Iran J Public Health, Vol. 50, No.7, Jul 2021, pp.1405-1415

Original Article

Prediction of Mumps Incidence Trend in China Based on Difference Grey Model and Artificial Neural Network Learning

***Jin Jia* [1], *Mingming Liu* [2], *Zhigang Xue* [1], *Zhe Wang* [1], **Yu Pan* [3]**

1. *Information Center, The First Affiliated Hospital of Harbin Medical University, Harbin 150001, P.R. China*
2. *Clinical Laboratory, The First Affiliated Hospital of Harbin Medical University, Harbin 150001, P.R. China*
3. *Medical Department, The First Affiliated Hospital of Harbin Medical University, Harbin 150001, P.R. China*

***Corresponding Author:** Email: ppphhh6789@163.com

(Received 18 Jul 2020; accepted 21 Sep 2020)

Abstract

Background: We aimed to compare the prediction efficiency of back propagation (BP) network and grey model (GM) (1.1) for mumps infectious diseases and compare the application effect of the two models.

Methods: By calculating the average incidence rate of mumps in January 2014 -2016, we conducted the modeling of the BP time series, GM (1,1) grey model and the combination models of them, and predicted the incidence rate in June 2016 in comparison with the actual one. We compared the quarterly incidence rate to test the two prediction models, and compared the advantages and disadvantages of these models.

Results: R value of BP model was 68.45%, for GM (1,1) was 58.49%, and for combined forecasting model was 86.95%. We used the principal component analysis clustering method to control the samples, and found that the samples were close to the population mean. We found that the GM (1.1) model was more suitable for the prediction of mumps infection mode. We carried out dimension reduction analysis on the model data, and the accuracy of the data after dimension reduction is within the range of Da. For the discrete degree of the data in the combined model, matlab pipeline was used to verify the reliability of the data and results. By calculation after manifold optimization small error probability was P=0.875 and semi mean relative error 2.43%.

Conclusion: BP, GM (1,1) is a better method for modeling the epidemic trend of mumps in China, but the efficiency of prediction is not as high as the combination of them.

Keywords: Mumps; Back propagation network; Grey model; China

Introduction

Mumps is a common respiratory infectious disease in children. It is caused by the virus and spreads through the respiratory tract; in the case of non-immunization, it is a periodic epidemic; winter and spring are the peak season of the disease; children are the host; the virus antigen type is single, relatively stable without mutation; there are safe and effective vaccines for prevention (1). If the immunization strategy is appropriate, coupled with active epidemiological and laboratory monitoring, it can be eliminated or even eliminated (2-4).

The common prediction methods of infectious diseases are:

1. Qualitative prediction ① epidemic control chart method: seasonal epidemic or periodic epidemic; ② ratio figure method: infectious diseases with normal distribution of incidence; ③ Delphi method: to provide reference;
2. Quantitative prediction ① time series prediction model grey model: short-term prediction for diseases with stable epidemic factors B ∽ J model (ARIMA): Suitable for non-stationary time series with $n < 50$, it is a high-precision short-term prediction model. ② multi factor model: Multiple regression model, neural network model;
3. Comprehensive prediction method (5,6). Transmission dynamics needs certain data to determine parameters (7). Due to the particularity of mumps, some parameters are difficult to determine, for example, it is difficult to calculate the quantity. Therefore, due to the limited original data, it is difficult to make high-precision prediction or prediction at all, it is difficult to predict and control mumps (8-10).

How to make a breakthrough in theory and find a better method is an urgent work to be solved (11,12). Grey theory can effectively deal with the system with significant uncertainty and few data, and find out the rules of data from the disordered and limited discrete data, the basic principle of the grey prediction model is to determine the best fitting curve through the point group generated by the accumulation of the original sequence (13-15). Some work has been done on the prediction of mumps with cm (1,1) grey model. In this paper, a new method based on Back propagation (BP) neural network and GM (1,1) is proposed, This model overcomes the shortcoming of large sample data required by the conventional BP neural network. It not only takes advantage of the advantages of less sample data required by GM (1,1) grey model and high short-term prediction accuracy, but also has the advantages of parallel calculation, strong fault tolerance and strong adaptive ability of BP neural network. About one third of the world's population has a history of mumps infection (16).

