

A mathematical model for the spread of Omicron virus based on grey prediction algorithm

Haotian Lin ^{*,#}, Jianze Lin [#], Kewei You [#]

School of Ocean Information Engineering, Jimei University, Xiamen, China

* Corresponding author: linhaotian@jmu.edu.cn

[#]These authors contributed equally.

Abstract. After two years of the Newcastle epidemic, the emergence of a mutant strain of Omicron virus has once again raised a high level of alarm worldwide, and it has rapidly replaced the Delta strain as the major strain in the world epidemic with high infectivity. In order to study the spread of the epidemic in a qualitative and quantitative manner, this paper uses relevant mathematical modeling ideas and methods, combined with the background of big data, to conduct relevant research and analysis in order to provide guidance for the actual fight against the epidemic and facilitate scientific research and response measures. In this paper, a gray prediction algorithm is used to predict the future epidemic data. A metabolic algorithm is introduced to recursively correct the prediction residuals and build a multilevel gray model. In this paper, to investigate the role of vaccines on epidemics, the R values under each dose of vaccine were calculated by selecting data from several countries and regions using fitted curve idealization. The prediction of the data yields the conclusion that vaccination is indeed directly effective in controlling the spread of the epidemic.

Keywords: Gray prediction algorithm, Omicron, Gray prediction Prediction.

1. Introduction

In the current global scenario of neo-coronavirus, the Omicron strain, a new variant of the novel coronavirus, has become the dominant strain worldwide with a strong infectious and immune escape capability, the most complete neutralizing "escapee" scientists have ever seen. People who have previously been infected with other strains of NIV and those who have received vaccination boosters are at risk of re-infection with the Omicron variant [1, 2]. Therefore, it is imperative to determine the transmission coefficient of Omicron virus and to predict the future epidemiological development of the epidemic, taking into account the epidemiological data from various countries and regions around the world [3].

In this paper, develop a mathematical model to predict future epidemic data based on existing epidemic data, a series of control measures, and relevant factors such as cure rates, severe disease rates, and mortality rates, and discuss how the current vaccination situation will affect the spread of the Omicron epidemic. Specific recommendations are made for practical measures to prevent the outbreak.

2. Model assumptions and notation

2.1. Assumptions[4]

- 1 Assume that no strict preventive and control measures were carried out at the beginning of the epidemic in Hong Kong.
- 2 Assume that all data related to the epidemic given in different places are accurate.
- 3 Assume that different vaccines are equally effective in the prevention and control of the epidemic.
- 4 Assume that the overall effect of different population attributes (e.g., average age, form of government organization) on the spread of the epidemic can be ignored in each region.
- 5 Assume that populations that have never received any type of vaccine are susceptible.

6 Assuming average incubation period of omicron virus and average time from onset of infection to the next unit of infection for infected individuals, etc.

7 Assuming that data from the Hong Kong outbreak prior to February 10 do not have a direct impact on data from February 10 onwards.

8 Assuming that the ongoing mutation of the virus is slow and gradual.

2.2. Notations

Important notations used in this paper are listed in Table 1.

Table 1. Notations

Symbols	Description
R_0	The basic regeneration number of Omicron virus
R	Effective number of regenerations of Omicron virus at a site
c	Residuals
δ	Correction factor for residuals
S	Infected
I	Vulnerable to infection
N	Total population in an area
S_t	Number of susceptible population in an area

3. World Epidemic Analysis and Forecast and Vaccine Value Assessment

3.1. Visualization of data related to the world epidemic

Data on worldwide outbreaks over a 90-day period from January 31 (Figure 1) were selected for analysis [5].

