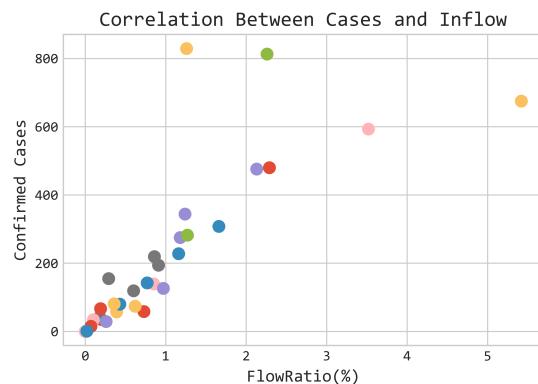


Figure 4: Correlation Between Cases and Inflow

4.2.2 Iterative Equations

Like classic SEIR, first we have γ as the chance of getting infected when an uninfected person is close to infectious one, a as the probability of a susceptible individual staying close to an exposed or infectious person, b as the infection efficiency and $h = \gamma abI$, where h stands for the mathematical expectation of daily infection ratio among P . The exposed transitions to infectious at a rate of θ . The total infectious individuals die at a proportion of k and recover at a proportion of r per day.

As we mentioned before, incubation period is a random variable with exponential distribution with parameter θ . We also assume the birth rate equals death rate and thus the total population stays constant. Therefore, the basic form of our model's differential equations can be easily inferred as (1) ~ (5), where $j \in \{0, 1, 2\}$:

$$S_t^{(j)} = P_t^{(j)} - E_t^{(j)} - I_t^{(j)} - R_t^{(j)} - D_t^{(j)} \quad (1)$$

$$\frac{dE_t^{(j)}}{dt} = hS_t^{(j)} + (f_{t,i}^{(j)} - f_{t,o}^{(j)} - \theta) E_t^{(j)} \quad (2)$$

$$\frac{dI_t^{(j)}}{dt} = \theta E_t^{(j)} - (r + k + f_{t,o}^{(j)} - f_{t,i}^{(j)}) I_t^{(j)} \quad (3)$$

$$\frac{dR_t^{(j)}}{dt} = r I_t^{(j)} \quad (4)$$

$$\frac{dD_t^{(j)}}{dt} = k I_t^{(j)} \quad (5)$$

Considering that the number of a certain group varies little every 24 hours, we use difference equations (6) ~ (10) instead and iterate once a day in our implementation corresponding to computational efficiency.

$$S_t^{(j)} = P_t^{(j)} - E_t^{(j)} - I_t^{(j)} - R_t^{(j)} - D_t^{(j)} \quad (6)$$

$$\Delta E_t^{(j)} = E_{t+1}^{(j)} - E_t^{(j)} = hS_t^{(j)} + (f_{t,i}^{(j)} - f_{t,o}^{(j)} - \theta) E_t^{(j)} \quad (7)$$

$$\Delta I_t^{(j)} = I_{t+1}^{(j)} - I_t^{(j)} = \theta E_t^{(j)} - (r + k + f_{t,o}^{(j)} - f_{t,i}^{(j)}) I_t^{(j)} \quad (8)$$

$$\Delta R_t^{(j)} = R_{t+1}^{(j)} - R_t^{(j)} = r I_t^{(j)} \quad (9)$$

$$\Delta D_t^{(j)} = D_{t+1}^{(j)} - D_t^{(j)} = k I_t^{(j)} \quad (10)$$

According to our first two assumptions, we have three cases for different regions:

1. **Wuhan** There are little population flow into Wuhan city, meaning we can set $f_{t,i}^{(0)} = 0$ and thus simplify our difference equations (7) and (8) as following:

$$\Delta E_t^{(0)} = E_{t+1}^{(0)} - E_t^{(0)} = hS_t^{(0)} - (f_{t,o}^{(0)} + \theta) E_t^{(0)} \quad (11)$$

$$\Delta I_t^{(0)} = I_{t+1}^{(0)} - I_t^{(0)} = \theta E_t^{(0)} - (r + k + f_{t,o}^{(0)}) I_t^{(0)} \quad (12)$$

2. **Hubei except for Wuhan** A large proportion (up to two thirds) of Wuhan's outflow population does not leave Hubei province, causing the epidemic situation in other regions inside Hubei deeply influenced by this group. Simultaneously, we can reasonably leave out the outflow in this case as the density of virus carriers is much lower.

So just let $f_{t,o}^{(0)} = 0$ in equations (7) and (8), and we will obtain the simplified equations specifically:

$$\Delta E_t^{(1)} = E_{t+1}^{(1)} - E_t^{(1)} = hS_t^{(1)} + \left(f_{t,i}^{(1)} - \theta \right) E_t^{(1)} \quad (13)$$

$$\Delta I_t^{(1)} = I_{t+1}^{(1)} - I_t^{(1)} = \theta E_t^{(1)} - \left(r + k - f_{t,i}^{(1)} \right) I_t^{(1)} \quad (14)$$

When considering the whole Hubei Province, we add the two cases above together.

- 3. **Other Provinces** As virus carrier flows from these regions are ignored, we still have $f_{t,o}^{(2)} = 0$ and thus get the simplified expressions for ΔE_t and ΔI_t the same as equations (13) and (14).

4.2.3 Parameter Determination

In our modified SEIR model, we need to choose suitable values for five parameters, which is θ , r , k , γ , and b . As will be shown in Section 5, some of them tend to have visible impacts on our models performance.

- Some researchers have given their predictions of 2019-nCoV's incubation period, ranging from five to eight days, as mentioned in [7]–[9]. Without loss of generality, we take one week as the incubation period, and thus $\theta = 1/7$.
- The average mortality and survival rate has leveled off in the past few days, and we take 3.5% for both of them. And we assume a one-week recover period $\lambda^{-1} = 7$ for each 'infectious' on average. Based on the theory of exponential distribution, we have $k = 0.035\lambda = 0.5\%$ all the time, $r = 0.035\lambda = 0.5\%$ in the first 30 days and $r = 2 \times 0.035\lambda = 1.0\%$ as the more infected people are diagnosed and treated earlier.
- According to recent news, the spread ability of 2019-nCoV is astonishing, as a large proportion of the infectious got infected during a very short stay with other carriers. Thus we choose a large constant, letting $\gamma = 0.75$.
- Till now there have not been many works focusing on the infection efficiency of the virus, and from collected information we only infer that its b is no less than that of SARS in 2003. Therefore, we use $b = 0.6$, same as the value often used for SARS.

4.2.4 Basic Reproductive Number

Here we introduce the concept of basic reproductive number R_0 , which an important concept in the evaluation of infectious diseases. In SEIR model without dynamic birth and death, R_0 is defined by Equation (15):

$$R_0 = \frac{\gamma abP}{\lambda} \quad (15)$$

Where γ , a , b , P , and λ have the same definitions as in this article.