In recent years, artificial neural network has been applied to medical research field and achieved good results. In this paper, BP neural network and grey theory are applied to the prediction of mumps infectious diseases, and through the establishment of prediction model, the feasibility of BP neural network and grey theory in forecasting epidemic infectious diseases is compared. BP neural network is composed of input layer, hidden layer, output layer and the connection between neurons in each layer. As shown in Fig. 1, BP algorithm divides the learning process of the network into two alternating processes: Forward propagation and back propagation (17).

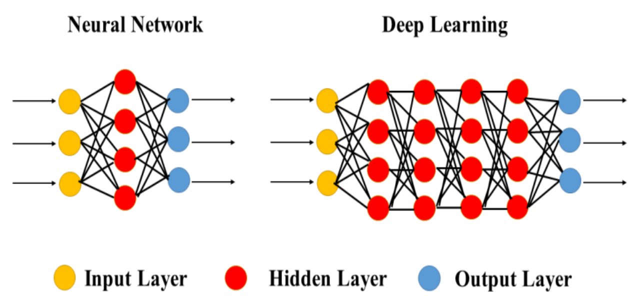

Fig. 1: The architecture of simple Deep Neural Networks

If the sum of squares of forward propagation output error fails to reach the expected precision, the weights and thresholds of neurons in each layer are corrected along the negative gradient direction of the error, and so on, until the total square error of the network reaches the expected accuracy.

In this study, we compared the quarterly incidence rate to test the two prediction models, and compared the advantages and disadvantages of the two models.

Methods

Data sources

The epidemic data of mumps in China were from China disease monitoring information report management system (http://www.chinacdc.cn/jkzt/crb/xcrxjb/201810/t20181017_195160.html). Demographic data from the annual demographic report of the National Health and Health Commission s http://www.nhc.gov.cn/.

1. Included subjects: ① literature on the effect of MUV vaccination on mumps prevention in China (excluding Hong Kong, Macao Special Administrative Region and Taiwan); ② the research design method was not limited, and the mumps antibody was detected before and after immunization for Chinese children (< 15 years old); ③the detection methods were hemagglutination inhibition (HI) test or enzyme-linked immunosorbent assay (enzyme-linked immunosorbent assay) (4) antibody positive conversion, including antibody negative before immunization, antibody positive after immunization, or antibody titer ≥ 4-fold increase after immunization.

2. exclusion criteria: ① animal experiment research; ② molecular biology research; ③ foreign research; ④ pharmacological research,

Construction of BP network model

The BP network model was mainly composed of input layer, hidden layer and output layer. The number of hidden layers adopted the following formula: $\sqrt{H+M} \leq$ L $\leq \sqrt{H+M}+\alpha$, which L is the number of hidden layers, H is the number of input layers, M is the number of output layers, and α is a constant of 1-10. It can be seen that the value of hidden layer L is 2-11, and the best hidden layer is selected according to the training results. The transfer function between neurons in the hidden layer was Tansig type; the transfer function in the output layer was Purelin type. Through a lot of experiments, this paper finally determines that the hidden layer is 5, which is a three-layer neural network with 1-5-1 topology (8).

GM (1,1) model construction

In GM (1,1) modeling, the original data was assumed to be Xo (i) = [x_0 (1), x_0 (2), x_0 (n)]. In order to overcome the strong randomness of the original data, the original data is processed by 1-AGO in the prediction, that is, x_1 (t) = x_1 (t-1) + x_0 (t). According to the relevant assumptions of grey theory, the time response equation can be obtained. After the parameter vector is calculated, it can be substituted into the formula to get the prediction model. The grey model is actually an accumulation generated number model, so it needs to be reduced after the prediction value is obtained.

Prediction model construction

Build BP network to predict mumps epidemic

Based on Jan 2014 to Jan 2016, the time of each month was the input layer, and the incidence rate of each month is output based on the observation BP network prediction model. Jan 2014 to Jan 2016 was selected as the input layer parameter of training samples every month, and then Jan 2016 to Apr 2019 was selected as the input layer parameter of each prediction sample. The incidence rate of each month was output layer. The input and output layers were normalized as shown in Fig. 1.

Build the grey theory model to predict the mumps epidemic

The G (1,1) model of mumps could be calculated by the above method on Jan. 12th, 2013, as shown in Table 1.

