The method of parameter reassessment for prediction of NCP(2019-nCoV) spread

Jin-Li Guo (✉ phd5816@163.com)
University of Shanghai for Science and Technology

Ya-Zhi Fu
University of Shanghai for Science and Technology

Research Article

Keywords: NCP, coronavirus, 2019-nCoV, epidemic spread, Logistic curve

Posted Date: April 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-24700/v1

License: ☕️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
The method of parameter reassessment for prediction of NCP(2019-nCoV) spread

Jin-Li Guo, Ya-Zhi Fu

Business School, University of Shanghai for Science and Technology, Shanghai, 200093, China

Abstract

This paper proposes a conversion rate prediction method and a parameter reevaluation method based on Logistic curve (S-curve) to predict the spread of NCP (the Novel coronavirus pneumonia). According to the statistical data, we use the conversion rate prediction method to predict the spread of NCP. The prediction accuracy is quite high. By fitting the cumulative number of NCP sufferers with the logistic curve, the average estimation method of the limit number is proposed to predict the spread of NCP and the limit number of sufferers. This paper also assessing the effectiveness of prevention and control measures with the dynamic estimation of the infection probability of NCP. Based on the Markov property, the parameter reevaluation method proposed in this paper avoids over-fitting the theoretical curve and improves the accuracy of prediction. This research idea is not only suitable for Logistic curve regression, but also for other regression prediction problems.

Keywords: NCP; coronavirus; 2019-nCoV; epidemic spread; Logistic curve.
PACS: 02.10.Ox, 89.75.Hc, 89.75.Fb

1. Introduction

In the past two years, breakthroughs have been made in the network science study. In 1998, Watts and Strogatz established the Small World (SW) network model [1, 2]; In 1999, Barabasi and Albert proposed the concept of Scale Free (SF) network [2, 3]. The discovery of small-world effects and scale-free properties has universal significance in complex networks. As a result, it has attracted widespread attention. The theory and application of complex networks have developed rapidly [4, 5]. The discovery of scale-free and small worlds has important implications for the understanding of disease spread [4]. On December 29, 2019, Hubei and Wuhan’s health departments received a report on cases of aggregated unknown cause pneumonia from a local hospital [6]. A metagenomic analysis of the patient's lung lavage fluid revealed the presence of coronavirus. This virus was temporarily named as 2019-nCoV. On January 8, 2020, the new coronavirus was identified as the pathogen of this outbreak. On January 15th, the Chinese National Health Commission issued materials for diagnosis, treatment, prevention and etc. of pneumonitis associated with a new coronavirus infection. On January 20, 2020, the Chinese State Council agreed to incorporate the new coronavirus pneumonia into the management of the Infectious Diseases Law and the Sanitary Quarantine Law. Besides, the Council would start the nationwide emergency prevention and control work [6]. On February 8, at the press conference of the Joint Prevention and Control of the Chinese State Council, the spokesperson issued that the English name of

1Corresponding author. E-mail address: phd5816@163.com (J.Guo).
the novel coronavirus-infected pneumonia is "Novel coronavirus pneumonia", "NCP" for short. If a person is regarded as a node, as long as one person can inhale the other's breath, there is an edge between them. This is a small and scale free network. Respiratory diseases can easily spread on such networks. NCP outbreaks in China during the Spring Festival, the large flow of people in all directions has turned a relatively static network into a dynamic network that is randomly connected on the original scale-free network with a greater probability. Such network has a huge number of nodes which have large degrees, and the disease spreads very quickly on such interpersonal networks. Although since January 23, 2020, Wuhan and other cities have suspended urban public transport, but the number of NCP sufferers has risen sharply still. As of 24:00 on February 6, 2020, a total of 31,161 cases have been confirmed; 1,540 cases have been discharged; 636 cases died; 26,359 suspected cases; 314,028 close contacts have been traced and 1,860,045 close contacts are still in medical observation.