Based on the parameters of our model, we can compute R_0 of 2019-nCoV as following:

$$R_0 = \frac{0.75 \times 0.6 \times P}{1/7 \times P} = 3.15$$

Which is generally consistent with the given values in other works.

The basic reproductive number R_0 is the initial mathematical expectation of infections brought by a single infected individual, which measures the spread ability of virus. As time goes by, the reproductive number R_0 might have different values, and the sufficient and necessary condition for continuous spread of an infectious disease is that it has its $R_0 \geq 1$. If $R_0 < 1$, the virus become extinct eventually. The R_0 of SARS is estimated in the range 2 to 4, but this virus is more contagious and we have better medical conditions, too. Above all, we think our 3.15 would be a reasonable estimate.

4.3 Implementation and Result

4.3.1 Model Validation

First we validate our modified SEIR model on some provinces' statistics. However, we find that the predictions are not exactly consistent with in some provinces, which is reasonably caused by the inaccuracy of population flow proportion between provinces. The uncertainty in the spread process of the disease should also be taken into consideration. For example, one infected person may influence all in his/her carriage and the ratio may deviate from the collected data. Therefore, we modify the population flow into specific provinces in order to make more precise predictions.

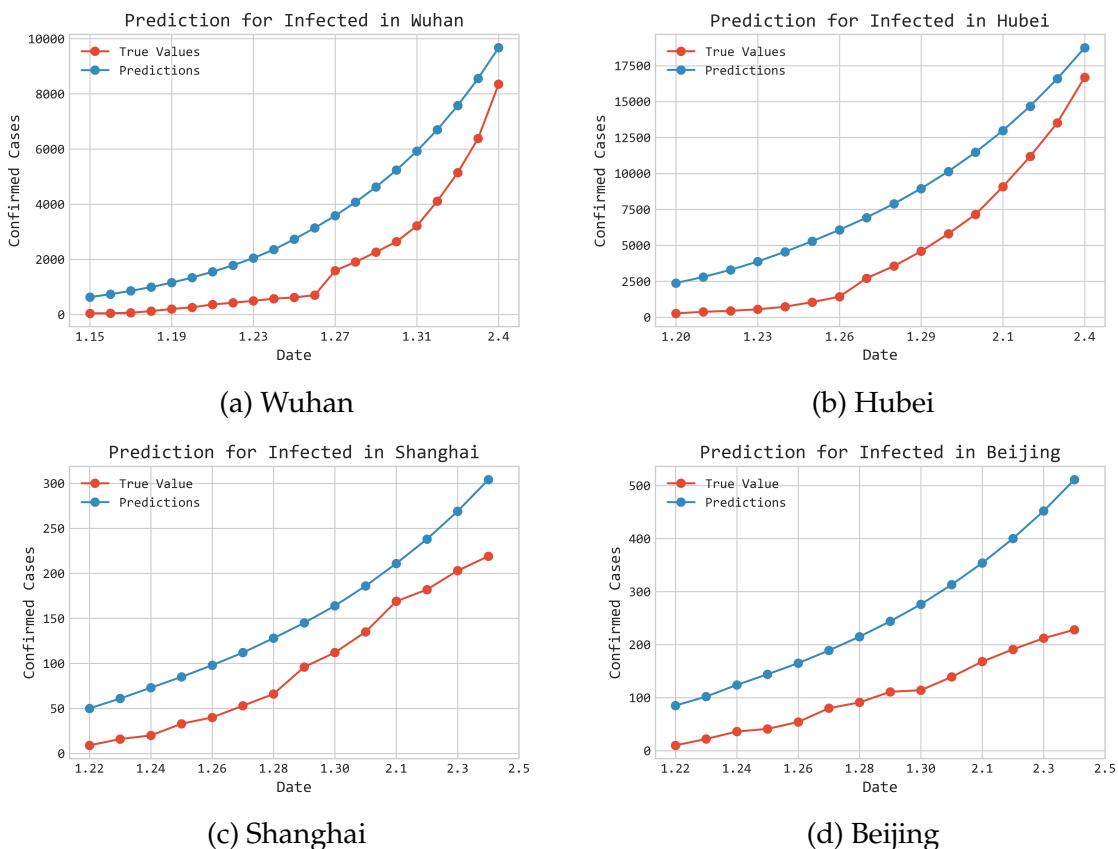


Figure 5: Model Validation on the Datasets of Some Regions

Predictions for Wuhan and several provinces are shown in Figure 5. We point out that we predict the infected cases in Figure 5 and the confirmed case is not equal to the infected case because some infected cases are not confirmed timely. That is to say, it is reasonable

that the predicted number is higher than that of confirmed cases publicly announced. Therefore, it is justified that our model is highly consistent with reality of last two weeks.

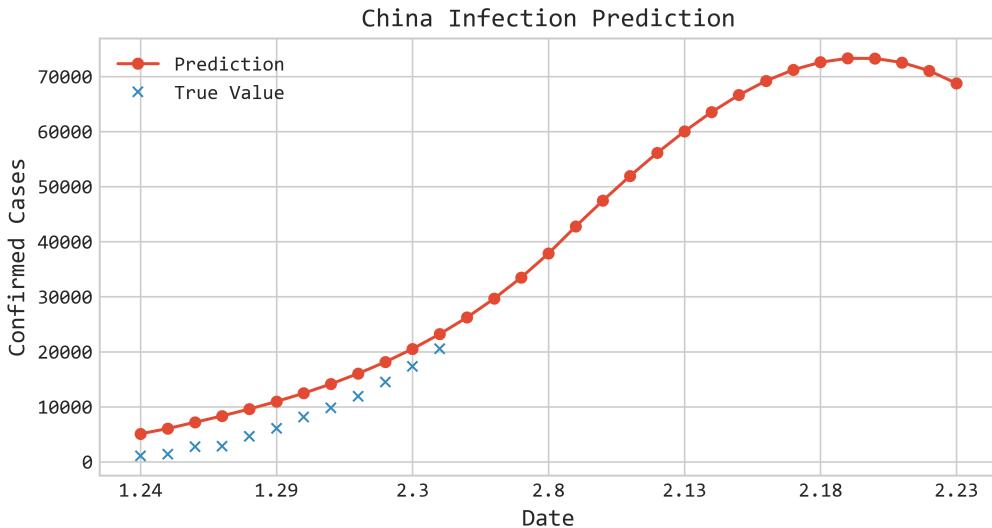


Figure 6: Domestic Infection Prediction

4.3.2 Domestic and Global Predictions

With its effectiveness proved, the modified compartmental SEIR now can be used to predict the domestic and global infections in the next period of time.

Considering domestic infections, we divide Chinese mainland into 31 provinces, and then implement our model using corresponding iteration equations on each of them. Finally we take the summation of all provinces' predicted value as our prediction for domestic infections. Our predictions can be seen in Figure 6.