Table 1: GM (1,1) model for mumps epidemic prediction

Disease species	***GM(1.1)***
Incidence rate of male mumps	$\hat{x}_1(t+1) = 9.42e^{0.00528t} - 9.36$
Incidence rate of female mumps	$\hat{x}_1(t+1) = -13.758e^{-0.0032t} + 13.798$

Statistical analysis

The time series of DPS755 and BP neural network were used to establish the prediction model of mumps data and combined forecasting models were established on average daily incidence rate. Incidence rate of January 2016 to April 2019 was predicted and compared with actual situation. BP time series, GM (1,1) and combination prediction model were analyzed using dps7.05 software. Spss19.0 (Chicago, IL, USA) was used to calculate correlation coefficient.

Results

Epidemic situation of mumps

From Jan, 2016 to Apr, 2018, 430 thousands cases of mumps infectious diseases were reported in China, of which 150000 cases were in 2016; 150000 in 2017 and 120000 in 2018, and the total number of people in 2016-2018 was 137 million, with a total incidence rate of 312.93/ 100,000..

Two-dimensional clustering feature analysis of two methods

Feature selection tests, combined with different original sequence features and topology (network) features, could obtain higher accuracy, sensitivity, and specificity. The accuracy rate of the network attribute feature (Fig. 2A) was 84.43%, and the sensitivity and specificity were close to 78.24% and 90% respectively. We observed that the standardized and filtered network features (Fig. 2B) had the highest accuracy (84.76%), a sensitivity of 77.77%, and a specificity of 91.71%. Among the main sequence features, BP and GM performed well and could be followed up for testing, as shown in Fig. 2C.

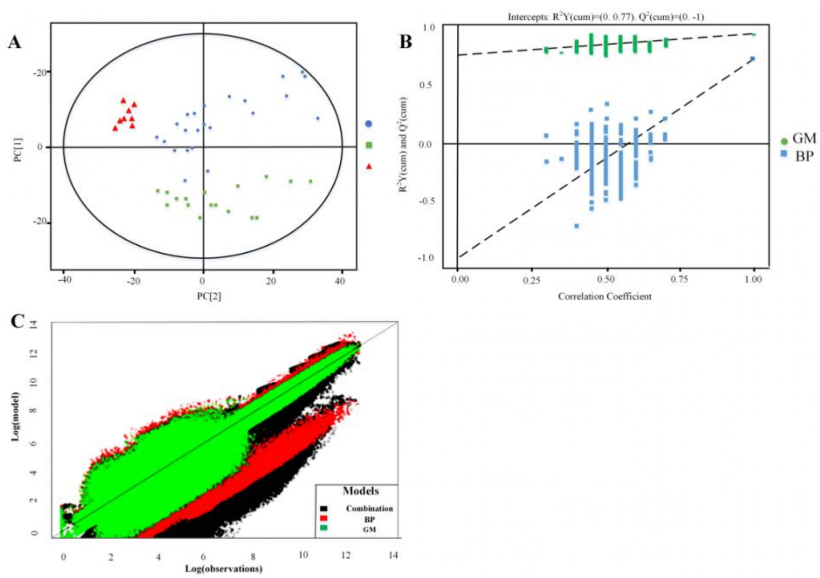

Fig. 2: Error function optimization after two-dimensional clustering of two groups of models (A). Scatter diagram of OPLS-DA scores of BP model to GM model, (B). Results of BP model 's replacement test for opls-da model of GM model, (C). BP model to GM model's binary permutation optimization function difference screening volcanic map

Time distribution

From the time distribution, the incidence rate showed a decreasing trend. The median monthly incidence rate in Jan 2016-Apr 2018 was 29.90/10,000,000, and the median daily incidence rate was 0.96/100,000.

Grey model (GM 1,1) for predicting the incidence rate of mumps in China

The mumps data from 2016 to 2018 in China predicted by grey model was taken as the original sequence x_0, and the cumulative generation sequence x and the adjacent mean equal weight generation sequence were calculated according to the formula in the method. The prediction results are showed in Table 2, Fig. 3. R value was 58.49%.