With the rapid spread of NCP, there is an urgent need to predict the speed of transmission and the limit number of people infected. A large number of scholars have studied this problem by far. Based on 425 confirmed cases, the characteristics of new coronary pneumonia were studied, the doubling time and basic regeneration number of new coronary pneumonia were estimated, and the distribution of epidemiological delay time was analyzed \[7\]. The paper \[8\] estimated regeneration value of new coronary pneumonia based on SEIR model and predicted the development of the epidemic. The paper \[9\] predicted the number of potential cases in Wuhan at the very beginning of the outbreak. Besides, the epidemiological characteristics of new coronavirus 2019-nCoV transmission were studied in the paper \[10, 11\]. Based on the characteristics of 2019-nCoV development until February 1st, the SIR model the paper \[12\] was modified to solve the dynamic equation of virus evolution. With the number of susceptible regeneration, current infection rate and latent infection rate, the paper also studied the changing trend of infected sufferers analyzed the influence of government administrative actions on the trend change.

The purpose of this article is to establish a new coronavirus transmission model, predict the cumulative limit of confirmed NCP sufferers (who has been tested by the nucleic acid kit and the result was positive) and enrich theoretical research related to new coronal pneumonia at the same time. It should be noted that the confirmed sufferers in this article are all recognized by nucleic acid kit. This paper is organized as follows: In section 2, data collection and interpretation; In section 3, proposing a method for predicting the conversion rate of new coronary pneumonia spread by defining a conversion rate of suspected diagnosis; In section 4, using the Logistic curve (also known as the S curve) to estimate the limit number of confirmed NCP sufferers; In Section 5, evaluating the effectiveness of prevention and control measures through the dynamic estimation of infection probability; In Section 6, proposing a secondary estimation method of Logistic curve parameters to predict the daily trend of cumulative NCP sufferers; Section 7 is the conclusion of this paper.

2. Data sources and collection

We collected the number of close contacts, suspects, infected individuals, cures, deaths and people under medical observation published on the websites of National Health Commission of the People's Republic of China and Wuhan Municipal Health Commission from January 15, 2020 to February 6, 2020, which are shown in Table 1.

| Date       | Close contacts | Suspects | Infected individuals | Cures | Deaths | Under medical observation |
|------------|----------------|----------|----------------------|-------|--------|--------------------------|
| 2020/1/15  | 763            | 41       | 12                   | 2     | 119    |                          |
| 2020/1/16  | 763            | 45       | 15                   | 2     | 98     |                          |
| 2020/1/17  | 763            | 62       | 19                   | 2     | 82     |                          |
### 3. Conversion rate prediction method of NCP spread

The Chinese government has been reacting appropriately, adopting strict prevention and control methods and continuously publishing statistics on the number of suspected and confirmed patients every day. With these transparent information, we can use the conversion rate of suspected and confirmed cases to make predictions. Let us denote by $I$ the cumulative number of confirmed NCP sufferers (Who has been through the nucleic acid test and the result were positive), $I_{0}$ the cumulative infection number of the day, $I_{0}$ the number of the day before, $y$ the number of the suspected sufferers of the day, the conversion rate $\eta$ can be expressed as
\[ \eta = \frac{I_t - I_0}{y}, \]  
\[ P = I_t + y\eta, \]  
(1)  
(2)

And the prediction of the cumulative number of infected individuals of the next day \( P \) are given by

| \( t \) | Data    | \( y \) | \( I \) | \( \eta \) | Prediction number | relative error rate % |
|-------|---------|--------|------|-------|-----------------|----------------------|
| 1     | 2020-1-21 | 163   | 440  |       |                 |                      |
| 2     | 2020-1-22 | 393   | 571  | 0.80368098 |                 |                      |
| 3     | 2020-1-23 | 1072  | 830  | 0.65903308 | 887             | 6.8489911           |
| 4     | 2020-1-24 | 1965  | 1287 | 0.42630597 | 1536            | 19.384884           |
| 5     | 2020-1-25 | 2684  | 1975 | 0.35012723 | 2125            | 7.5793029           |
| 6     | 2020-1-26 | 5794  | 2744 | 0.28651267 | 2915            | 6.222357            |
| 7     | 2020-1-27 | 6973  | 4515 | 0.30566103 | 4404            | 2.457267            |
| 8     | 2020-1-28 | 9239  | 5974 | 0.20923562 | 6646            | 11.255011           |
| 9     | 2020-1-29 | 12167 | 7711 | 0.18800736 | 7907            | 2.5434823           |
| 10    | 2020-1-30 | 15238 | 9692 | 0.16281746 | 9998            | 3.1622529           |
| 11    | 2020-1-31 | 17988 | 11791| 0.13774774 | 12173           | 3.2398644           |
| 12    | 2020-2-1  | 19544 | 14380| 0.14392929 | 14269           | 0.7732526           |
| 13    | 2020-2-2  | 21558 | 17205| 0.14454564 | 17193           | 0.0700147           |
| 14    | 2020-2-3  | 23214 | 20438| 0.14996753 | 20321           | 0.5719008           |
| 15    | 2020-2-4  | 23260 | 24324| 0.16739898 | 23919           | 1.6635988           |
| 16    | 2020-2-5  | 24702 | 28018| 0.15881341 | 28218           | 0.7127573           |
| 17    | 2020-2-6  | 26359 | 31161| 0.12723666 | 31941           | 2.5031576           |
| 28    | 2020-2-7  |       |      |        | 34515           |                      |