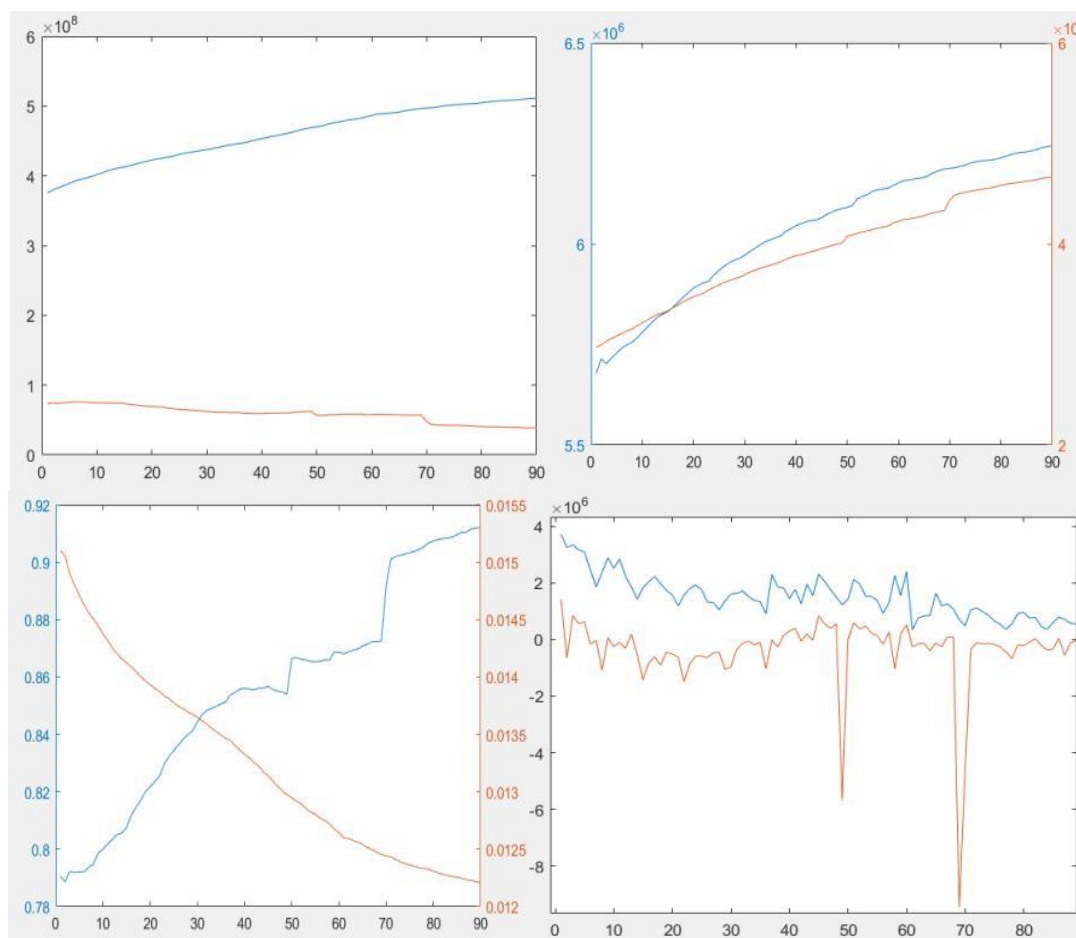


Figure 1. Various data of the world epidemic

Assuming recent mortality = new deaths/new confirmed diagnoses, there is recent mortality as shown in Figure 2.

Overall, the rate of transmission of the epidemic is gradually decreasing and the cure rate is rapidly increasing with the low virulence mutation of the virus and the related medical technology level, while the mortality rate is clearly having a clear downward trend. In this paper, we do not explore other properties of the virus, but only model the analysis from the perspective of infection.