Table 2: Incidence rate of mumps GM (1,1) model fitting results

Time	*Sequential*	*Incidence rate*	*Predicted value*	*Residual*
201601-201603	1	66.5	65.5	-1
201604-201607	2	65.9	68.9	3
201608-201610	3	72.6	69.8	-2.8
201611-201701	4	76.9	74.2	-2.7
201702-201704	5	78.6	80.2	1.6
201705-201707	6	85.4	81.3	-4.1
201708-201710	7	78.6	84.9	6.3
201711-201801	8	85.9	88.9	3
201802-201804	9	90.5	95.8	5.3

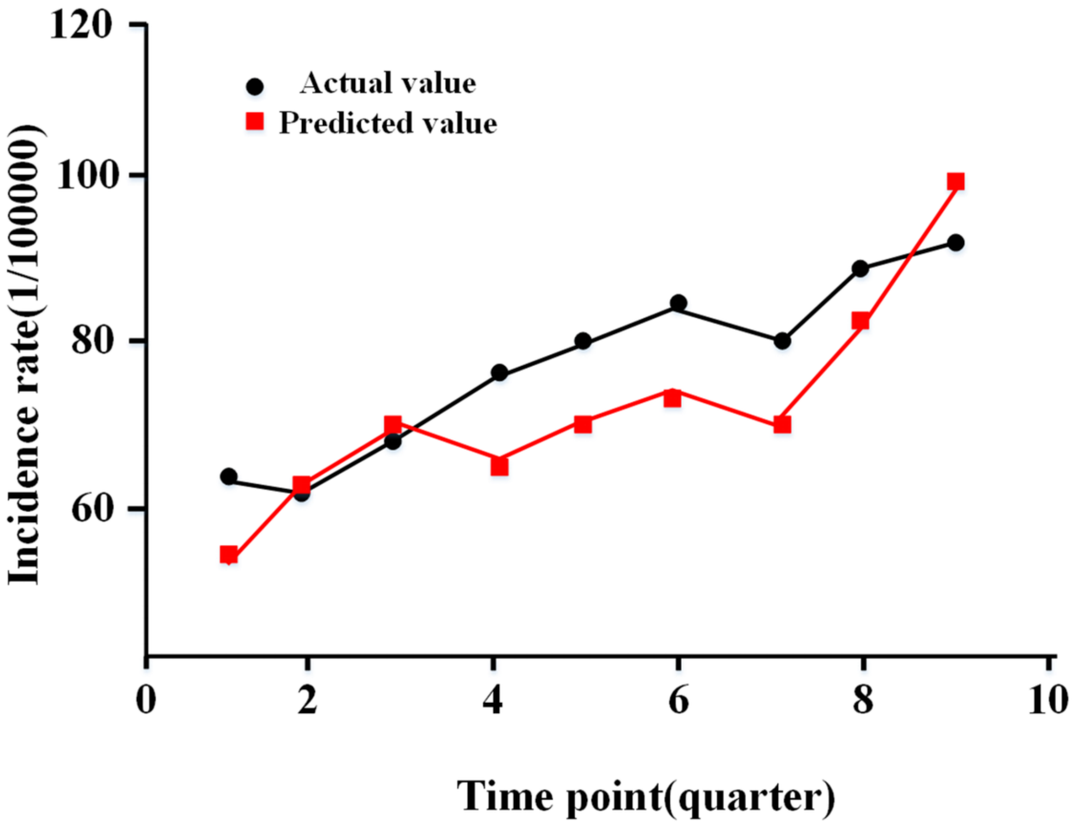

Fig. 3: Fitting diagram of GM (1,1) model prediction results. The predicted value of the autocorrelation function of mumps incidence in China is red, and the actual value is black

Prediction of incidence rate of mumps in China by BP network

Based on the data of the incidence rate of mumps in China in 2016-2018 (Table 3), the BP algorithm was applied to predict the incidence rate of mumps in 2018-2019. The process was implemented by Matlab software and R value was 68.45%. Fig. 4 shows that the BP prediction method was closer to the real value.

Table 3: Incidence rate of mumps BP model fitting results

Time	*Sequential*	*Incidence rate*	*Predicted value*	*Residual*
201601-201603	1	66.5	58.7	-7.8
201604-201607	2	65.9	66.5	0.6
201608-201610	3	72.6	75.6	3
201611-201701	4	76.9	70.1	-6.8
201702-201704	5	78.6	72.3	-6.3
201705-201707	6	85.4	74.6	-10.8
201708-201710	7	78.6	70.2	-8.4
201711-201801	8	85.9	81.5	-4.4
201802-201804	9	90.5	98.6	8.1