It can be seen from Table 2 that the accuracy of using conversion rate method to predict NCP propagation is very high, and the relative error rate between most prediction data and real data is within 5%. The shortcoming of the conversion rate prediction method of NCP spread is that it cannot be used for long-term prediction. We will discuss the medium- and long-term prediction method of the NCP sufferers number later.
4. NCP spreading model and limit number estimation

Let $S$ denote the susceptible state, the state of an individual before being infected. The individual is likely to be infected by a neighbor. Let $I$ denote the infected state in which an individual has been infected. This individual can infect its neighbors with a certain probability $\gamma$. Since we are predicting the cumulative number of confirmed sufferers, it can be assumed that people will remain infected once they are infected. This is the classic SI model of infectious diseases \[4\]. From this model, it can be seen that the cumulative number of sufferers meets the Logistic growth curve. Since the Logistic curve is S-shaped, it is also called "S-curve". The Logistic curve equation is

$$I(t) = \frac{L}{1 + e^{a - \gamma t}} \quad (2)$$

Where $t$ is a time series; $I(t)$ indicates the cumulative number of confirmed sufferers at time $t$; $\gamma$ is the probability that an individual becomes infected with the 2019-nCoV virus; coefficient $a$ is constant; $L$ is the limit number of sufferers. As long as the values of parameters $a$, $\gamma$, $L$ in the model are determined, we can use the model to predict the number of confirmed NCP sufferers. We can estimate $L$ according to the three-point method and the four-point method \[13\].

What is the maximum number of people who can be infected by this new coronal pneumonia? Wuhan "closed the city" on January 23, 2020, indicating that the country attached great importance to the epidemic on January 20 (According to the academician, Nanshan Zhong on live TV), and that all measures were basically implemented by January 27. Therefore, we estimate the limit data based on January 27 as $L = 84754$.

According to linear regression analysis, we can get

$$\gamma \approx 0.4059, \quad a \approx 8.0245$$

Thus, the Logistic curve equation can be obtained as follows:

$$I(t) = \frac{84754}{1 + 3657.37e^{-0.4059t}} \quad (3)$$

![Fig. 1. The cumulative number of infected individuals and Logistic curve fitting graph.](image)

The data are collected from January 15 to February 6, 2020 according to Table 1, and the limit number obtained by the three-point method is $L = 36457$. However, as of February 9, 2020, the cumulative number of confirmed sufferers has reached 40171, indicating that the theoretical curve is overfitting and it is not conducive to the prediction of future data. The question how to estimate the limit number reasonably is raised. Let us denote by $C_n$ the cumulative number of confirmed sufferers on the day $n$. We propose the average estimation method: starting from the first three data, estimating $L_i$, $i = 1, 2, ..., n$ one by one with the three-point or four-point method \[7\], taking the last non-positive $L_m$ as the benchmark and then averaging the following limit numbers as an estimate of $L$. This can be expressed as
5. Evaluating the effectiveness of prevention and control through dynamic estimation of infection probability

The incubation period of NCP is 14 days. The number of infected individuals has been counted since January 15, 2020. We dynamically estimate the infection probability from January 28.

