European Journal of Preventive Medicine
2023; 11(1): 11-15
http://www.sciencepublishinggroup.com/j/ejpm
doi: 10.11648/j.ejpm.20231101.13
ISSN: 2330-8222 (Print); ISSN: 2330-8230 (Online)

Research on the Application of Grey Modeling Theory in the Prediction of Future Epidemic Trend of Infectious Diseases

Li Ming Quan

Nanchong City Committee Office of the Chinese People's Political Consultative Conference, Nanchong, China

Email address:
617738477@qq.com

To cite this article:
Li Ming Quan. Research on the Application of Grey Modeling Theory in the Prediction of Future Epidemic Trend of Infectious Diseases. *European Journal of Preventive Medicine*. Vol. 11, No. 1, 2023, pp. 11-15. doi: 10.11648/j.ejpm.20231101.13

Received: January 31, 2023; **Accepted**: February 25, 2023; **Published**: March 9, 2023

Abstract: Grey theory is a kind of grey system theory founded by Professor Deng Jurong, a Chinese scholar in the eighties century. It is a new method to study the uncertainty problem with few data and poor information. In recent years, grey modeling forecasting method has been widely used in industry, agriculture, science and technology, economic and social development planning and analysis, hydrology, geology, breeding and natural disaster prediction and so on. The prediction of epidemic trend of infectious diseases is an important work in the prevention and control of infectious diseases. Therefore, the author of this paper studied the application of grey modeling theory in the prediction and prediction of epidemic trend. Objective: To introduce the grey modeling prediction theory into the prediction of epidemic trend of infectious diseases, and to provide a new prediction method of epidemic trend of infectious diseases. The methods are as follows: The GM(1.1) model was established with the grey system theory, and the grey modeling was carried out on the short and medium term confirmed cases and the wave time of the epidemic outbreak in China in 2021, and the simulation accuracy and precision were analyzed. Results: The average simulation accuracy of the established short-term prediction model reached 73.2% and the prediction accuracy reached 99.85%. The average accuracy of the simulation value of the model reached 97.67% and the prediction accuracy reached 97.33% for the time of wave occurrence in the epidemic. This paper also analyzes the reasons for the low simulation value and prediction accuracy of the grey model forecasting method for the medium and long term confirmed cases of the epidemic.

Keywords: Gray Modeling, Infectious Disease, Fashion Trend, Forecast, Research

1. Introduction

In recent years, grey modeling forecasting method has been widely used in industry, agriculture, science and technology, economic and social development planning and analysis, hydrology, geology, breeding and natural disaster prediction and so on. Grey modeling prediction is a new method to study the problems of few data, poor information and uncertainty, mainly through the generation and development of "part" of the known information, extract valuable information to achieve the correct description of the system operation behavior, evolution law and effective monitoring [1]. The epidemic of human infectious diseases is an inevitable accidental event in nature. Its occurrence and prevalence are affected by many factors, and its development and change are also controlled by many factors. The frequency and intensity of confirmed cases and epidemic waves are specific indicators of the development trend of infectious diseases. The increasing frequency and intensity of epidemic waves indicate that the epidemic of infectious diseases is becoming more and more serious. If the number of confirmed cases decreases, the frequency and intensity of outbreaks indicate a trend towards a decline in the prevalence of infectious diseases [2]. With the available data materials, the author conducted a GM(1,1) gray modeling study on the number of short-term confirmed cases and epidemic wave occurrence of the outbreak and development of the novel coronavirus pneumonia in China, and applied the grey modeling theory to the prediction of the future epidemic trend of infectious diseases. The situation is reported below [3].

2. Materials and Methods

2.1. Source and Arrangement of Materials

The number of confirmed cases of COVID-19 in this article is from the daily Epidemic bulletin on the official website of the Health Commission of the People's Republic of China. The extraction period of the number of confirmed cases in the quiet period of the epidemic was from August 1, 2021 to August 25, 2021, with an interval of 5 days (short term in this paper) [4], that is, August 1 was used as the base number, and the subsequent dates of 5 and 10 were used as the data extraction date, and the number of confirmed cases was used as the original data for the short-term grey modeling prediction of the quiet period. Data on confirmed cases 30, 40, and 60 days apart (referred to here as the medium and long term of the epidemic) are from 1 January 2021 to the end of August 2021. In the study, we sorted the 2021 dates into raton numbers, with each number having a corresponding date and the number of newly confirmed COVID-19 cases on that day [5]. In this study, the daily new crown new confirmed cases of pneumonia, the percentage increase or decrease month-on-month, when one day quarter-on-quarter ratio > 0 and sequential ratio is too high, to wave ratio when determining outbreak occurred, the date for the outbreak wave time moment, take out the for should be the serial number [6]. grey modeling of wave time of the original sequence. The original data used in the study are summarized in Table 1.