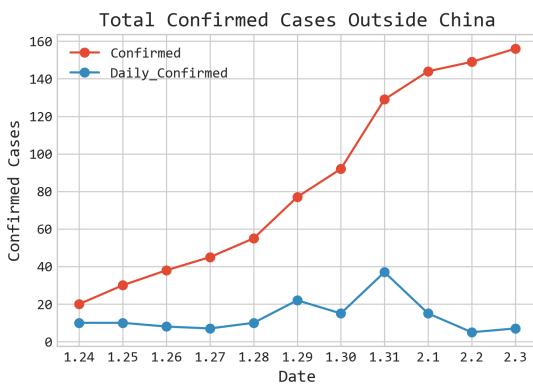


Figure 7: Total Confirmed Cases Outside China

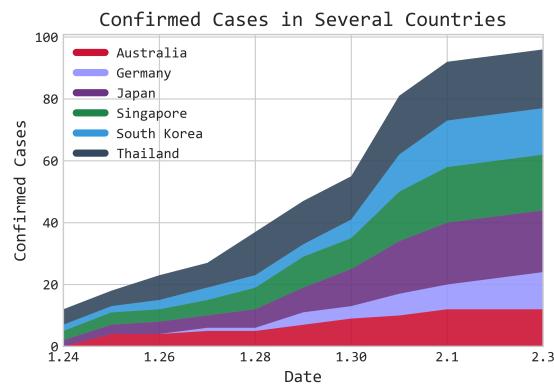


Figure 8: Confirmed Cases in Some Countries

As for global infections, we first analyze the situation from Jan. 24 to Feb. 3 because the influence of 2019-nCoV came to emerge outside China around Jan. 24.

As is shown in Figure 7, the total confirmed cases outside China were on a daily increase of less than 40. Figure 8 indicates confirmed cases in countries suffering relatively seriously from 2019-nCov, mainly attributed to geological reason and international population flow.

Generally speaking, the confirmed cases is under 20 cases in a single country. Therefore, outside China, the outbreak of 2019-nCov is discretely distributed and kept under control.

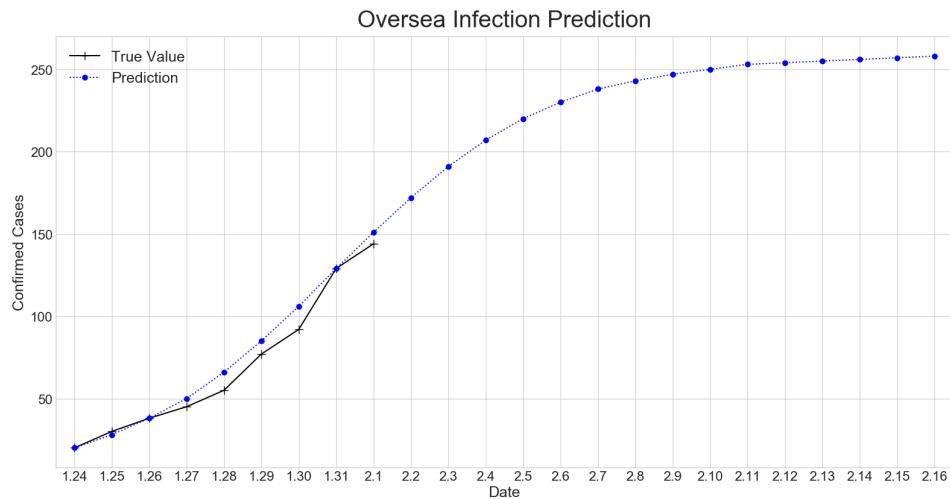


Figure 9: Oversea Infection Prediction

Considering the poor statistics from other countries and the influences brought by potential uncertain factors, we choose the grey prediction model Grey Verhulst to forecast the trend of oversea infections because it can take into account the spread saturability of virus under powerful control, and then obtain the result in Figure 9, which indicates that the number of oversea infections is about to level off in the next two weeks and the main reasons might be early discovery, timely isolation and travel restriction from China.

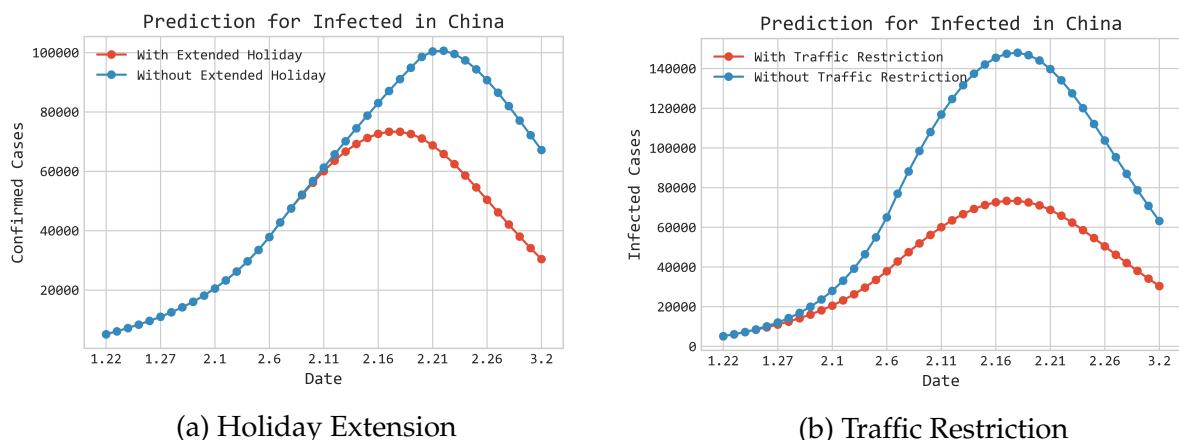


Figure 10: Quantification of Measures' effect

4.3.3 Quantification of Measures' Effect

We have presented our model's prediction in Section 4.3.2 with major control measures taken by the Chinese government, including travel restriction and holiday extension. To analyze the effectiveness of these measures, in this part we will exhibit our predictions without the influence of them as comparison.

As illustrated in Figure 11c, policies like quarantine, travel limitations and holiday extension do blunt the outbreak's impacts. Up to 68% potential infections have been cut according

to our speculation. To be more specific, travel limitations arrest the soaring infections during traditional Chinese new year and the extended holiday put off the massive influx of people in the city which could get the outbreak out of control. Some other measures like temperature check stations and suspension of public transportation play their parts too. But for all these prompt measures, the virus could have claimed far longer lives.

