Analysis on the Development Trend of COVID-19 Outbreak in Beijing Based on the Cluster Analysis and SEIR Model

Mingming Jiang[1], Longhui Zu*[2],Zhaolin Li[1] Zhihuan Zhao[1],Yingyu Pan[1],Weili Liu[1],Shuolin Yang[1]

[1]Mechanical& electronic Engineering College, Shandong AgricultureAnd Engineering University,Jinan,China

[2]Shandong Academy of Chinese Medicine,Jinan,China

851815697@qq.com

2020 Chinese Automation Congress (CAC) | 978-1-7281-7687-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/CAC51589.2020.9327560

Abstract---**[Purposes] Through the cluster analysis and the SEIR Model established, to study the trend of COVID-19 outbreak in all the regions of Beijing and analyze the prevention and control effects to provide basis for its prevention and control. [Methods] We collected the data related to COVID-19 outbreak in Beijing from January 21 to March 31, 2020, used SPSS19.0 and MetaboAnalysis software to conduct a cluster analysis on the outbreak data of the cases in all the regions of Beijing, the cases from others places to Beijing and the imported cases and built a SEIR Model through Matlab to simulate the development trend of COVID-19 outbreak in Beijing. [Results] According to the cluster analysis, the high incidence was mainly distributed in Chaoyang District, Haidian District, Changping District, Fengtai District, Daxing District, Xicheng District and among the people from other places to Beijing; according to the SEIR Model, in mid-to-late February, the inflection point of the local outbreak appeared but there were still the enormous risks of imported cases and the prevention and control was still in a severe situation. [Conclusion] At present, the outbreak situation has tended to be gentle but the comprehensive and strict prevention and control shall still be strengthened to strictly prevent the imported cases and resolutely curb the spread of the outbreak.**

Key words---Novel coronavirus; Cluster analysis; SEIR Model; Imported cases; Outbreak prevention and control

I. INTRODUCTION

In December 2019, a novel coronavirus caused an outbreak of pneumonia in Wuhan City, Hubei Province, which rapidly spread across the country. WHO named the novel coronavirus-infected pneumonia as corona virus disease 2019 (COVID-19) on February 11, 2020 [1]. As of 24:00 on April 19, 2020, 82,747 accumulatively confirmed cases had been reported nationwide, of which 77,084 accumulative cases had been cured and discharged from hospital, and there had been 1,583 imported accumulatively confirmed cases, of which 742 accumulative cases had been healed and discharged from hospital [2]. As of April 19, 2020, there had been the confirmed cases in 209 countries and regions overseas, and there had been 2,240,735 accumulatively confirmed cases [3]. The outbreak in China has been basically controlled, and the situation has continuously tended to a good prospect but the foreign countries are in a rapidly-rising outbreak period so the risks of imported cases have greatly increased, and it is undoubtedly a major hidden danger for China.

After the outbreak of the epidemic, many scholars at home and abroad carried out mathematical modeling and epidemiological studies on covid-19, and provided the basic epidemiological parameters such as the basic regeneration number, peak time, epidemic scale, and evaluated the probability distribution model[4], SIS model, SIR model[5] and SEIR model[6] have been established to evaluate and predict the epidemic situation, but most of the studies have not included the influence of the lurkers on the epidemic situation. In this paper, we add the latent state on the basis of SIR model, and the lurk has a certain probability to infect the susceptible, and has a certain probability to transform into an infected person.

After January 21, 2020, the outbreak in Beijing tended to be serious, plus the imported confirmed cases were found in Beijing on February 29, in order to seek the spread rules for the outbreak and to provide the basis for the prevention and control, a SEIR transmission dynamics model was established in the paper to conduct a fitting study on the data of COVID-19 outbreak in Beijing from January 21 to March 31, 2020 and conduct the cluster analysis on the cases in all the regions of Beijing, the cases from other places to Beijing and the imported cases so as to provide the theoretical support for the reasonable prevention and provide the basis for scientifically and effectively preventing and controlling the COVID-19.