Table 1. *Original data table.*

The number of daily confirmed cases with an interval of 5 days between 2021.8.1 and 2021.8.30 during the quiet period of the epidemic	98 124 111 51 20 26
2021.1.1-2021.8.30 epidemic The serial number of the wave occurrence	9 48 59 77 92 112 150

2.2. Research Methods

This study applied the basic theory of grey system, established GM(1,1) grey model, calculated the parameters of the time response function (prediction model), established a calculation model of the time response function series, and then calculated the simulation value and residual error according to the model, analyzed and judged the accuracy of the model [7], and finally carried out the actual verification of the model application, and analyzed the accuracy of the model used in the actual prediction.

3. Modeling and Analysis

3.1. Prediction of Short-Term Confirmed Cases

Starting from August 1, 2021, during the quiet period (a period of low epidemic fluctuation), data of confirmed cases will be collected every 5 days, based on the number of confirmed cases on August 1, until August 25, 2021. These confirmed cases were used to form the original series of short-term forecast of confirmed cases in chronological order [8]. the1-AGO accumulation of the original sequence is generated to form the cumulative sequence: $X^{(1)} = [X^{(1)}(1), X^{(1)}(2), X^{(1)}(3), X^{(1)}(4)\ X^{(1)}(5), X^{(1)}(6)]$, the accumulation generation number listed as $X^{(1)} = [98, 222, 333, 384, 404, 434]$, lose to the mean sequence: $Z^{(1)} = [Z^{(1)}(2), Z^{(1)}(3), Z^{(1)}(4), Z^{(1)}(5), Z^{(1)}(6)]$, close to the mean value of $Z^{(1)} = [160, 277.5, 385.5, 394, 417]$. The GM (1, 1) model of basic model for $X^{(0)}(k) + a\,Z^{(1)}(k) = b$, $\hat{a} = [a, b]^T$ column for a parameter, and $X^{(0)}(k) + aZ^{(1)}(k) = b$, the least squares estimate of parameter meet bty $\hat{a}=[B^TB]^{-1}B^Ty$

$$B = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ -Z^{(1)}(4) & 1 \\ -Z^{(1)}(5) & 1 \\ -Z^{(1)}(6) & 1 \end{bmatrix} = \begin{bmatrix} -160 & 1 \\ -277.5 & 1 \\ -358.5 & 1 \\ -394 & 1 \\ -417 & 1 \end{bmatrix}$$

$$y = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ X^{(0)}(4) \\ X^{(0)}(5) \\ X^{(0)}(6) \end{bmatrix} = \begin{bmatrix} 124 \\ 111 \\ 52 \\ 20 \\ 26 \end{bmatrix}$$

Where B^T is the transpose of B. Pass Calculated $\hat{a} = [a, b]^T = [0.422304, 206.351612]^T$ is a = 0.422304, b = 206.351612. The GM (1, 1) model $X^{(0)}(k+1) + aZ^{(1)}(k) = b$ for time response sequence$\hat{X}^{(1)}(k+1)=(X^{(0)}(1)-b/a)e^{-ak} + b/a = (98-488.632862)e^{-ak} + 488.632862 = -390.632862e^{-0.422304k} +488.632862$ The simulated value and residual error are calculated according to the time response series. See Table 2 for specific data. It can be seen from the data in*the* table that the total proportion of the absolute residual error to the original value is 1.34, the average residual error is 0.268, and the average simulation accuracy reaches 73.2%, which meets the basic requirements of the simulation accuracy of gray modeling The established.

Table 2. *Residual analysis of simulated values of confirmed cases 5 days apart during the calm period.*

Raw data	98	124	111	51	20	26
Simulated value	98	134.56	88.21	57.82	37.41	24.85
Absolute residual value	0	10.56	22.79	6.82	17.41	1.15
Residual error rate		0.09	0.21	0.13	0.87	0.04

Timeresponse function is used to predict the confirmed cases on August 30, 2021 at the next time point with a 5-day interval. According to the above established model, 26.97 confirmed cases are expected to occur. According to the data extraction method in this study, the next 5-day time interval should be August 30, 2021. 27 confirmed cases actually occurred on that day [9], almost equal to the predicted data of 26.96, with a prediction accuracy of 99.85%. Grey modeling was used to predict the calm period of the epidemic.