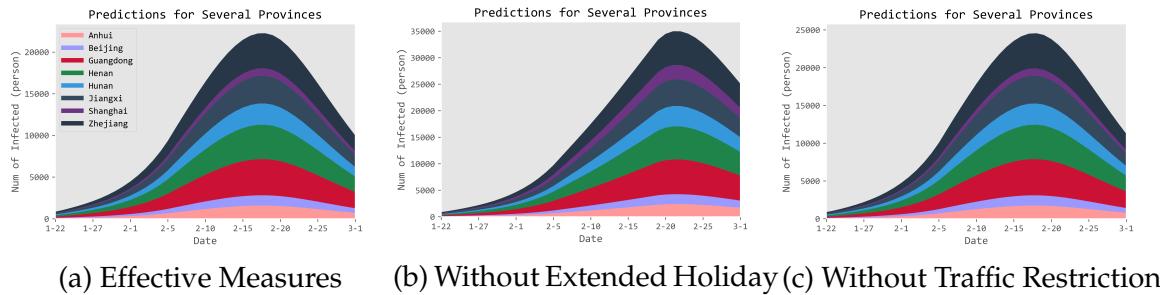


Figure 11: Predictions for Several Provinces in China

Nevertheless, the reduction came with a cost. The lockdown — encircling roughly 50 million people — is unprecedented in scale and experts have questioned its ethics. Wuhan and its home province, Hubei, have borne the brunt of the epidemic as the sudden shutdown of transportation links into and around the area slowed the shipping of vital medical supplies. The scarcity partly contributed to the the difference in death rates of regions. The fatality rate in Wuhan is 4.1 percent and 2.8 percent in Hubei, compared to 0.17 percent elsewhere in mainland China.

Let's discuss the negative impacts in a larger scale. The Spring Festival break used to be a golden week, a boom for tourism and other tertiary industries. The tourism is in deep recession after coronavirus hit the whole country. It sharply slashed the revenue of tourism, over 70 billion dollars based on a report of 2019 by State Tourism Agency. The industry is also in depression. They have to delay resuming work to avoid the risk of getting employees infected. In final analysis, the pathogen put the whole economy under severe strain.

4.4 Emergence of Various Factors

If we put this problem into a bigger picture, many possible factors should be taken into consideration, including presence of secondary infection, discovery of effective treatment method and returning after Spring Festival.

4.4.1 Presence of Secondary Infection

There are two cases for secondary infection: non-permanent immunity and virus mutation. In case of non-permanent immunity, we return the recovered individuals to the susceptible stage immediately after their recovery, and thus our model is to become SEIS. In this case, the iterative equations (6)~(10) should be rewritten as following, ignoring the group of R :

$$\begin{aligned}
S_t^{(j)} &= P_t^{(j)} - E_t^{(j)} - I_t^{(j)} - D_t^{(j)} \\
\Delta E_t^{(j)} &= E_{t+1}^{(j)} - E_t^{(j)} = h S_t^{(j)} + \left(f_{t,i}^{(j)} - f_{t,o}^{(j)} - \theta \right) E_t^{(j)} \\
\Delta I_t^{(j)} &= I_{t+1}^{(j)} - I_t^{(j)} = \theta E_t^{(j)} - \left(r + k + f_{t,o}^{(j)} - f_{t,i}^{(j)} \right) I_t^{(j)} \\
\Delta D_t^{(j)} &= D_{t+1}^{(j)} - D_t^{(j)} = k I_t^{(j)}
\end{aligned}$$

We then implement our model according to the derivation above. The result is shown in Figure 12. As the number of recovered individuals is relatively small compared to the population of whole region, this type of secondary infection does not make a significant difference to our prediction.

In another case, the trend of epidemic situation will turn around immediately. The mutation of a virus means the increase of infecting chance γ to γ' for a susceptible individual to get infected. It at last leads to a rise in basic reproductive number R_0 . We simulate the case in which we increase γ by 0.015 per day since Feb.10 until $\gamma' = 1.0$, and plot the curve in Figure 13 compared with the original one. The consequence of virus mutation is much more serious, as the epidemic situation will be hard to control in a relatively long period of time.

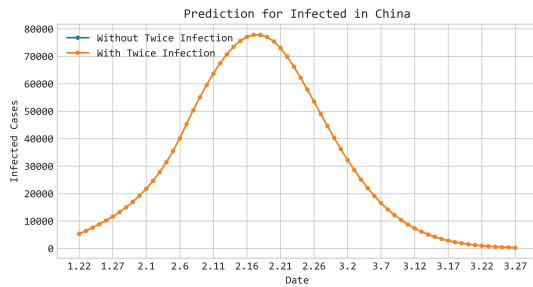


Figure 12: Prediction with Twice Infection

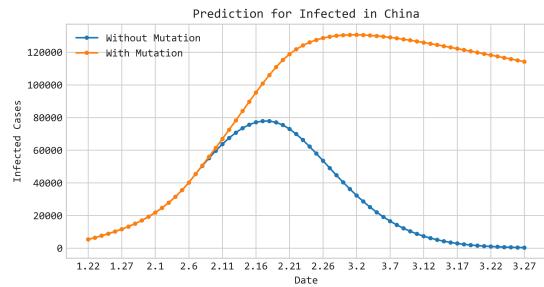


Figure 13: Prediction with Virus Mutation

4.4.2 Discovery of Effective Treatment Method

When an effective treatment method is developed for a certain disease, the major and direct result will be a sharp rise in its recover ratio. Therefore, we assume the recover rate r of 2019-nCoV will increase to 0.05, which means our treating efficiency is ten times as before.

We rerun our model with r raised at different timestamps, and thus obtain a curve showing the influence of new treatment and its timing in Figure 14.

The influence of effective treatment is clear: the earlier it is discovered, the earlier China can control the situation and the less total infections. More importantly, the shortage of test kits is a bigger issue. If we don't have the actual number of infections, then we may make the wrong prediction and underestimate the situation.

4.4.3 Returning after Spring Festival

The traditional transportation in China also involves the large-scale returning after Chinese New Year, which means another peak of population flow between regions. Because of the control measures of Chinese government, the traffic pulse has been buffered. To restore

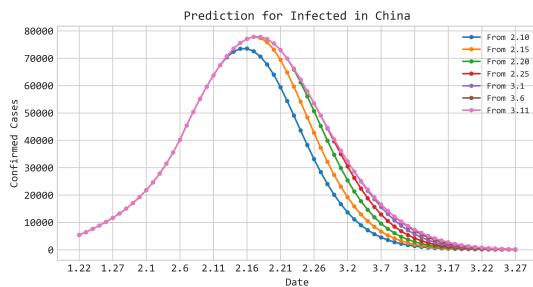


Figure 14: Prediction with New Medicine

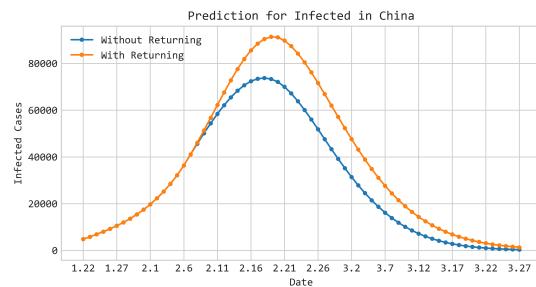


Figure 15: Prediction with Returning to Work

its impact on the spread of epidemic, we let a (thus h) rise a bit after the Spring Festival and then level down to the same value as our original model. Figure 15 exhibits the influence of the returning rush.

From the picture, we can see that the big returning after festival reasonably leads to a higher peak of infections, which also proves the effectiveness of our countries' measures.