\[
L = \begin{cases} 
\frac{1}{n-m} \sum_{j=m+1}^{n} L_j & \text{if } \frac{1}{n-m} \sum_{j=m+1}^{n} L_j > C_n \\
\max\left\{\frac{1}{n-m} \sum_{j=m+1}^{n} L_j, L_n\right\} & \text{if } \frac{1}{n-m} \sum_{j=m+1}^{n} L_j \leq C_n
\end{cases}
\] (4)

6. Parameter reassessment method and prediction of NCP sufferers

When using theoretical models to fit actual data, the pursuit is to minimize the average distance between each actual point and the corresponding theoretical point. The advantage of this is that the actual data is near the theoretical curve in the average sense; the disadvantage is that although the data fits well, the prediction results are not definitely good. This major public health emergency broke out in Wuhan. Medical resources were temporarily strained and the previous testing reagents, equipment, and personnel could not keep up. With the availability of medical resources later, the number of confirmed sufferers has gradually increased and such data has Markov property. Assuming that we have collected \( n \) cumulative confirmed cases, the number of confirmed sufferers in the future is not relevant with the cumulative number of confirmed sufferers before the number \( m \). At number \( i \), the cumulative number of

Fig. 2. Dynamic estimation of infection probability

After the outbreak of the new coronavirus, level one response to major public health emergencies have been initiated across the country since January 29, 2020. It can be seen from Figure 2 that since January 28, 2020, the probability of new coronavirus infection has been decreasing over time, indicating that the national joint prevention and control measures have taken effect.
confirmed sufferers is \( y_i \). The theoretical number of sufferers at time \( t_i \) is 
\[
I(t_i) = \frac{L}{1 + e^{a-\gamma t_i}}.
\]
Assuming the first parameter estimation has determined \( a \) and \( \gamma \), we can get the limit parameter \( L \) by the following equation
\[
\min \frac{L}{\sum_{i=m}^{n} \left( \frac{L}{1 + e^{a-\gamma t_i}} - y_i \right)^2}
\]
(5)
Solving the equation (5), we get \( L \)
\[
L = \frac{\sum_{i=m}^{n} \frac{y_i}{1 + e^{a-\gamma t_i}}}{\sum_{i=m}^{n} \frac{1}{(1 + e^{a-\gamma t_i})^2}}
\]
(6)
Assuming that the first parameter estimation has determined \( L \) and \( \gamma \), we can get the limit parameter \( a \) by the following equation
\[
\min \frac{L}{\sum_{i=m}^{n} \left( \frac{L}{1 + e^{a-\gamma t_i}} - y_i \right)^2}
\]
(7)
Solving the equation (8), we get \( a \)
\[
\sum_{i=m}^{n} \frac{y_i e^{a-\gamma t_i}}{(1 + e^{a-\gamma t_i})^2} = L \sum_{i=m}^{n} \frac{e^{a-\gamma t_i}}{(1 + e^{a-\gamma t_i})^2}
\]
(8)
When \( m = n \),
\[
a = \frac{\gamma}{n} t_n + \ln\left(\frac{L}{y_n} - 1\right)
\]
(9)
Assuming that the first parameter estimation has determined \( L \) and \( a \), we can get the limit parameter \( \gamma \) by the following equation
\[
\min \frac{L}{\sum_{i=m}^{n} \left( \frac{L}{1 + e^{a-\gamma t_i}} - y_i \right)^2}
\]
(10)
Solving the equation (11), we get \( \gamma \)
\[
L \sum_{i=m}^{n} \frac{t_i e^{a-\gamma t_i}}{(1 + e^{a-\gamma t_i})^3} = \sum_{i=m}^{n} \frac{y_i t_i e^{a-\gamma t_i}}{(1 + e^{a-\gamma t_i})^2}
\]
(11)
When \( m = n \),
\[
\gamma = \frac{1}{t_n} \left( a - \ln\left(\frac{L}{y_n} - 1\right) \right)
\]
(12)
According to Table 1, making \( n = 23 \) and \( m = 16 \). Using equation (6) to estimate \( L \), we can obtain the cumulative diagnosed prediction data of NCP transmission as shown in Table 3. The fitting result between the actual cumulative sufferers and the cumulative predicted sufferers is shown in Figure 3.