II. DATA AND METHODS

A. Data sources

The data in the paper come from the official statistical data of Beijing as of March 31, 2020, and all the data can be obtained from the official website of Beijing Municipal Health Commission[7]. According to the official report, early in the morning of January 20, Beijing announced two confirmed COVID-19 cases; therefore, in the paper, January 20, 2020 is defined as the 0^{th} day of the COVID-19 outbreak in Beijing. For the data analysis, the software Microsoft Excel, SPSS19.0, MetaboAnalysis and Matlab 2012a were used.

B. Establishment of the SEIR model

SEIR dynamics model[8][9] is as shown in Figure 1, and it classifies the research objects as S, E, I and R: S – susceptible, represents the potentially-infected people; E – exposed, represents the people who have been infected but have still not show the symptoms of infection; I – infected, represents the people who show the symptoms of infection and can infect the susceptible population contacting with them in the certain probability; R- removed, represents the people who will no longer be influenced by the infectious disease (the people who are cured, die or are effectively quarantined).

S , E_t , I_t and R_t are the number of the susceptible, the exposed, the infected and the removed at the moment of t

t respectively and meeting the formula $N = E_t + I_t + R_t + S_t$, where N is the total number.

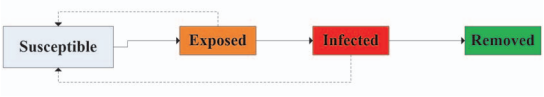

Figure 1 SEIR dynamic model

The differential equations [9] [10]are as follows:

$$S_t = S_{t-1} - r\beta I_{t-1}\frac{S_{t-1}}{N} - r_1\beta_2 E_{t-1}\frac{S_{t-1}}{N}$$

$$E_t = E_{t-1} - \alpha E_{t-1} + r\beta I_{t-1}\frac{S_{t-1}}{N} + r_1\beta_2 E_{t-1}\frac{S_{t-1}}{N},$$

$$I_t = I_{t-1} + \alpha E_{t-1} - \gamma I_{t-1}$$

$$R_t = R_{t-1} + \gamma I_{t-1},$$

t is the number of spread days of COVID-19; r represents the number of the people in contact with the infected; r_1 represents the number of the people with whom the exposed contact every day; β represents the probability that the susceptible are infected after the contract; β_2 represents the probability that the exposed infect the susceptible; α represents the probability that the exposed turn to the infected; γ represents the recovery probability of the infected; S_t/N represents the proportion of the susceptible at the time of r.

C. Parameter determination

SEIR Model is highly sensitive to the parameter settings so the unreasonable parameter settings will generate the greater errors to the model prediction [8]. On the base that this virus has an incubation period and has the certain infectivity during the incubation period and National Health Commission said that the average incubation period of the virus was around 7 days and that the longest incubation period could reach 14 days so in the paper, the incubation period is assumed as 12 days, and according to the research of Reference [8], α may be set as the reciprocal of the incubation period $\alpha = 1/12 = 0.083$. It is assumed that the number of the susceptible with whom the infected contact is the same as that of the susceptible with whom the exposed contact, that is, $r = r_1 = 6$, and then $\beta = \beta_2 = 0.048$ and the recovery rate $\gamma = 0.10$ will be obtained by calculation according to the national data announced by National Health Commission and Reference [11]. The total population of Beijing is 21.536 million according to Baidu Baike.

III.RESULTS

A. Development characteristics of the outbreak

Figure 2 shows the variation trends of the number of accumulatively confirmed cases and the number of daily new cases in Beijing, with two obvious characteristics: first, it can be seen from the curve of the number of daily new cases that from the first report of confirmed cases on January 20 to February 2, the number of daily new cases was on the rise and reached the maximum value 29 on February 2, and after that (except March 23), the number of daily new cases gradually declined; thus, it is speculated that the outbreak had reached the inflection point of growth on February 2, indicating that the situation took a turn for the better and that the anti-epidemic measures were effective. At the same time, it can be seen from the curve of the number of accumulatively confirmed cases that it developed rapidly before February 12 and the growth slowed down and approached to a smooth and steady state from February 12 to March 18. Second, there were significant fluctuations in the number of daily new cases from February 29 when there were new imported confirmed cases, and it reached the maximum value 32 by March 23, including 31 imported cases and 1 case associated with the imported cases, and later, the curve of the number of accumulatively confirmed cases presented a trend of slow growth again, indicating that the outbreak overseas accelerated spreading and Beijing still continuously had the risks of imported cases.