To sum up, all signs point to the possibility that WHO will terminate the PHEIC earlier than 3 months routine review.

5 Model Analysis

5.1 Sensitivity Analysis

In Section 4.2.3, we chose values for some parameters in our model according to recent reports and statistics. Here we are going to discuss the sensitivity of them, utilizing the reproductive number R defined in Section 4.2.4. Before detailed discussion, we formalize the sensitivity $S(x)$ of a variable x in this article as following:

Definition Given the peak of infections $I_p(x = x_0)$, the sensitivity of variable x at $x = x_0$ is:

$$S_x(x_0) = \left[\left(\frac{dI_p(x)}{I_p(x)} \right) \Big/ \left(\frac{dx}{x} \right) \right] \Big|_{x=x_0} = \left[\left(\frac{dI_p(x)}{dx} \right) \left(\frac{x}{I_p(x)} \right) \right] \Big|_{x=x_0}$$

Since the peak is only accessible through numeric simulation, we take its difference as its differential, just as what we did with the equations in Section 4.2.2.

5.1.1 The Sensitivity of b and γ

These two parameters are closely related to the reproductive number R , and thus have huge impacts on the peak of infections as shown in Figure 16a and 16b. We also evaluate their sensitivity and impact on the reproductive number R numerically:

$$S_b(b = 0.6) = \frac{134000 - 40000}{0.65 - 0.55} \times \frac{0.6}{74000} \approx 7.62$$

$$S_\gamma(\gamma = 0.75) = \frac{120000 - 45000}{0.8 - 0.7} \times \frac{0.75}{74000} \approx 7.60$$

$$\frac{dR}{db} = \frac{\gamma a P}{\lambda} = 5.25$$

$$\frac{dR}{d\gamma} = \frac{abP}{\lambda} = 4.2$$

Here we can summarize an important inference: The parameters in SEIR model tend to have high sensitivity since tiny changes could lead to a large difference in the reproductive number R . This can be proved on the contrary by looking into the sensitivity of total population P : Similarly we have $\frac{dR}{dP} = \frac{\gamma ab}{\lambda} < 10^{-6}$, so even if we double P to more than 2×10^7 , the infections are quite the same as before.

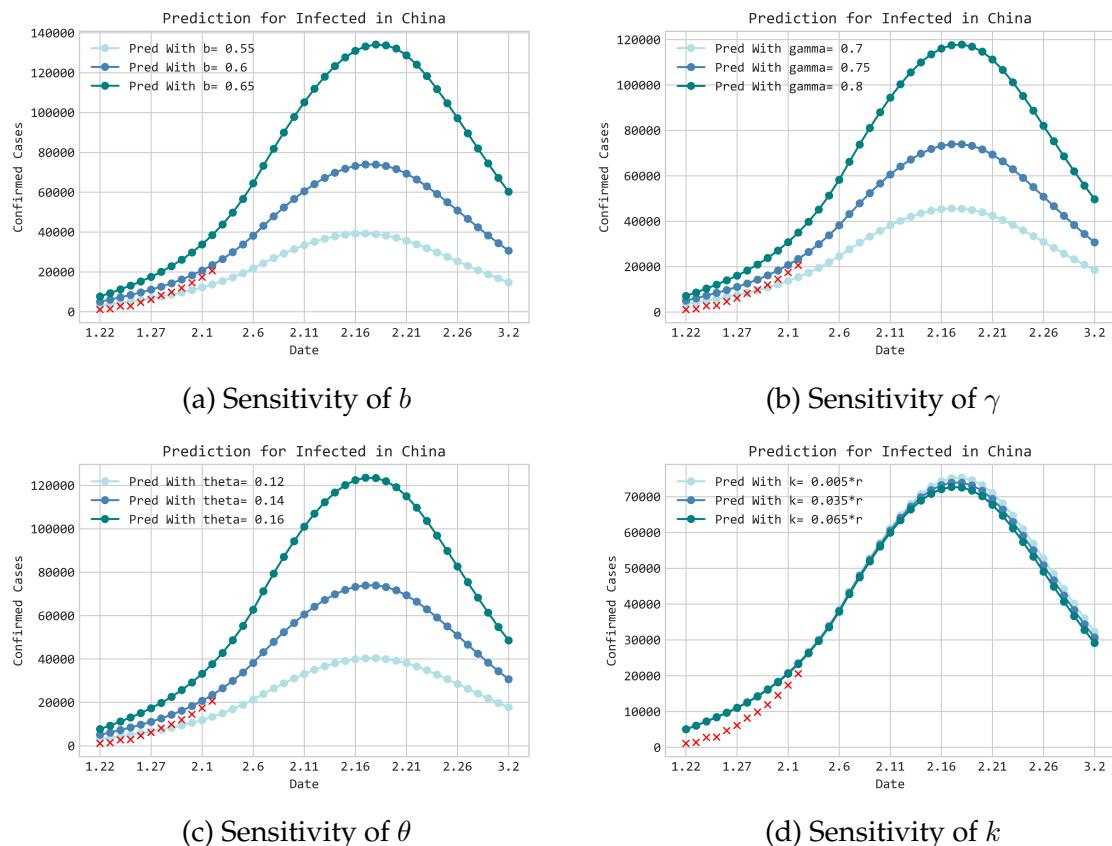


Figure 16: Sensitivity Analysis

5.1.2 The Sensitivity of θ and k

These two parameters do not appear in the expression of R . However, their sensitivities are in different orders of magnitude, as presented in Figure 16c and Figure 16d. Numerically we have:

$$S_\theta(\theta = 0.14) = \frac{123000 - 40000}{0.16 - 0.12} \times \frac{0.14}{74000} \approx 3.93$$

$$S_\gamma(\gamma = 0.75) = \frac{75000 - 72000}{(0.065 - 0.005) \times r} \times \frac{0.035 \times r}{74000} \approx 0.02$$

We think that the reason for the huge difference in the sensitivities of these two parameters might be that increasing θ significantly accelerate the process of E to I , and thus raise the chance of S to E . However, as recovered and dead individuals are both removed from the chain of infection as the immunity is permanent, the rise of k does nearly nothing to our prediction.

5.2 Strengths and Weaknesses

5.2.1 Strengths

1. Collect plenty statistics from official source and make full use of them. Our data is mainly obtained from the Health Commission of a region or country, which has high credibility. With reliable and useful data, we can build, validate and apply our MoC-SEIR model confidently.
2. Take various factors into consideration. To address the disadvantage of classic SEIR, we introduce several factors to our modified model, including population flow, secondary infection and the appearance of effective treatments, and quantify their impacts on the epidemic scenario. In this way, we successfully depict the spread process of 2019-nCoV and give some reasonable predictions to its trend with influences of different factors.
3. Offer high value in application. Based on our analysis on the reproductive number, we propose many trains of thought to help prevent the virus spreading among crowds in different perspectives.
4. Have flexibility and adaptability. Our model can adapt to the realistic situation better than the raw SEIR model and more parameters means a stronger expression ability.