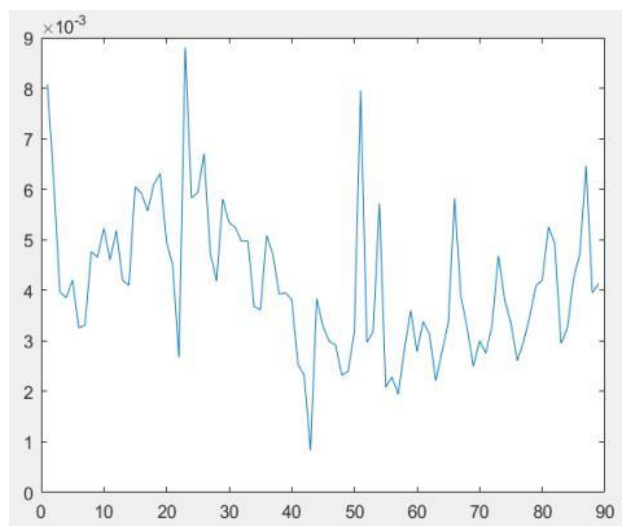


Figure 2. Recent mortality

3.2. Gray projections [6, 7] of world epidemics

The main feature of gray forecasting is that the model uses not the original data series, but the generated data series. The core system is the Grey Model (GM), which is a method of modeling the original data to obtain an approximate exponential pattern by cumulative generation (or other methods). The data were fitted to the cumulative diagnosis, existing diagnosis and mortality data (Table 2), and it was found that there was an exponential trend in the original data, and the interrelationship between the data was not clear.

Table 2. Table of P-values for different fitting methods of correlation data

Fitting method	Linear	Secondary	Index
Cumulative confirmed diagnosis	0.0369	0.0501	0.0333
New confirmations	0.0247	0.0475	0.0225
Mortality rate	0.0159	0.0443	0.0209

It can be seen that all three types of data conform well to the linear and exponential trends, but the specific information about their systems is still unclear, so the choice to consider the exponential trend, the choice of gray prediction, the establishment of the original data to build the GM (1,1) model for whitening prediction.

$$\begin{aligned} X(1) &= (X(1)(1), X(1)(2), \dots, X(1)(n)) \\ &= (x(0)(1), x(0)(1) + x(0)(2), \dots, x(0)(1) + \dots + x(0)(n)) \end{aligned} \quad (1)$$

$$z(1) = (z(1)(2), z(1)(3), \dots, z(1)(n)) \quad (2)$$

Establishing the grey differential equation

$$x(0)(k) + az(1)(k) = b, \quad k = 2, 3, \dots, n, \quad (3)$$

The corresponding whitening differential equation is given by

$$\frac{dx^{(1)}}{dt} + ax(1)(t) = b \quad (4)$$

Due to the large amount of data, the direct use of the GM (1,1) model is not effective, so here we add the new metabolic algorithm, the specific process is to construct the GM (1,1) model again for the residuals for selective correction, in the process of constantly adding new data to the residuals cycle correction, to achieve the best fit effect.

$$x(t + 1) = b \times \text{ert} \mp 6 \times b \times \text{ert} \quad (6 = 0, 1) \quad (5)$$

The basic idea of performing cyclic residual correction is that we establish a new gray equation for the residuals, calculate the residual pair with the residual's gray equation of the secondary residuals, and then perform residual correction on the original data. For a specific item, if the prediction result does not meet the test result, we make a secondary correction to the residuals, otherwise no correction is made, and so on and so forth until all predictions are completed. And how to make prediction correction it, we need to start in the known data segment on the continuous gray prediction, while ensuring that each data segment on the known data, and then calculate the posterior difference test number, the error of the larger items for correction. This is repeated until the prediction is actually started, and then a more accurate fit is obtained.

The formula for the posterior difference test number is

$$C=S1/S2 \quad (6)$$

On this basis, this paper proposes an idea of multiple corrections using program recursion, i.e., if the residual correction and the quadratic correction of the residuals still cannot pass the test, the higher-order correction equation continues to be built on that interval segment until the data segment can pass the test.

With a significant improvement in the global epidemic, a slowdown in the trend of cumulative confirmations, and a decrease in both new confirmations and mortality rates, the paper is optimistic that, with the right measures, there is hope that the epidemic will move into a post-epidemic era with localized epidemics and manageable mortality.

3.3. Exploring the role of vaccines in influencing epidemics

In order to investigate the effect of vaccines on the epidemic, selected several countries and regions, conducted the corresponding tests, and then used the same first method to establish the relationship between the number of doses of vaccine per capita and the effective regeneration number R[8,9] within 28 days after vaccination. It can use this study to investigate the intuitive relationship between the number of vaccinations and the corresponding transmission capacity of the epidemic.

The relevant data for the five regions are given in Table 3 below.