3.2. Grey Modeling Prediction of Medium and Long-Term Confirmed Cases of the Epidemic

In order to study the application of grey system modeling prediction theory in the medium and long term prediction of epidemic development. We studied the accuracy of grey modelling forecasts for the number of new confirmed cases over the medium to long term over the next 30, 40, and 60 days of the pandemic. In the period from January 1, 2021 to the end of August 2021, we selected the data of confirmed cases at the corresponding time points in the period by using the method of 30 days, 40 days and 60 days, respectively, to compose the original series of modeling, and established the simulation prediction function, and analyzed its accuracy. The results showed that the simulated data were far from the actual data, and the average difference between the simulated data and the original data was 122.1%, 98.02% and 33.7%, respectively [10]. The established models did not have satisfactory simulation accuracy and prediction effect. Therefore, we believed that the grey GM(1,1) modeling prediction method was not accurate in the prediction of medium and long-term confirmed cases of infectious diseases. Can not get the predicted satisfactory results.

3.3. Grey Modeling Prediction of Epidemic Wave Occurrence Time

Epidemic wave refers to the phenomenon of a sudden increase in confirmed cases at a certain time during an epidemic. If people can grasp the epidemic pattern and the approximate time of epidemic waves as early as possible, we can make all kinds of preparations in advance, reduce the possible harm and loss caused by the epidemic, and strengthen people's initiative to overcome the epidemic. The study on the wave occurrence of COVID-19 in this paper takes the real-time epidemic situation of COVID-19 in China as the sample [11]. In this study, the daily new crown new confirmed cases of pneumonia in the annulus is the percentage increase or decrease of calculation, when one day quarter-on-quarter ratio > 0 and sequential ratio is too high, to wave ratio when determining outbreak occurred, the date for the outbreak wave time moment, take out the for should be the serial number, grey modeling of wave time of the original sequence. We observed and analyzed the epidemic situation and the occurrence of confirmed cases in China during the period from January 1 to the end of June 2021. According to statistics, the number of the last epidemic wave occurrence before the end of June 2021 is 150, and the time is May 30, 2021. In the study, the confirmed cases in this period were calculated on a sequential basis (the proportion of the increase or decrease of the confirmed cases on that day compared with the previous day). According to the preliminary observation and analysis, more than ten wave outbreaks occurred during this period. In the study, the sequential ratio was set as ʒ, and the date of $ʒ \geq 0.8$ was taken as the outbreak time of the epidemic wave. If there was another value of $ʒ \geq 0.8$ within the next 3-5 days, it was counted as a wave time, and the moment of the first occurrence of $ʒ \geq 0.8$ was taken as the outbreak time of the epidemic wave. In this study, we extracted the serial numbers of occurrence times conforming to the regulations and $ʒ \geq 0.8$ to form the original series of epidemic wave occurrence. According to the original series, the cumulative series and the adjacent mean series were respectively generated, and the GM(1,1) model was established, whose basic type was $X^{(0)}k)+a_{ʒ}z^{(1)}(k)=b_{ʒ}$. $\hat{a} = [a_{ʒ}, b_{ʒ}]^T$ for the parameter columns, because.

$X^{(0)}(k) + a_{ʒ}z^{(1)}(k) = b_{ʒ}$, the least squares estimates of parameters satisfy $\hat{a} = [B_{ʒ}^T B_{ʒ}]^{-1} B_{ʒ}^T Y_{ʒ}$, the method (3.1) in this paper is used again for matrix calculation, through calculated $\hat{a} = [-0.226235, 39.68049]^T$, $a_{ʒ} = -0.226235$, $b_{ʒ} = 39.68049$, the number of time response as:

$$\hat{X}^{(1)}(k+1)=[(X^{(0)}(1) - b_{ʒ}/ a_{ʒ})e^{-ak} + b_{ʒ}/a_{ʒ}]$$

$$=[9-39.68049/-0.226235]e^{0.226235k}+39.680409/-0.226235$$

$$=184.3950e^{0.226235k} - 175.3950.$$

Then the simulation values and precision analysis were calculated according to the time response function. The simulated values and relative residual rates obtained through calculation are shown in Table 3. The data from the table shows that the simulated value corresponds to the original data fluctuation, good consistency, and the difference between the simulated value and the original value is very small. The sum of residuals is 0.1340, the average residuals is 2.23%, and the average accuracy of the simulated value reaches 97.67%. The time response series (model) established was used to predict the possible time of the next epidemic wave. Through calculation, it was known that the time number of the next epidemic wave was 181.92, about 182, and the corresponding time number was July 1, 2021. Through tracking, observation and recording, the actual occurrence point of the next epidemic wave was the time number 187. The time point was July 6, 2021. Although the actual occurrence time of epidemic wave was 5 days later than the predicted occurrence time, the accuracy of prediction reached 97.33%. The grey modeling was used to predict the occurrence of epidemic wave in infectious diseases.

Table 3. *Residual analysis of simulated values at the time of wave epidemic.*

Raw data	9	48	59	77	92	112	150
Simulated value	9	46.81	58.70	73.60	92.28	115.7	145.92
Absolute residual value	0	1.19	0.3	3.4	0.003	3.70	4.08
Residual error rate	0	0.025	0.005	0.044	0.000	0.033	0.027

4. Conclusion and Discussion

4.1. Applicability of Grey Modeling Theory in Prediction of Epidemic Trend of Infectious Diseases

The epidemic of human infectious diseases is an inevitable accidental event in nature. Most of them are emergencies [12]. Because they are emergencies, they are also an uncertain event with few data and poor information. The global epidemic of novel coronavirus is one of a variety of human infectious diseases. The application of grey modeling theory to predict the future trend of epidemic occurrence and development is consistent with the requirements of grey modeling theory. Moreover, in this study, the prediction of short-term confirmed cases and epidemic wave occurrence predicted by grey GM(1,1) modeling has achieved good results [13]. Therefore, the grey GM(1,1) modeling theory is scientific and practical in predicting the future trend of infectious disease epidemic.

4.2. Simulation Accuracy and Prediction Accuracy of the Prediction Model of the Future Epidemic Trend of Infectious Diseases by Grey Modeling Theory

In the prevention and control of major epidemics, if the number of confirmed cases and the time of wave occurrence can be known in advance, people can improve the initiative of epidemic prevention and control and take measures as early as possible, which can effectively reduce the harm and loss caused by the epidemic. In this study, the development trend of the COVID-19 epidemic in China in the relevant period of 2021 was studied. The content of the study was to use the grey GM(1,1) modeling theory to model the confirmed cases on the next 5 days, 30 days, 40 days and 60 days respectively, and to introduce the disaster theory in the grey system into the wave prediction of the epidemic [14]. Research results show that in 5 cases of the data to build the forecasting model of GM (1, 1) $\hat{X}^{(1)}(k+1) = -390.632862e^{-0.422304k} + 488.632862$, for simulating the raw data, and the simulation value of average residual rate is 0.268, The average simulation accuracy reached 73.2%. The prediction accuracy of the short-term prediction model based on the data of confirmed cases for five days reached 99.85%, and the result was satisfactory. The accuracy of the simulation value and the prediction accuracy of the epidemic wave prediction model were 97.77% and 97.33%, he prediction model of epidemic wave GM (1,1) established this, tim efunction is $\hat{X}^{(1)}(k+1)= 184.3950\ e^{0.226235\ K} - 175.3950$, with a model for simulating the raw data, and the simulation values of the average residual rate is 2.23%, the average simulation accuracy up to 97.67%. The prediction accuracy of the grey GM(1,1) model is 97.33%. Both the short-term prediction of confirmed cases and the waves time prediction of epidemic achieved satisfactory results.

4.3. Analysis of Reasons for the Poor Effect of Grey Modeling Theory in Predicting Medium and Long Term Confirmed Cases of the Epidemic

The epidemic development process of infectious diseases is often not a simple linear process, but a wavy epidemic development process. In this process, there will be many wave fronts and troughs. For a period of time, the confirmed cases of the epidemic fluctuate up, and then fluctuate down after reaching the peak value. Some of the original data may be at the moment of upward fluctuation of the epidemic, while others may be at the moment of downward fluctuation of the epidemic development. After forming the series of these original data, the internal coherence is poor, the direction conflicts, the trend is not obvious, and the dispersion is large. The GM(1,1)) model established by this method is difficult to reflect the law of epidemic trend, and its simulation data and model prediction data are both poor in accuracy, which cannot objectively reflect the real situation of the occurrence of medium and long-term confirmed cases.