5.2.2 Weaknesses

1. Some real conditions and exceptions are neglected. We simplify our model by only considering the outflow from Wuhan, while population flow between other regions might entangle the model a little bit. Also, the situation in some region has been interfered by accidental factors, which are ignored in our model.
2. Lacking detailed oversea data and analysis. As the outbreak took place suddenly, there are few statistics and survey abroad. Therefore, we use grey prediction to forecast the trend out of China, failing to give concrete analysis.
3. The robustness of our model is not very strong. As a common disadvantage of the models of infectious disease, the simulation result is highly correlated to some parameters, which ultimately changes R_0 , as seen in our modified model.

6 Application

As discussed in Section 4 and Section 5, we find several major factors having enormous impacts on the spread of virus, which provide us many useful trains of thought to protect ourselves and help to control virus spreading. Here we list and explain mathematically some of them.

- **Supply enough test kits.** At the very beginning of the outbreak, there was huge scarcity of test tools which kind of led to potential carriers spreading virus everywhere. What's worse, the shortage could give us underestimation of the current situation and give rise to more fatalities.
- **A 14-days quarantine is still needed in both individual and collective scale.** The basic function of quarantine is to rule out the possibility of being potential carriers. It also cuts population flow and serves as a buffer against the reflow peek.
- **Decrease population density in public and cut population flow.** We have presented the direct impact on the epidemic situation of a as it changes the basic reproductive number R_0 . The reduction of population density and flow leads to a decline in a , which in turn lowers the reproductive number and inhibit the spread of virus. *This might be realized mainly by Chinese government.*
- **Improve personal protection against virus.** This, obviously, is aimed at reducing γ , which has a similar effect as a . As we protect ourselves better, we would have a lower chance of being infected when staying close to infectious individuals. *This might be done mainly by single person.*
- **Encourage the production of medical supplies and the researches on treatment.** Lastly there are ways to make r higher. If infected people are cured more efficiently, then less healthy individuals might be infected. At the same time, the discovery of effective treatment, as shown in Section 4.4.2, can also help a lot to control and improve the situation. *These might be accomplished by Chinese government, related enterprises and scientific research teams together.*

7 Conclusion

In this paper, we look deep into the development and trend of the epidemic situation of 2019-nCoV. We first present the simulation results of classic SEIR and then modify it with extra features taking the effect of population flow brought by traditional transportation and control measures taken by the Chinese government into consideration, which leads to the proposition of our **modified compartmental SEIR model**.

Based on our modified model, we give a prediction of domestic infections in the next period of time, which indicates that the increasing of newly infected individuals is about to stop by the end of February. **Grey Verhulst model** is then introduced to forecast oversea infections as few statistics and factors are known to us, suggesting the spread trend has been stemmed abroad. Next we quantify the effectiveness of Chinese government's measures by changing corresponding features in our model and look into their losses in the economic aspect. We speculate that the virus devastate the tertiary industry most for the Spring Festival is such a boon for the industry. Emergence of two types of powerful factors are also mathematically analyzed. In total, our conclusion is that WHO is expected to terminate the PHEIC in the end of March.

Measures, in either personal or national perspective, are fully discussed, which might reduce the chance of virus spread between individuals. In future works, the subtle influences of expositors and population variation can be explored. We hope our model offers an insight into what's happening in China and assist China CDC with making informed decisions.

8 Blog

Five Key Ways to Combat 2019-nCov

The outbreak of Coronavirus has claimed over 500 lives and the number continues to climb. This time period is traditional Chinese new year when families get together and have a reunion dinner. The disaster pushes China into a economic downturn, industrial production stagnating and stores closed except those providing the basic life support and have something to with medicine.

Nonetheless, the Chinese government's prompt response and well management prevent the situation from getting worse. Here we will analyze some policies based on the result of our model and explain why some actions are recommended.

1. **Cities are increasingly going under lockdown.** On Jan. 23 The first city, Wuhan, declared to lock the city, stop running public transportation and limit the cars on the road, followed by other cities in Hubei. This policy played an indispensable role in preventing a broader outbreak. It also limits the inner spreading in cities by banning festival gatherings. By our calculation, the infectious has been cut up to 68% due to these measures. So we recommend you to stay home and try to avoid go outside.
2. **Raise awareness across the public.** The Internet keeps people updated about the latest information about the virus and government's suggestions. Now almost everyone in the street wears a mask in China. Besides, people outside Hubei make all the way to help this areas plagued by virus. The donations made by civilians actually ease the shortage of supplies in way. Here we advise not attending any gatherings wash hands and keep your house well-ventilated.
3. **The virus is well tracked and early confirm cases are all linked to Wuhan.** This characteristic gives the Health Commission a clue and helps to locate the most potential carriers, those having been to Wuhan. Timely quarantine has been deployed and prevents more infection. If you know someone who has been to Wuhan or even Hubei, we suggest you to report to the community office or call 110 right now. If you just came back from Hubei or have been with guys from Hubei, it is necessary for you to have a 14-days quarantine for the sake of your family health.
4. **The unity and optimism prevails over adversity.** Over past 20 days, the outbreak has seen donations from people across the country to the areas plagued by virus. They send food, medical supplies and so on. Hospitals outside the Hubei dispatches medical team to ease the hospital staff shortage in Hubei. Hopefully, the outbreak has been brought under control and a downturn is expected in the near future.
5. **China has learned its lessons from the fight against SARS.** Now, hospitals are equipped with more advanced facilities and better-trained doctors and have more accurate ways to detect the virus. Most of all, the Internet makes the information reachable for everyone and keeps us updated about latest developments which prevents the social panic.

Now, the fight against the plague has stepped into a crucial stage. Where the plague leads us to is determined by you, me and everyone in this country. Follow the instructions above and we as a nation will make through this devastating disaster.