Table 3. Relevant data for the five regions

Region	Hong Kong (Feb.10)	Belgium	South Africa	United Kingdom	Singapore
Average dose	1.62	2.21	1.30	2.35	2.65
R	2.67	1.95	4.04	1.87	1.71

Due to the small number of data points and the strong dispersion, it is not convenient to use a very complicated way to fit the data, so we directly use the cftool toolbox of matlab to fit the data in a simple way and choose the curve that fits the best. The ideal R data for the corresponding dose were

obtained by fitting the curve with the coefficients given by the cubic polynomial curve and substituting the corresponding dose as follows Table 4.

Table 4. Ideal R

Average dose	0	1	2	3
R	25.43	6.36	2.03	0.98

Choose the power function to fit and give the fitting results. By fitting the curve and substituting the corresponding dose, the ideal R data under the corresponding dose are obtained as in Table 5.

Table 5. Ideal R

Average dose	0	1	2	3
R	\	7.86	2.09	1.67

RMSE (root mean square error) was introduced as a measure of the fitting effect. MSE is the average of the squared sum of the difference between the predicted and actual values, and RMSE is the quadratic root of MSE, which is a very effective measure of the fitting effect, and can intuitively show the difference in the fitting effect of each curve, which is calculated by the following formula.

$$MSE = 1/N \left(\sum_{i=1}^N (k_i - k_i')^2 \right) \quad (7)$$

$$RMSE = \sqrt{MSE} \quad (8)$$

The accuracy of the polynomial fit is higher than that of the power function, and the power function will fail in the calculation without vaccination, so the subsequent calculation army in this paper is based on the polynomial fit.

To verify the above findings and to investigate the change in R as the number of vaccination doses increases in the same region, we selected the number of doses per capita in Hong Kong for the five weeks after February 10 and the relationship between R (due to the limitation of the duration of the epidemic, we could only select the data for the corresponding four weeks for the calculation of the R index regression relationship). The relationship between the R-index and the per capita doses in Hong Kong for the five weeks after February 10 is shown in Figure 3 below.

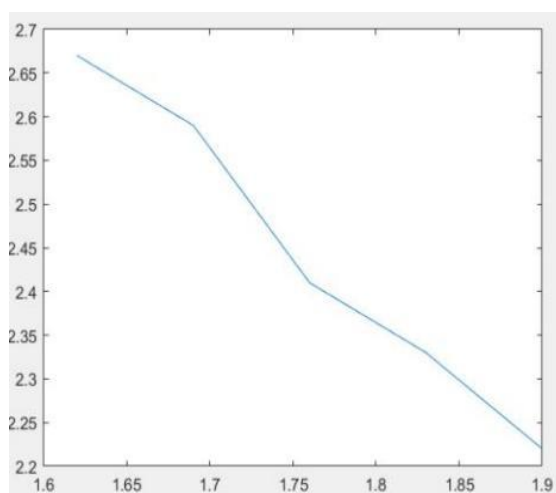


Figure 3. Graph of the relationship between the average number of doses of vaccination and the number of effective regenerations

The presence of a significant negative correlation between the average number of doses of vaccine and the number of effective regenerations in the same area also confirms the conclusions obtained in

the previous paragraph of the paper and does not differ significantly from the fit obtained in the above case. It is clear to conclude that vaccination, regardless of the type of vaccine, is effective in controlling the spread of the epidemic.

Looking at the data, we can see that the average number of doses administered in Hong Kong is now more than 2.12, and the epidemic has been controlled very effectively and the city can be restored to order soon.

The SEIR model [10] has a direct relationship with R . In this paper, we choose to build the SEIR model first, and then use the R values of the doses corresponding to the model given above to calculate the R_0 instead of the SEIR model, and finally derive the predicted values directly from the SEIR model. The predicted values were obtained, and finally the predicted values for each case were used to then establish the relationship between the vaccination doses and the epidemic data speculated in this paper. We consider that since the epidemic prevention capacity and virus properties also evolve dynamically, choosing different orders in time for the same region may be influenced by other factors, so we still choose the data of doses in the same time period in multiple locations as the basis for modeling.