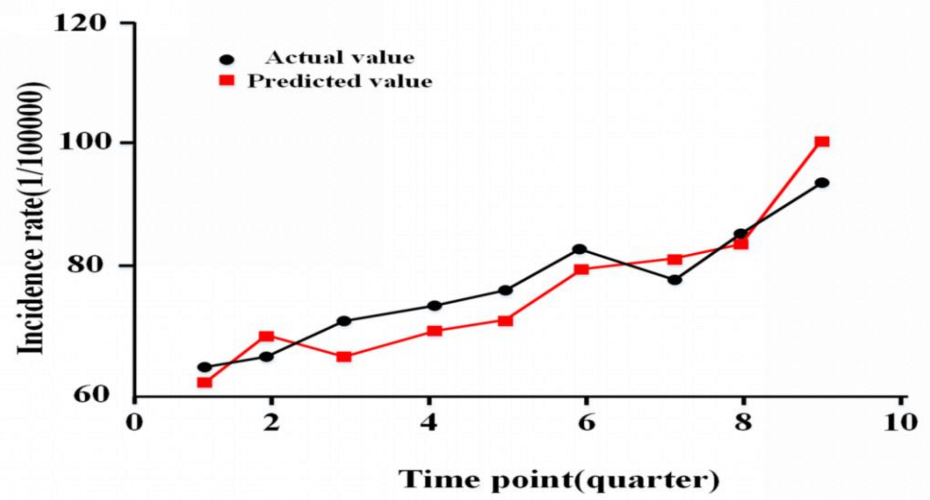

Fig. 4: Fitting diagram of BP network prediction results. The predicted value of the autocorrelation function of mumps incidence in China is red, and the actual value is black

Prediction of incidence rate of mumps in China by 4.3 combination model

The grey neural network combination model prediction results revealed R value as 86.95%. Table 4 was grey neural network combination model prediction result fitting graph. The results in Figure 5 illustrated that the fitting degree of BP was significantly higher than that of GM in the range of adaptive orbit prediction values.

Table 4: Grey neural network combination model prediction result fitting graph

Time	*Sequential*	*Incidence rate*	*Predicted value*	*Residual*
201601-201603	1	66.5	65.5	-1
201604-201607	2	65.9	68.9	3
201608-201610	3	72.6	71.5	-1.1
201611-201701	4	76.9	77.2	0.3
201702-201704	5	78.6	80.2	1.6
201705-201707	6	85.4	81.3	-4.1
201708-201710	7	78.6	79.2	0.6
201711-201801	8	85.9	88.9	3
201802-201804	9	90.5	89.5	-1

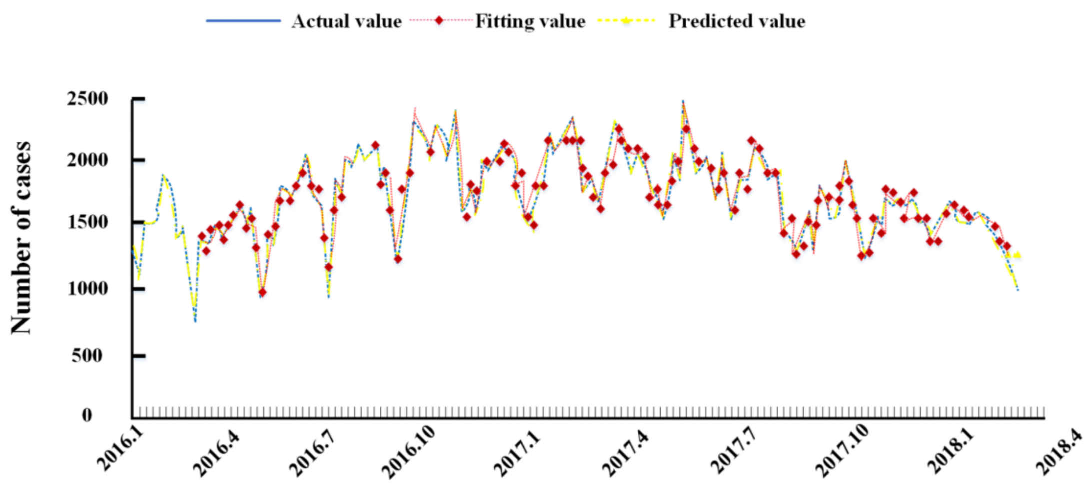

Fig. 5: Grey neural network combination model prediction result fitting graph

Comparison of BP, GM(1.1) and combination model of BP and GM

According to the GM time series model, BP model and their joint fitting and verification, the fitting degree of the two joint models was 92.34%, which was the highest (Table 5).