References

- [1] J. M. Read, J. R. E. Bridgen, D. A. T. Cummings, *et al.*, "Novel coronavirus 2019-ncov: Early estimation of epidemiological parameters and epidemic predictions," *medRxiv*, 2020. DOI: 10.1101/2020.01.23.20018549. [Online]. Available: <https://www.medrxiv.org/content/early/2020/01/28/2020.01.23.20018549>.
- [2] S. Zhao, L. Qianyin, R. Jinjun, *et al.*, "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *bioRxiv*, 2020. DOI: 10.1101/2020.01.23.916395. [Online]. Available: <https://www.biorxiv.org/content/early/2020/01/29/2020.01.23.916395>.
- [3] T. Biao, W. Xia, L. Qian, *et al.*, "Estimation of the transmission risk of 2019-ncov and its implication for public health interventions," *SSRN*, 2020. [Online]. Available: <https://ssrn.com/abstract=3525558>.
- [4] A. Siqi, Z. Guanghu, T. Fei, *et al.*, "Population movement, city closure and spatial transmission of the 2019-ncov infection in china," *medRxiv*, 2020. DOI: 10.1101/2020.02.04.20020339. [Online]. Available: <https://www.medrxiv.org/content/early/2020/02/05/2020.02.04.20020339>.
- [5] W. O. Kermack and A. G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics," *Proceedings of the Royal Society of London Series A*, vol. 115, no. 772, pp. 700–721, Aug. 1927. DOI: 10.1098/rspa.1927.0118.
- [6] globalcitizen, *2019-wuhan-coronavirus-data*. [Online]. Available: <https://github.com/globalcitizen/2019-wuhan-coronavirus-data>.
- [7] S. A. Lauer, K. H. Grantz, Q. Bi, *et al.*, "The incubation period of 2019-ncov from publicly reported confirmed cases: Estimation and application," *medRxiv*, 2020. DOI: 10.1101/2020.02.02.20020016. [Online]. Available: <https://www.medrxiv.org/content/early/2020/02/04/2020.02.02.20020016>.
- [8] J. A. Backer, D. Klinkenberg, and J. Wallinga, "The incubation period of 2019-ncov infections among travellers from wuhan, china," *medRxiv*, 2020. DOI: 10.1101/2020.01.27.20018986. [Online]. Available: <https://www.medrxiv.org/content/early/2020/01/30/2020.01.27.20018986>.
- [9] T. Liu, J. Hu, M. Kang, *et al.*, "Transmission dynamics of 2019 novel coronavirus (2019-ncov)," *bioRxiv*, 2020. DOI: 10.1101/2020.01.25.919787. [Online]. Available: <https://www.biorxiv.org/content/early/2020/01/26/2020.01.25.919787>.

Appendices

Appendix A Grey Verhulst Model

```

syms a b;
c = [a b]';

% Import initial statistics.
A = [20,30,38,45,55,77,92,129,144];
n = length(A);

%Constructing Data Matrices
for i = 2:n
    H(i) = A(i) - A(i - 1);
end
H(1) = [];

for i = 2:n
    C(i) = (A(i) + A(i-1))/2;
end
C(1) = [];

D = [-C; C.^2];
Y = H; Y' = Y';

% Using the least square method to compute the parameters.
c = inv(D*D')*D*Y;
c = c';
a = c(1); b = c(2);

% Make Predictions
F = []; F(1) = A(1);
for i = 2:(n+14)
    F(i) = (a*A(1))/(b*A(1)+(a - b*A(1))*exp(a*(i-1)));
end

disp('Predictions:');
F

% Plot the data points and the fitted curve.
t1 = 0:n-1;
t2 = 0:n+13;
figure(1);
plot(t1, A, 'ro'); hold on;
plot(t2, F, 'b--');
xlabel('Timestamp'); ylabel('Confirmed Cases');
legend('True Value','Prediction','location','northwest');
title('Oversea Infection Prediction'); grid;

```

Appendix B Data Preprocessing

```

# -*- coding: utf-8 -*-

#Dict: P,I,D,R
#Col: Time Series. Dec 1st, 2019 -->

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

plt.style.use('seaborn-whitegrid')
plt.rcParams['font.sans-serif'] = 'Consolas'
plt.rcParams['axes.unicode_minus'] = True
plt.rcParams['savefig.dpi'] = 600

time_stamp={}
Date_List=['0122','0123','0124','0125','0126','0127','0128','0129','0130',
           '0131','0201','0202','0203','0204']

for i,date in enumerate(Date_List):
    file='./Data/Province_Data/'+date+'.csv'
    data=pd.read_csv(file,sep='|',error_bad_lines=False)
    data.drop(['deaths','notes','sources'],axis=1,inplace=True)

```

```

for j,place in enumerate(data.place):
    if time_stamp.get(place) ==None :
        time_stamp[place]=[]
        for k in range(i):
            time_stamp[place].append(0)
        time_stamp[place].append(data.confirmed_cases[j])

China_Total=np.array(time_stamp.pop('CHINA TOTAL'))
Province=np.array(list(time_stamp))
Confirmed=np.array(list(time_stamp.values()))

data=pd.read_csv('./Data/Province_Basic_Data.csv',error_bad_lines=False)
Population=np.array(data.Population)
FlowRatio=np.array(data.FlowRatio,dtype=np.float32)
Cases=np.array(data.Confirmed_Cases,dtype=np.float32)

fig=plt.figure()
for i in range(34):
    if FlowRatio[i]>20:
        continue
    plt.plot(FlowRatio[i],Cases[i],'o',markersize=8)
plt.xlabel('FlowRatio(%)')
plt.ylabel('Confirmed Cases')
plt.title('Correlation Between Cases and Inflow')
fig.savefig('./figure/Relation_Ana.png')

```

Appendix C SEIR Model

```

# -*- coding: utf-8 -*-

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import generate_province_data as prov
import Prediction_China_No_Holiday as No_Holiday
import Prediction_China_Mutation as Mutation
import Prediction_China_Twice_Infection as Twice_Infection
import Prediction_China_No_GeoRestriction as No_GeoRestriction
import Prediction_Wuhan_Naive as Wuhan_Naive
import Prediction_China_Returning as Returning

plt.style.use('seaborn-whitegrid')
plt.rcParams['font.sans-serif'] = 'Consolas'
plt.rcParams['axes.unicode_minus'] = True
plt.rcParams['savefig.dpi'] = 600
plt.rcParams['figure.figsize'] = (6.0, 4.0)

Date_List=['0115','0116','0117','0118','0119','0120','0121','0122','0123',
           '0124','0125','0126','0127','0128','0129','0130','0131','0201',
           '0202','0203','0204']

#parameter
m=1
n=1
thita=0.14
P=11000000.
b=0.6

# Data For Wuhan
file='./Data/Wuhan_Data.csv'
Wuhan_data=pd.read_csv(file,error_bad_lines=False) #1.15-2.4

#Row: 0--S, 1--E, 2--I, 3--R, 4--D
#Total_person=P=1100w, P=S+E+I+R+D
#Col: Time Series. Dec 1st, 2019 -->

# description about population flow, travel restriction considered
data=pd.read_csv('./Data/Flow_Rate_Data.csv',error_bad_lines=False)
Flow_rate=data.Flow_rate
Wuhan=np.zeros([5,120],dtype=np.float32)
Wuhan[:, 0]=[P, 0., 1., 0., 0.]
for i in range(119):
    Wuhan[0,i+1]=P-Wuhan[1,i]-Wuhan[2,i]-Wuhan[3,i]-Wuhan[4,i]
    if i >= 67 and i <= 80:
        b -= 0.04      # Infectious intensity decreases as temperature goes up
h=b*0.75*Wuhan[2,i]/P