It should be noted that, in fact, since the epidemic prevention measures in each place are not only as simple as the number of vaccination doses, but the overall countries or regions with higher vaccination doses also have better anti-epidemic mechanisms, the overall number of vaccination doses still reflects a strong positive correlation with the development of the epidemic, and therefore, in this paper, R is considered to be directly related to the number of vaccination doses.

4. Conclusion

From the mathematical calculations and studies described above, some preliminary conclusions were drawn.

(i) Novel coronaviruses are indeed extremely infectious, possessing an infectiousness that is among the best of all diseases, even with universal vaccination and human intervention.

(ii) Novel coronaviruses are controllable and their risk of transmission can be effectively reduced by human intervention.

(iii) Vaccination remains one of the most effective means of epidemic prevention, and vaccination can be very effective in reducing the risk of transmission of the epidemic.

(d) Overall, the future remains positive, with infection rates declining and mortality rates declining even faster, and the epidemic, while difficult to eradicate completely, has the opportunity to become a low-risk disease with localized epidemics.

In addition, investigated the outbreak data for Hong Kong from February to March 2022 and used mathematical methods to develop and analyze models of infectious disease transmission. In this wave, the total number of cases reported by nucleic acid testing was 920019 and the cumulative number of deaths was 7690. Through our investigation, we found that the number of confirmed cases increased very rapidly, reaching a peak on March 3, when the number of confirmed cases reached 56,825 on March 3 alone. Based on the Hong Kong data, assuming that people who have not received a single dose of vaccine are susceptible, the R_0 is calculated to be about 10.12, which is highly consistent with Dr. Wenhong Zhang's conclusion ($R_0 = 9.5$) that even in Hong Kong, where vaccination rates are high, it is necessary to strictly isolate confirmed patients who may pose a public health risk.

Analyzed the number of deaths in the current outbreak in Hong Kong, which peaked on March 12, with as many as 292 deaths in a single day. This may be due to the lack of success in treatment over several days, but the number of confirmed cases and deaths have since leveled off, thanks to the Hong Kong government's efforts to control the epidemic, and only under strict conditions can the number of confirmed cases and deaths be effectively reduced slowly. This is a frightening figure: 7 out of every 100 people who get sick die. Under the tremendous pressure, the healthcare system in Hong Kong faced collapse, the death rate increased, and many patients could not be treated to the fullest

extent, but with the efforts of the Hong Kong people themselves and the establishment of a regional hospital and the arrival of medical teams from the mainland, Hong Kong eventually survived.

References

- [1] Zhang Xiaobao. Research progress of basic regeneration number and effective regeneration number indicators in infectious disease epidemiology [J].
- [2] Wallinga J, & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. [J]. *Proceedings of the Royal Society B: Biological Sciences*, 2007.
- [3] Zhao S. 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. [J]. *International Journal of Infectious Diseases*, 2020.
- [4] Li Q, Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... & Xing, X. Early transmission dynamics in Wuhan [J]. *New England Journal of Medicine*., 2020.
- [5] Wu J T, Leung, K., & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study.[J]. *The Lancet*, 2020.
- [6] Xie, N.M., A summary of grey forecasting models. *Grey Systems-Theory and Application*.
- [7] Liu, S. F., F. Jeffrey, and Y.J. Yang. Several Grey System Models which Used Most Commonly. in 11th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM). 2013. Greece.
- [8] Chen, D.B. and T. Zhou, Evaluating the effect of Chinese control measures on COVID-19 via temporal reproduction number estimation. *Plos One*, 2021. 16 (2).
- [9] Alimohamadi, Y., M. Taghdir, and M. Sepandi, Estimate of the Basic Reproduction Number for COVID-19: A Systematic Review and Meta-analysis. *Journal of preventive medicine and public health Yebang Uihakhoe chi*, 2020. 53 (3): p. 151 - 157.
- [10] Sampath, S. and J. Bose. Modeling Effect of Lockdowns and Other Effects on India Covid-19 Infections Using SEIR Model and Machine Learning. in 13th International Conference on Soft Computing and Pattern Recognition (SoCPaR) / 13th World Congress on Nature and Biologically Inspired Computing (NaBIC). 2021. World Wide Web, ELECTR NETWORK.