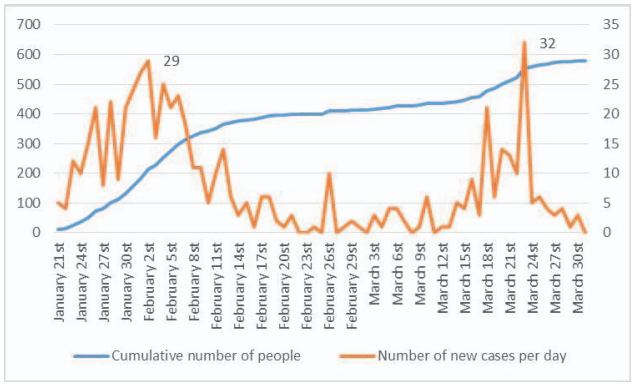

Figure 2 Cumulative number of confirmed cases and newly confirmed cases of COVID-19s

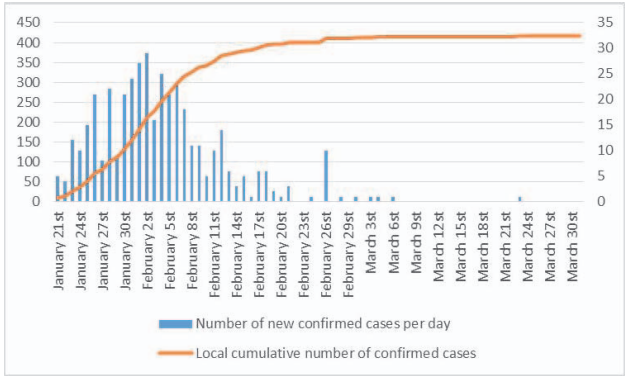

Figure 3 Cumulative number of confirmed cases and newly confirmed cases of COVID-19 in Beijing itself

Figure 3 shows the variation trends of the daily new cases and accumulatively confirmed COVID-19 cases except the imported cases in Beijing. It can be seen from the curves of the daily new cases and the accumulatively confirmed cases that the outbreak developed relatively quickly before February 12, that from February 13 to March 5, its growth slowed down and that it was in a smooth and steady state after March 5, which complies with the results predicted through the SEIR model, indicating that the outbreak in Beijing was gradually taking a turn for the better.

Figure 4 shows the variation trends of the daily new important COVID-19 cases and the accumulatively confirmed COVID-19 cases in Beijing from February 29 to March 31, 2020. From the first report of imported confirmed cases on February 29, 2020, the number of daily new confirmed cases rose in a wave type and after it reached the maximum value 31 on March 23, it declined in a wave type, indicating that Beijing still faced the great pressure of imported cases and that the imported cases still brought the greatest risks.

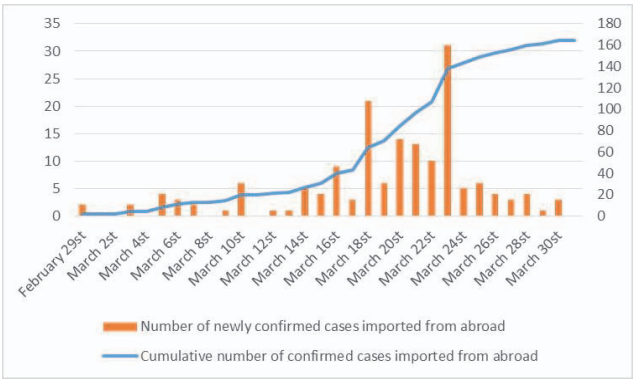

Figure 4 Cumulative number of confirmed cases and newly confirmed cases imported from abroad Cluster analysis

Cluster analysis is a mathematical statistics method to classify the observation objects according some quantitative characteristics, with the basic idea of clustering the regions where the outbreak variation trends are similar. In the paper, the cluster analysis is adopted to divide the cases in all the regions of Beijing, cases from other places to Beijing and the imported cases into different categories to comprehensively reflect and evaluate the outbreak situation in Beijing.