```

```

if i >= 80:
    h *= 0.08
elif i >= 67:
    h *= 0.36 # Most exposers show symptoms and get isolated
elif i >= 53:
    h *= 0.6 # Travelling restriction
elif i >= 31:
    h *= 0.98 # Notice publicated

if i <=31:
    r = 0.005
elif i <=67 :
    r = 0.01
else:
    r+=0.005

k=0.035*r

Wuhan[1,i+1]=Wuhan[1,i]+h*Wuhan[0,i]-thita*Wuhan[1,i]
-m*Wuhan[1,i]*Flow_rate[i]
Wuhan[2,i+1]=Wuhan[2,i]+thita*Wuhan[1,i]-r*Wuhan[2,i]-k*Wuhan[2,i]
-n*Wuhan[2,i]*Flow_rate[i]
Wuhan[3,i+1]=Wuhan[3,i]+r*Wuhan[2,i]
Wuhan[4,i+1]=Wuhan[4,i]+k*Wuhan[2,i]
Wuhan=Wuhan.astype(np.int)

#Data for Hubei
p=59000000-P
alpha=0.67
m=1
n=1
thita=0.14
P=11000000.
b=0.6

Hubei=np.zeros([5,120],dtype=np.float32)
Hubei[:, 0]=[p, 0., 0., 0., 0.]
for i in range(119):
    Hubei[0,i+1]=p-Hubei[1,i]-Hubei[2,i]-Hubei[3,i]-Hubei[4,i]
    if i >= 67 and i <= 80 :
        b -= 0.04

    h=0.75*b*Hubei[2,i]/p

    if i >= 80:
        h *= 0.08
    elif i >= 67:
        h *= 0.36
    elif i >= 53:
        h *= 0.6
    elif i >= 31:
        h *= 0.98

    if i <=31:
        r = 0.005
    elif i <=67 :
        r = 0.01
    else:
        r+=0.005

    k=0.035*r

    Hubei[1,i+1]=Hubei[1,i]+ h*Hubei[0,i]- thita*Hubei[1,i]
    + alpha*m*Wuhan[1,i]*Flow_rate[i]
    Hubei[2,i+1]=Hubei[2,i]+ thita*Hubei[1,i]- r*Hubei[2,i]- k*Hubei[2,i]
    + alpha*n*Wuhan[2,i]*Flow_rate[i]
    Hubei[3,i+1]=Hubei[3,i]+r*Hubei[2,i]
    Hubei[4,i+1]=Hubei[4,i]+k*Hubei[2,i]

Hubei=Hubei.astype(np.int)
Hubei[:, :]+=Wuhan[:, :]

file='./Data/Hubei_Data.csv'
Hubei_data=pd.read_csv(file,error_bad_lines=False) #1.20-2.4

#Other Province, Hubei excluded
China_Total_Pre=np.zeros([5,120],dtype=np.float32)
China_Total_Pre=Hubei
ShowProvince=['Beijing','Shanghai','Guangdong','Zhejiang','Henan','Hunan',
'Jiangxi','Anhui']
# Adjust the population flow
High_Province=['Gansu','Qinghai','Hainan','Shanghai','Yunnan','Tianjin']
Far_High_Province=['Shanghai']

```

```

Low_Province=['Sichuan','Jiangxi','Guangdong','Xinjiang','Hunan','Beijing',
              'Anhui','Shandong','Guangxi','Jiangsu','Hebei','Zhejiang',
              'Heilongjiang','Sichuan','Jiangxi','Fujian']
Far_Low_Province=['Sichuan','Jiangxi','Guangdong','Xinjiang','Zhejiang',
                   'Fujian','Anhui']
for j in range(34):
    Province=prov.Province[j]
    if(prov.Province[j]=='Hubei' or prov.Province[j]=='Hong Kong'
       or prov.Province[j]=='Macau' or prov.Province[j]=='Taiwan'
       or prov.Province[j]=='Tibet'):
        continue
    p = prov.Population[j] * 10000
    alpha=prov.FlowRatio[j]/100
    if Province=='Inner Mongolia':
        continue
    if Province=='Guangxi':
        alpha=0.85/100.
    if Province=='Hunan':
        alpha=3.52/100.
    if Province in High_Province:
        alpha/=3
    if Province in Far_High_Province:
        alpha/=2
    if Province in Low_Province:
        alpha*=3
    if Province in Far_Low_Province:
        alpha*=2
    if Province =='Fujian' or Province=='Jiangxi' or Province=='Shaanxi':
        alpha*=2
m=1
n=1
thita=0.14
b=0.6
Place=np.zeros([5,120],dtype=np.float32)
Place[:, 0]=[p, 0., 0., 0., 0.]
for i in range(119):
    Place[0,i+1]=p-Place[1,i]-Place[2,i]-Place[3,i]-Place[4,i]
    if i >= 67 and i <= 80:
        b -= 0.04
    h=0.75*b*Place[2,i]/p
    if i >= 80:
        h *= 0.08
    elif i >= 67:
        h *= 0.36
        h *= 0.6
    elif i >= 31:
        h *= 0.98
    if i <=31:
        r = 0.005
    elif i <=67 :
        r = 0.01
    else:
        r+=0.005
    k=0.035*r
    Place[1,i+1]=Place[1,i]+ h*Place[0,i]- thita*Place[1,i]
                + alpha*m*Wuhan[1,i]*Flow_rate[i]
    Place[2,i+1]=Place[2,i]+ thita*Place[1,i]- r*Place[2,i]- k*Place[2,i]
                + alpha*n*Wuhan[2,i]*Flow_rate[i]
    Place[3,i+1]=Place[3,i]+r*Place[2,i]
    Place[4,i+1]=Place[4,i]+k*Place[2,i]
Place=Place.astype(np.int)
China_Total_Pre+=Place

Date_List_tmp=['1.22','1.27','2.1','2.6','2.11','2.16','2.21','2.26','3.2']
fig = plt.figure(40)
plt.plot(range(len(prov.China_Total)), prov.China_Total, 'o-',
          label='True Values')
plt.plot(range(len(prov.China_Total)), China_Total_Pre[2,52:66], 'o-',
          label='Predictions') #1.22->2.4
plt.xticks(range(0,15,2),Date_List_tmp)
plt.xlabel('Date')
plt.ylabel('Confirmed Cases')
plt.legend(loc = 'best')
plt.title('Prediction for Infected in China')
plt.savefig('./figure/China_Validation.png')

China_Total_No_Holiday=No_Holiday.China_Total_Pre
Date_List_tmp=['1.22','1.27','2.1','2.6','2.11','2.16','2.21','2.26','3.2']

```

```
fig = plt.figure(41)
plt.plot(range(41), China_Total_Pre[2,52:93], 'o-',
          label='With Extended Holiday', markersize=5) #1.22->2.20
plt.plot(range(41), China_Total_No_Holiday[2,52:93], 'o-',
          label='Without Extended Holiday', markersize=5)
plt.xticks(range(0,41,5),Date_List_tmp)
plt.xlabel('Date')
plt.ylabel('Confirmed Cases')
plt.legend(loc = 'upper left')
plt.title('Prediction for Infected in China')
plt.savefig('./figure/China_Total_Holiday_Comp.png')
```