| \( t \) | Date      | \( I \) | Prediction number | Relative error | Relative error rate % |
|-------|-----------|--------|-------------------|--------------|-----------------------|
| 1     | 2020-1-15 | 41     | 42                | -0.86219     | 2.10                  |
| 2     | 2020-1-16 | 45     | 60                | -14.6213     | 32.49                 |
| 3     | 2020-1-17 | 62     | 85                | -22.9008     | 36.94                 |
| 4     | 2020-1-18 | 121    | 121               | 0.128008     | 0.11                  |
| Date       | Prediction | Actual | Error   | NCP Sufferers |
|-----------|------------|--------|---------|---------------|
| 2020-1-19 | 198        | 172    | 25.97   | 13.12         |
| 2020-1-20 | 291        | 245    | 46.28   | 15.90         |
| 2020-1-21 | 440        | 348    | 92.08   | 20.93         |
| 2020-1-22 | 571        | 494    | 76.82   | 13.45         |
| 2020-1-23 | 830        | 701    | 128.98  | 15.54         |
| 2020-1-24 | 1287       | 993    | 294.39  | 22.87         |
| 2020-1-25 | 1975       | 1402   | 573.11  | 29.02         |
| 2020-1-26 | 2744       | 1973   | 771.15  | 28.10         |
| 2020-1-27 | 4515       | 2763   | 1752.47 | 38.81         |
| 2020-1-28 | 5974       | 3842   | 2132.11 | 35.69         |
| 2020-1-29 | 7711       | 5294   | 2417.45 | 31.35         |
| 2020-1-30 | 9692       | 7204   | 2487.89 | 25.67         |
| 2020-1-31 | 11791      | 9648   | 2142.77 | 18.17         |
| 2020-2-1  | 14380      | 12664  | 1716.25 | 11.93         |
| 2020-2-2  | 17205      | 16222  | 982.55  | 5.71          |
| 2020-2-3  | 20438      | 20208  | 229.71  | 1.12          |
| 2020-2-4  | 24324      | 24419  | -95.39  | 0.39          |
| 2020-2-5  | 28018      | 28603  | -584.86 | 2.09          |
| 2020-2-6  | 31161      | 32512  | -1351.24| 4.34          |
| 2020-2-7  | 35962      |        |         |               |
| 2020-2-8  | 38856      |        |         |               |
| 2020-2-9  | 41182      |        |         |               |
| 2020-2-10 | 42988      |        |         |               |
| 2020-2-11 | 44353      |        |         |               |
| 2020-2-12 | 45365      |        |         |               |
| 2020-2-13 | 46102      |        |         |               |
| 2020-2-14 | 46635      |        |         |               |
| 2020-2-15 | 47016      |        |         |               |

**Fig. 3.** Cumulative actual number of NCP sufferers and Logistic prediction curve.

Fig. 3 is a graphical display of Tab. 3. It seems that the fit result of Fig. 3 is not as good as Fig. 1.
But the most recent 5 data in Fig. 3 fit better than in Fig. 1, and its prediction ability is stronger.

At the time of writing this article’s model, the data of NCP was only updated to February 6, 2020. The new coronavirus pneumonia data has been updated to February 12 now. Therefore, we can use these newly published data to test the specific prediction effect of the Logistic curve prediction method and the parameter reassessment method. It should be noted that on the 12th, 13332 clinically diagnosed cases in Hubei were included in the cumulative number of confirmed sufferers for the first time, while the confirmed patients were all recognized by nucleic acid kit in the past. Therefore, the number of cumulative diagnoses on the day increased by 15,152. In order to unify the data to test the effect of the prediction model in this paper better, the cumulative number of confirmed patients on the 12th day in our statistics does not include clinical individuals. The results are shown in Tab. 4 and 5.

| Date   | Prediction number | Relative error rate % |
|--------|-------------------|-----------------------|
| 2020-2-7 | 34546             | 32139                | 6.97 |
| 2020-2-8 | 37198             | 33401                | 10.21 |
| 2020-2-9 | 40171             | 34319                | 14.57 |
| 2020-2-10 | 42638            | 34973                | 17.98 |
| 2020-2-11 | 44653            | 35433                | 20.65 |
| 2020-2-12 | 46472             | 35753                | 23.06 |

Comparing Tab. 4 and Tab. 5, we can clearly see that the Logistic curve method’s relative error rate is unstable. Its relative error rate on February 12 is as high as 23.06%. While the relative error rates of parameter reassessment method are all within 5%. It’s easy to draw the conclusion that the prediction effect of parameter reassessment method is quite better.