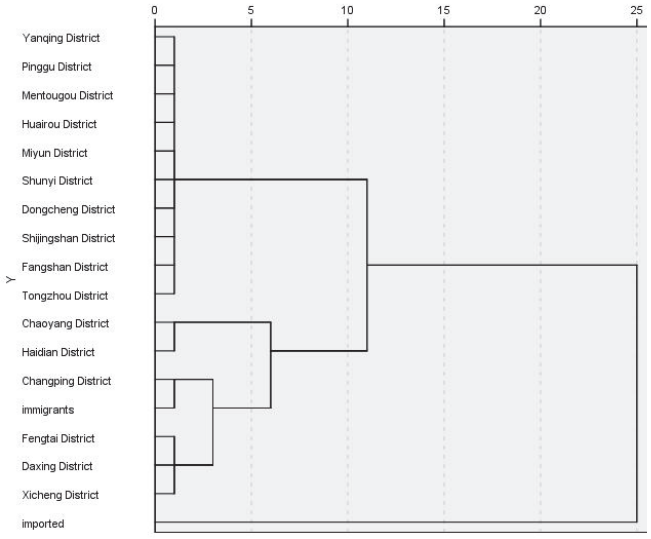

Figure 5 Regional clustering results of cumulative confirmed cases

The data of samples obtained (the variation trends of the cases in all regions of Beijing, the cases from other places to Beijing and the imported cases over time) were imported to SPSS19.0 software for the Q-type cluster analysis [12], and for the clustering result, see Figure 5 and Figure 6. According to the result, when Euclidean distance was 15, the samples were divided into two categories including the imported cases and the cases in other regions, and when Euclidean distance was 10, except the imported cases, the samples were divided into two categories, including the category of the cases in Chaoyang District, Haidian District, Changping District, Fengtai District, Daxing District and Xicheng Distrct and the cases from other places to Beijing, where there were more than 25 confirmed cases, indicating that in such regions in a serious situation and with high incidence, the outbreak prevention and control work was more different than that in other regions, and the category of the cases in other regions where there were less than 20 confirmed cases. According to the results of cluster analysis in Fig. 5, Xicheng District, Chaoyang District, Haidian District, Fengtai District, foreign coming to Beijing, Changping District, Daxing District and overseas input in the lower half of the above figure are defined as group A, which are marked as A1, A2,..., respectively Dongcheng District,Tongzhou District, Shijingshan District, Fangshan District, Shunyi District, Huairou District, Miyun District, Mentougou District, Yanqing District and Pinggu District in the upper half of the region are defined as group B, which are marked as B1, B2, …,B10. The two-way hierarchical clustering analysis of time and region is carried out by using metaboanalysis software, and the clustering heat map is used for more intuitive representation.The horizontal axis direction represents different regions, and the "column" corresponding to each area represents the change degree of the epidemic situation in the region with time; the vertical axis direction represents the outbreak days, and the change of "line" corresponding to a time point represents the difference degree of epidemic situation in different regions; the top and left broken line clusters represent the clustering process [13].

The heat map of Figure 6 shows that the epidemic situation can be divided into two categories according to the severity of the epidemic situation. The color of group A on the left side (red) is significantly darker than that of the area in group B (blue), indicating that the epidemic situation in group A is more serious than that in area B, which is consistent with the cluster analysis in Figure 5. The number of cases in Yanqing District (B9) and Pinggu District (B10) were 1 and 0, which showed that the epidemic situation was the lightest, while the color of imported cases (A8) gradually deepened 60 days after the outbreak, indicating the process of the number of cases from scratch.