Table 5: Comparison of three main model parameters

Simulation name	*Fit simulation fit (%)*	*r*	*Sum of squares of validation model residuals*
GM	48.15	0.58	5234.55
BP	72.81	0.68	3482.34
Combination	92.34	0.86	2905.2

Discrete neural network model with self-feedback in system

The combined forecasting model was obtained based on GRNN by running the program in MATLAB. The weight coefficients of GM (1,1) model with residual correction and ARIMA (1,0,1) * (1,1,0) 2 model changed with time in the combined forecasting model based on GRNN. Due to the black box principle of neural network, the specific formula of the combined model could not be obtained. $\beta=1/2$, $\alpha_{12}= -1$, $\alpha_{13} =-1$. Obviously sin' (0) =1, arctan' (0) = 3, sin "(0) = arctan" (0) = 0, sin "(0) =-1 <0, arctan" (0). Derived from if D makes $\alpha_{21} = 0.50$, easy to get DD + if makes 0.49 easy If D = end <D, then (0.0,0) is gradually stable. If a21 = 0.51, easy to get D = such as> D: then (0,0,0) is gradually divergent. There was a periodic solution branched from (0,0,0). The above conclusions are verified in Fig.6. It could been seen all the prediction models were in the appropriate threshold range, and the comparison results were credible.

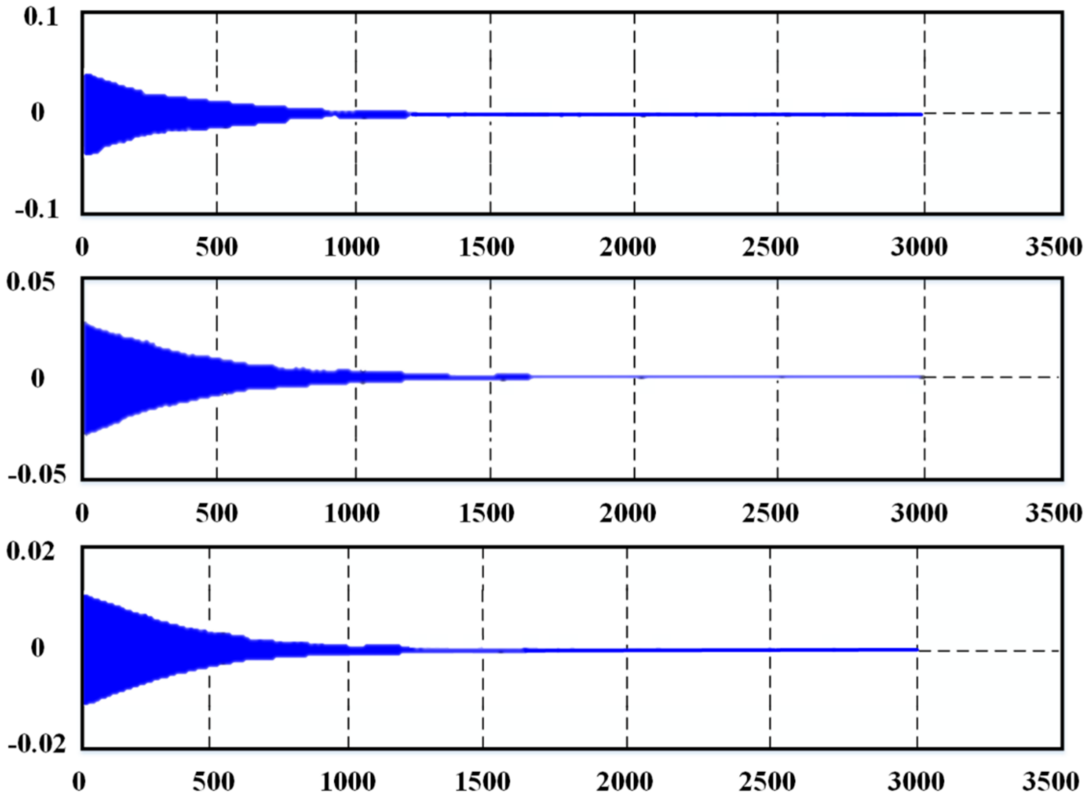

Fig. 6: Trajectory of fitting function x1, x2, x3 when sin "(0) = 0.50

Discussion

Based on the incidence rate of mumps in China, we used grey system theory to establish the model and predict the incidence rate of mumps. The posterior difference ratio was C-0.26, small error probability (P) was equal to 0.875 and semi mean relative error was 2.43%, which showed that the model had good accuracy grade, the prediction result was ideal, and the extrapolation result was credible.