7. Conclusion

The paper is based on the situation that NCP statistical standards are unchanged. We use the Logistic model to study the spreading trend of NCP. By proposing a parameter reassessment method, the spreading trend of the new coronary pneumonia has been analyzed and predicted. The effectiveness of prevention and control measures is evaluated by dynamic estimation of infection probability. The prediction results in this article indicate that the spread of NCP will start to stabilize in mid-February 2020 and the epidemic will be basically controlled in late February or early March. By comparing the prediction results of the Logistic curve method and the parameter reassessment method, the validity of the parameter re-evaluation method proposed in this paper is verified. The results of this article have certain reference value for the formulation of joint prevention and control policies. This paper proposes
a parameter reassessment method based on the data’s Markov property, which avoids over-fitting the theoretical curve and improves the accuracy of the prediction. The research idea is not only suitable for the Logistic curve regression, but also for other regression prediction problems.

Acknowledgments

The authors acknowledge the supports of National Natural Science Foundation of China (Grant No. 71571119).

References

[1] Watts, D. J., Strogatz, S. H., Collective dynamics of "small-world" networks. J. Nat. 393, 440-442 (1998).
[2] Albert, R., Barabási, A. L. Statistical mechanics of complex networks. J. Rev. Mod. Phys. 74, 47–97 (2002).
[3] Barabási, A. L., Albert, R. Emergence of scaling in random networks. J. Sci. 286, 509-512 (1999).
[4] Wang, X.F., Li, X., Chen, G.R. Introduction to the Complex Networks (Higher Education Press), 8-10 (Beijing, 2012).
[5] Lu, J.H., Tan, S.L. Games and evolutionary dynamics on the Complex Networks (Higher Education Press), 53-57(Beijing, 2019).
[6] Wuhan Municipal Health Commission. Wuhan Municipal Health and Health Commission’s Briefing on the Current Pneumonia Epidemic Situation in Our City. (2020). http://wjw.wuhan.gov.cn/front/web/showDetail/20191231080898.
[7] Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. J. N. Eng. J. Med. (2020). DOI: 10.1056/ENJMoal2001316.
[8] Zhou, T. et al. Preliminary prediction of the basic reproduction number of 2019-nCoV. J. J. Evi. Med. (2020). DOI: 10.1111/jebm.12376.
[9] Imai, N. et al. Estimating the potential total number of novel coronavirus cases in Wuhan City, China. J. Imp. Col. Lun. (2020). Available online: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/2019-nCoV-outbreak-report-22-01-2020.pdf.
[10] Read, J.M., Bridgen, R.E.&jJessica, Cummings A.T&Derek, Ho, A.,Jewell, C. P. Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. J. medRxiv. (2020). DOI: 10.1101/2020.01.23.20018549.
[11] Imai, N. et al. Transmissibility of 2019-nCoV. J. Imp. Col. Lun. (2020). Available online: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf.
[12] Yu, Z., Zhang, G. Q., Liu, Q. Z., Ly, Z.Q. 2019-nCoV epidemic assessment and forecast based on time-varying parameter-SIR model. J. J. Uni. Ele. Sci. Tec. Chi. (2020). Available online: http://kns.cnki.net/kcms/detail/51.1207.t.20200207.1546.002.html.
[13] Yin, Z.Y. Research on Fitting Method of Logistic Curve. J. Math. Sta. Man. 21, 41–45 (2002). DOI: 10.13860/j.cnki.sltj.2002.01.010.
Figures

Figure 1

The cumulative number of infected individuals and Logistic curve fitting graph.

Figure 2

The infection probability over time from January 26 to February 9, 2020.
Dynamic estimation of infection probability

Figure 3

Cumulative actual number of NCP sufferers and Logistic prediction curve.