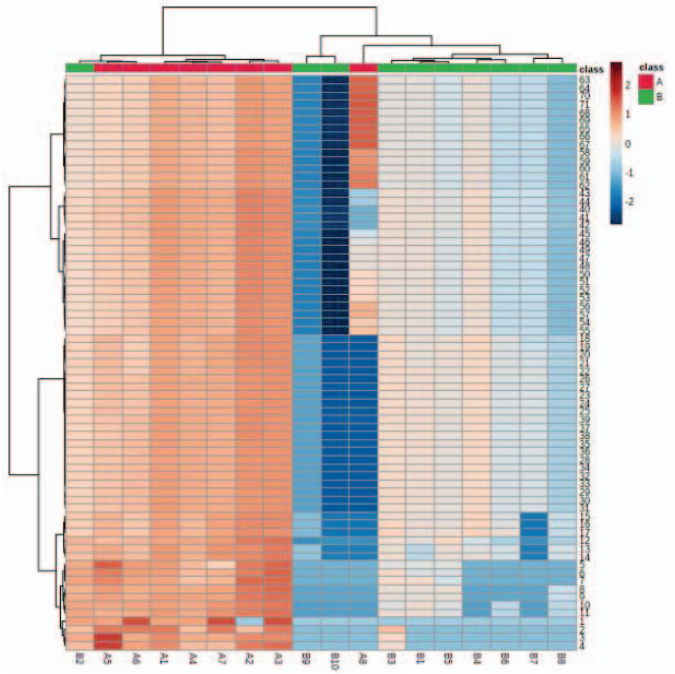

Figure 6 Heat map of regional clustering results of cumulative confirmed cases

B. Dynamics model analysis on the outbreak in Beijing based on the SEIR model

After the relevant parameters were determined according to the differential equations in Section 1.2, the SEIR model for COVID-19 in Beijing was established by Matlab programming, and it is as shown in Figure 7. It can be seen from the figure above that the peak rate of the exposed would come first, which was around one week earlier than the peak value of the infected, and the inflection point of outbreak had come. Around the 40th day of the outbreak (February 28), the

number of the infected would reach the peak value of 623, and after that the number of cases would be gradually reduced, and around the 90th day, the outbreak would be basically stable until it disappeared. It was relatively consistent with the data "On February 12, except the imported cases, there have been 415 accumulatively confirmed cases, and there have been 295 present confirmed cases" reported on the official website of Beijing Municipal Health Commission and the situation that there were around 415 accumulatively confirmed cases stably after that and that the present confirmed cases kept reducing. However, the predicted peak value of the infected in Beijing shown in the model was higher than that of the actual infected, and the peak value appeared slightly later than the actual time possibly because the infected and the exposed of COVID-19 in Beijing contacted with more susceptible persons than that assumed in the paper and the probability that the exposed turned to the infected was lower than the probability assumed in the paper. Therefore, during the outbreak, the prevention and control measures should still be strengthened, and the personnel flow should still be strictly controlled.

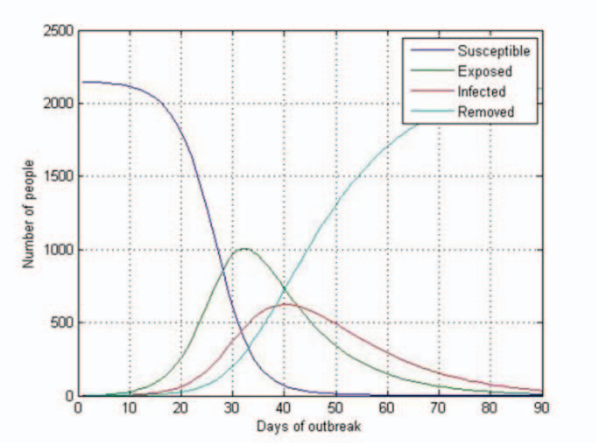

Figure 7 SEIR model for COVID-19

1. Parameter sensitivity analysis

In order to evaluate the influence of government control and travel restriction on the model, the paper analyzes the relevant parameters of the model. Assuming that the government's control is weakened, the number of infected persons contacting susceptible persons and the number of latent persons contacting susceptible persons every day increase, both of which are set to $r=r_1=10$, then the model prediction is shown in Figure 8. The epidemic scale will reach a peak of 707 cases on the 30th day, which is 13.3% higher than the expected actual scale and 10 days ahead of the expected actual time. The above results show that the government can effectively control the spread of the epidemic by strengthening control and restricting travel to reduce the number of contacts.