4.4. Some Observations

In this study, the grey system theory is applied to predict the short, medium and long term epidemic and epidemic wave. First, when predicting the future epidemic trend of infectious diseases, GM (1,1) model should be established by using the latest actual confirmed cases, so that the original series can better reflect the internal relationship of data and the objective law of events [15]. In the establishment of grey GM(1.1) model for short-term prediction, the time interval of data extraction can be determined according to the length of the calm period of the epidemic, and the fixed model should be less used to forecast at multiple time points, so as to avoid low accuracy of prediction. Second, the data should be updated on a rolling basis, new parameters of the model should be calculated with the new data obtained, and a new prediction model should be established accordingly, so as to make the model more able to reflect the new inherent law of epidemic changes, so as to improve the simulation and prediction accuracy of the model. Third, the GM (1,1) model established in this study failed to achieve the expected effect in the medium and long term prediction of the epidemic trend of the novel coronavirus pneumonia. Due to the limitations of conditions, no residuals model correction study was conducted, and no other modeling methods of grey theory were used to make up for it. This needs to be further studied and explored in predicting the future epidemic trend of infectious diseases with the grey modeling theory [16]. Third, a variety of prediction models can be combined in the future trend prediction of epidemic and development of infectious diseases, to make up for the defects and deficiencies of a single prediction method, so as to make the prediction work more perfect and achieve the desired effect.

References

[1] Deng Julong, Foundation of Grey Theory, Huazhong University of Science and Technology Press, 2002. 02.

[2] Liu Sifeng Xie Naiming et al. Grey System Theory and Its Implications, Science Press, 2008.

[3] Dang Yaoguo, Liu Sifeng, Wang Zhengxin, Lin Yi et al. Research on Decision Model of Grey Prediction, Science Press, 2009. 12.

[4] YU Xu. Application of Grey correlation analysis Method in the layout of slope position shift monitoring points [J]. Water conservancy and hydropower letters, 2022 lancet (S1): 14-17 + 24. DOI: 10.15974/j.carol carroll nki SLSDKB. 2022. S1.004.

[5] Wu Quan, Yang Yanmei, Fu Yu. Prediction of total urban water use in Guiyang based on grey system theory [J]. Journal of Nanchang Institute of Technology, 2022, 41 (01): 40-46.

[6] Yu Guo, Li Haitao, Chen Yanru, Zhan Weiyun, Wang Chenyu, Zhang Dongming. Prediction of natural gas reserves growth trend in Sichuan Basin based on multi-cycle model [J]. Natural Gas Exploration and Development, 2021, 44 (01): 30-37.

[7] Chu Xianliang, Cui Tengfei, Xiang Tao, Li Yifan. Grey and time series model prediction [J]. Surveying, Mapping and Spatial Geographic Information, 2019, 42 (08): 199-202. (in Chinese).

[8] Li Ming Quan. Epidemic Characteristics and Prevention and Control Implications of COVID-19 in China [J]. European Journal of Preventive Medicine, 2021, 9 (4).

[9] Li Ming Quan. Prevention and Control Effects of Different Prevention and Control Mechanisms on the Epidemic of COVID-19. European Journal of Preventive Medicine. Vol. 10, No. 6, 2022, pp. 110-114. doi: 10.11648/j.ejpm.2022.10.06.11

[10] Zhao X F. (2011). Overview of grey system theory. Journal of Jilin Provincial Institute of Education (03), 152-154.

[11] PANG D Y. Trend Prediction of Transboundary Water Pollution Based on Grey Modeling [J]. Western Leather, 2019, 41 (08): 69-71. (in Chinese).

[12] Jia L Y. Grey increment model of population prediction and its application. Nanjing University of Information Science & Technology, 2006.

[13] PENG Xiaoxiao. Research on Prediction and Pre-Control Countermeasures of Construction Safety Accidents [D]. Ocean University of China, 2012.

[14] Zhai Xidong. Research on Prediction Model of Port Container Throughput [D]. Dalian University of Technology, 2006.

[15] Sun Jing. Research on Product Modeling Design Method Based on Image [D]. Wuhan University of Technology, 2007.

[16] Ma Le. Research on Grey Theory Modeling Method [D]. Dongbei University of Finance and Economics, 2005.

Biography

Li Ming Quan, male, was born in 1957 with a postgraduate degree in economics, senior animal husbandry division. Long - term engaged in agricultural and animal husbandry economic research work.