The incidence rate of mumps in China continues to decline with an average annual rate of 2.83%. We analyzed the following reasons might be related: 1. with the implementation of the national infectious disease prevention and control plan, the incidence rate of tuberculosis in the whole country is decreasing gradually. 2. Since the strategy of infectious disease control has been carried out in China, we have focused on the detection and treatment of patients with mumps, and strengthened cooperation in medical treatment and prevention. Changes are closely related to time. But when predicting the incidence of infectious diseases in a one-dimensional time series, due to the different application conditions of different models, the prediction models fitted to the same infectious disease are also different, so the accuracy of fitting in different models is also different.

In this study, both BP network and GM (1,1) time series models were used to predict the incidence of mumps in China from 2016 to 2018 to compare the accuracy of BP network model and GM (1,1) model in predicting the incidence of infectious diseases (18,19). The BP neural network was trained with the predicted value as the input and the original data as the expected value, and the corresponding weights and thresholds were obtained. It could be predicted that the incidence rate of tuberculosis in China would continue to decrease year by year in the next few years. This was also the result of implementing the measures to prevent mumps in China (20, 21). In view of the difficulty in establishing the transmission dynamics of mumps disease, a combined model based on GM (1,1) grey model and BP neural network model was proposed in this paper.

It not only took advantage of the advantages of grey theory, such as less original data, simple principle, high short-term prediction accuracy, but also has the advantages of parallel calculation and strong adaptive ability of neural network. The combined model of grey model and BP neural network model could predict the number of mumps, which was better than the single GM (1,1) grey model, and could be used for the prediction of mumps. When the grey GM (1,1) model was used to predict, the sample size was small, and the randomness of the time series data was weakened by the ashing of the original data, but the prediction result was not good for the volatility fitting of the actual data. The advantage of grey model was that it could contain the comprehensive effect of multiple factors that affected the occurrence and development of disease in time variables. The prediction result of the grey model was a grey differential equation, which had a good prediction effect for the data with little volatility. However, when the data with large volatility were encountered, the prediction effect was not good. Grey model GM (1,1) had a good short-term prediction effect on incidence rate of mumps, and new information was needed for long-term prediction.

BP network model could reflect the volatility of the original data in the prediction, but its result was very random, and the model parameter setting had a great influence on the result (20, 21). Different parameters made different convergence speed of the network, and there might be the phenomenon that the prediction result had a large deviation or even no result. Theoretically, if the excitation function of the output layer was a linear function, then the BP network model could approach to any continuous function grey neural network combination model and had the advantages of grey model and neural network model (5, 13, 14). The prediction result could reflect the volatility of the original data well, and also weakened the randomness of the data. It was suitable for prediction problems in medical application.

The incidence of infectious diseases was affected by natural factors such as climate and geography, and the social factors such as living conditions, occupation, and social system. Therefore, the incidence rate fluctuated greatly and was difficult to predict. A single prediction method usually considers one aspect of the data, so there was a big error between the prediction results and the actual values. The combination model combined the advantages of the two models in a reasonable way, and the prediction results of the model were closer to the measured values. In this paper, the incidence rate of mumps in China was taken as an example, and the results were analyzed by combination model prediction. The combination model of grey neural network had good prediction effect and could be extended to other prediction problems in medical field.

Conclusion

The incidence rate of mumps in GM (1,1) and the error rate of external preprocessing can be reduced. Therefore, grey correlation analysis and principal component analysis are used to identify the factors. Then, GM is used to predict the rate of error. According to the data predicted by the experiment, when mumps is modeled by grey prediction and rolling grey prediction, the incidence of mumps is low. We can improve the accuracy of the external prediction of the unemployment rate. We can use grey prediction to do research. We can use GM (1,1) as far as possible, and the grey correlation factor of GM (1, n) is best combined with its joint degree analysis. However, the external prediction part of GM (1, n) can predict too many periods, or the value of 5 can be trained with the theory of neural like or other related categories, because of the influence error of rolling factor prediction, instead of substituting a normal 0.5, in that way the result may be better.

Ethical considerations

Ethical issues (including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgements

No funding was received in this study.

Conflict of interest

The authors declare that there is no conflict of interest.