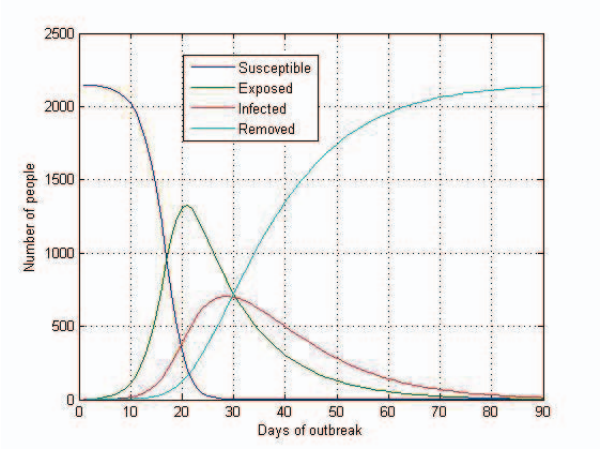

Figure 8 SEIR model of COVID-2019 at $r=r_1=10$

Based on the fact that the virus has a latent period, and the incubation period is infectious to a certain extent, It is necessary to explore the influence of incubation period on the model. According to the research in literature [14], the average incubation period is 5.2 days. Therefore, this paper assumes that the incubation period is 5 days, and the probability of a latent person turning into an infected person is $\alpha=1/5=0.2$. As shown in Figure 9, the epidemic scale will reach a peak of 763 cases on the 37th day, which is 22.3% higher than the predicted actual scale and 3 days ahead of the expected actual time. The results showed that the incubation period had a certain influence on the epidemic scale.

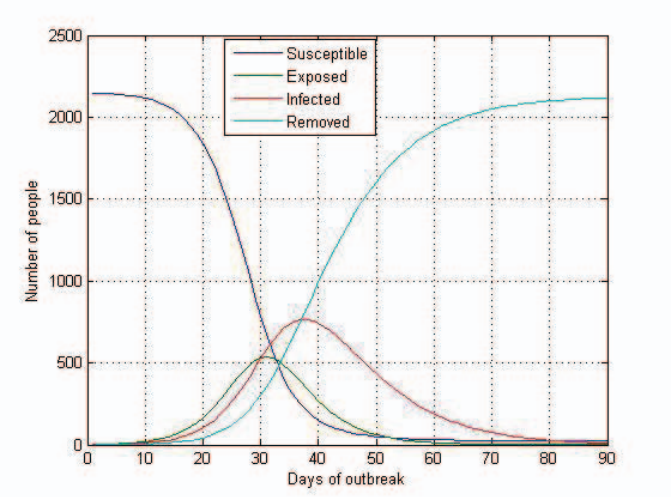

Figure 9 SEIR model of COVID-2019 at $\alpha=0.2$

III. DISCUSSIONS

In the study, the cluster analysis was firstly conducted on the confirmed cases in all regions of Beijing, the confirmed cases from other places to Beijing and the imported confirmed cases based on SPSS19.0, MetaboAnalysis software, and then Matlab software was used to build a SEIR model with the infectivity during the incubation period to calculate the key parameters required by the data calculation model released by National Health Commission to predict the COVID-19 outbreak data in Beijing this time and simulate the outbreak development trend. According to the model, the peak value of the exposed this outbreak would come around one week earlier than the peak value of the infected, and the inflection point would appear in mid-to-late February. However, the model would differ from the reality inevitably. In the outbreak process, "super spreaders" and the situation that the individual

rehabilitees became positive again have appeared but the model cannot estimate the impact of “super spreaders” and “the rehabilitees become positive again” on the outbreak[15]. We hope that the present model can be improved in the future research process to solve the impact of “super spreaders” and “rehabilitees become positive again” on the model.

COVID-19 is more infectious than SARS and MERS [16]. In the context of spread acceleration worldwide, Beijing, like other domestic regions such as Guangzhou and Shanghai, also faces the enormous risks of imported cases [17][18]. Since March, the COVID-19 outbreak in Beijing has shifted from new local cases to the mainly imported cases so the prevision and control has still been in a severe situation and the prevention and control shall be made the normalized work.

REFERENCES

[1]people.cn. WHO Named the Novel Coronavirus-infected Pneumonia As “Covid-19” [EB/OL]. (2020-02-12)[2020-02-14]. http://world.people.com.cn/n1/2020/0212/c1002-31582421.html

[2] Official Website of National Health Commission. Notification of the Latest Situation of COVID-19 As of 24:00 on April 19. http://www.nhc.gov.cn/xcs/yqtb/202004/fceb99a885cd436980656f292a1c7282.shtml.

[3]Zhang Lin,Fitness of the Generalized Growth to the 2019 Novel Coronavirus Data[J]. Journal of University of Electronic Science and Technology of China,2020,49(03):345-348.

[4]CHEN J H, CHEN S Y, LUH P H, et al. Modeling chronic hepatitis B virus infections with survival probability metrics[J]. Operations Research for Health Care, 2017, 12:29-42.

[5]YU Zi,ZHANG GuiQing,LIU QingZHen,et al.The Outbreak Assessment and Prediction of COVID-19 Based on Time-varying SIR Model[J], Journal of University of Electronic Science and Technology of China, 2020,49(03):357-361.

[6] LIN Jun-feng, Assessment and Prediction of COVID-19 Based on SEIR Model withUndiscovered People[J], Journal of University of Electronic Science and Technology of China, 2020,49(03):375-382.

[7]Official website of Beijing Municipal Health Committee, http://wjw.beijing.gov.cn/.

[8]Fan Ruguo, Wang Yibo, Luo Ming, et al. COVID-19 Spread Model and Inflection Point Prediction and Analysis Based on SEIR [J]. Journal of University of Electronic Science and Technology of China 2020, 49(03): 369 – 374.

[9] Xu Juannian, Ma Ran, Zhao Zhongrui, et al. Preliminary Prediction of the COVID-19 Outbreak Trend in Shandong Province [J]. Journal of Chongqing Normal University (Natural Science), 2020, 37(02): 101-106.

[10] Xu Gongxian, Feng Enmin, Wang Zongtao, et al. SEIR Dynamics Model for SARS Epidemic Disease and Its Parameter Identification [J]. Journal of Natural Science of Heilongjiang University, 2005(04): 459 – 462 + 467.

[11] Geng Hui, Xu Anding, Wang Xiaoyan, et al. Analysis on the Role Played by the Relevant Intervening Measures on the COVID-19 Outbreak Based on the SEIR Model [J]. Journal of Jinan University (Natural Science & Medicine Edition), 2020, 41(02): 175 - 180.

[12] Wu Song and Pan Faming. Completed SPSS Statistical Analysis [M]. Tsinghua University Press, 326 – 329.

[13]YI DaLi,LIGaoming,LENG Huaiming,Cluster analysis on epidemic status of COVID-19 among different provinces[J/OL], Journal of Chongqing Medical University, 1-6[2020-08-01].https://doi.org/10.13406/ j.cnki.

cyxb.002386.

[14] LI Q, GUAN X, WU P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus –Infected pneumonia[J/OL]. New England Journal of Medicine, 2020 [2020-01-31]. https://www.nejm.org/doi/pdf/10.1056/ NEJMoa 2001316.

[15] Zhao Mengjiao, Liu Zhong, Li Zhan, et al. Survey on a Clustering Outbreak of COVID-19 in Ji’nan City [J]. Journal of Shandong University (Health Sciences), 2020, 58(04):44-48.

[16] Chan JF, Yuan SF, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster[J]. Lancet, 2020, 395(10223): 514-23.

[17] Cheng Lin and Xu Huilan. Analysis of the Impact Factors of the Interval from the Entry of the Imported COVID-19 Case to the Confirmation in Guangdong [J/OL]. Journal of Southern Medical University: 2020 40(5): 741 -745.

[18] WHO. WHO Director-General's opening remarks at the media briefing on COVID-19-23 March2020[EB/OL] [2020.3.23]. https://www.who.int/dg/who-director-general-s-openingremarks-at-the-media-briefing-on-covid-19---23-march-2020.