Reference

1. Chen ZL, Wang XF (2018). [Nondestructive detection of total nitrogen content in leaves of Santalum album based on ST-PCA-BP neural network.]. *Ying Yong Sheng Tai Xue Bao,* 29(5):1551-1558. [In Chinese].
2. Huang L, Huang J, Wang W (2018). The Sustainable Development Assessment of Reservoir Resettlement Based on a BP Neural Network. *Int J Environ Res Public Health,*15(1): 146.
3. Liang W, Sun F (2019). Competing Endogenous RNA Network Analysis Reveals Pivotal ceRNAs in Adrenocortical Carcinoma. *Front Endocrinol (Lausanne),*10:301.
4. Liu L, Zhao T, Ma M, et al (2016). A new gene regulatory network model based on BP algorithm for interrogating differentially expressed genes of Sea Urchin. *Springerplus,* 5(1):1911.
5. Lyu J, Zhang J (2019). BP neural network prediction model for suicide attempt among Chinese rural residents. *J Affect Disord,*246:465-473.
6. Shi Y, Li Y, Cai M, et al (2019). A Lung Sound Category Recognition Method Based on Wavelet Decomposition and BP Neural Network. *Int J Biol Sci,* 15(1):195-207.
7. Wang L, Zhou W, Xing Y, et al (2018). A Novel Neural Network Model for Blood Pressure Estimation Using Photoplethesmography without Electrocardiogram. *J Healthc Eng,* 2018:7804243.
8. Wen H, Xie W, Pei J (2016). A Structure-Adaptive Hybrid RBF-BP Classifier with an Optimized Learning Strategy. *PLoS One,*11(10):e0164719.
9. Belyaletdinova IK, Mitrofanova IV, Kozlovskaya LI, et al (2020). Cases of aseptic meningitis after vaccination against mumps in Russia (2009-2019). *Public Health,* 186:8-11.
10. Hamidi H, Zarrineh M, Es-Haghi A, et al (2020). Rapid and sensitive determination of neomycin and kanamycin in measles, mumps, and rubella vaccine via high-performance liquid chromatography-tandem mass spectrometry using modified super-paramagnetic Fe3O4 nanospheres. *J Chromatogr A,* 1625:461343.
11. Wang F, Chen R, Jiang Q, et al (2020). Roles of Sialic Acid, AXL, and MER Receptor Tyrosine Kinases in Mumps Virus Infection of Mouse Sertoli and Leydig Cells. *Front Microbiol,*11:1292.
12. McPherson TD, Ramirez E, Ringness M, et al (2020). Mumps Cases Disproportionately Affecting Persons Living with HIV Infection and Men Who Have Sex with Men - Chicago, Illinois, 2018. *MMWR Morb Mortal Wkly Rep,* 69(28):909-912.
13. Filippi F, Patrizi A, Sabattini E, et al (2020). Pityriasis lichenoides triggered by measles-mumps-rubella vaccine injection. *J Dtsch Dermatol Ges,* 18(7):758-760.
14. Yang L, Grenfell BT, Mina MJ (2020). Waning immunity and re-emergence of measles and mumps in the vaccine era. *Curr Opin Virol,* 40:48-54.
15. Salehzadeh O, Vachon M, Sabourin N, et al (2020). Vertically tapered waveguide spot size converters fabricated via a linewidth controlled grey tone lithography for InP-based photonic integrated circuits. *Opt Express,* 28(16):23523-23533.
16. Qiu JS, Luo L, Jin RG (2020). [Grey model prediction of occupational diseases in Changsha]. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi,*38(7):508-511.
17. Lawson B, Suppiah S, Rota PA, et al (2020). In vitro inhibition of mumps virus replication by favipiravir (T-705). *Antiviral Res,*180:104849.
18. Zhou S, Liu N, Shen C, et al (2019). An adaptive Kalman filtering algorithm based on back-

propagation (BP) neural network applied for simultaneously detection of exhaled CO and N2O. *Spectrochim Acta A Mol Biomol Spectrosc,* 223:117332.

19. Zhu Q, Ding J, Ren W, et al (2019). [Study on Risk Assessment Model of in Vitro Diagnostic Reagent Adverse Events Based on BP Neural Network]. *Zhongguo Yi Liao Qi Xie Za Zhi,*43(2):136-139. [In Chinese].
20. Zha WT, Li WT, Zhou N, et al (2020). Effects of meteorological factors on the incidence of mumps and models for prediction, China. *BMC Infect Dis,*20(1):468.
21. Hassan J, Kelly J, De Gascun C (2020). Presence of low mumps-specific IgG in oral fluids is associated with high mumps viral loads. *J Clin Virol,* 129:104517.