

# Public Health

## Predicting the Trend of Infectious Diseases Using Grey Self-memory System Model: a Case Study of Incidence of Tuberculosis

--Manuscript Draft--

<b>Manuscript Number:</b>	
<b>Article Type:</b>	Original Research
<b>Keywords:</b>	Grey prediction theory; Grey self-memory system model; Infectious diseases; Incidence prediction
<b>Corresponding Author:</b>	Xiaojun Guo, Ph.D. Nantong University Nantong, Jiangsu CHINA
<b>First Author:</b>	Xiaojun Guo, Ph.D.
<b>Order of Authors:</b>	Xiaojun Guo, Ph.D. Houxue Shen Sifeng Liu Naiming Xie Yingjie Yang Jingliang Jin
<b>Abstract:</b>	<p><b>Background</b></p> <p>The prediction and early warning of infectious diseases is an important work in the field of public health. In order to predict the trend of infectious diseases affected by many uncertain factors, this study constructed the grey self-memory system prediction model. And tuberculosis was taken as an example to predict and analyze its incidence, so as to verify the effectiveness and application of this method.</p> <p><b>Methods</b></p> <p>By organically coupling the self-memory algorithm with the traditional mean GM(1,1) model, the tuberculosis incidence statistics of China from 2004 to 2018 was selected for prediction analysis. Because the self-memory equation uses the multiple time-points initial field to replace the single time-point's, it overcomes the weakness of the traditional grey prediction model which is sensitive to the initial value. Meanwhile, by comparing with the other traditional forecasting methods, three accuracy check indexes (MSE, AME, MAPE) were conducting to error analysis.</p> <p><b>Results</b></p> <p>Based on the comprehensive analysis of the three accuracy indexes, the GM self-memory model has significant forecasting advantages compared with the mean GM model and the other three statistical forecasting methods. With the help of the self-memory algorithm, the GM self-memory model makes full use of the historical data at multiple time-points, and its single-step rolling prediction accuracy is significantly better than other prediction methods. Therefore, the incidence of tuberculosis in China in the next year can be predicted as 55.30 (unit: 1/10<sup>5</sup>).</p> <p><b>Conclusion</b></p> <p>The grey self-memory system prediction model can closely capture the individual random fluctuation phenomenon in the whole evolution trend of the uncertain system. It is appropriate for predict the future evolution trend of the incidence of infectious diseases, and is worth popularizing and applying to other similar public health prediction problems.</p>
<b>Suggested Reviewers:</b>	Xuerui Tan, Ph.D.

Dean, Shantou University  
tanxuerui@vip.sina.com  
Prof. Xuerui Tan has been an active expert in the fields of grey systems theory and medical science.

Liping Zhang, Ph.D.  
Associate Director, Xinjiang Medical University  
zhanglp1219@163.com  
Dr Liping Zhang has been an active researcher in the fields of epidemiology and health statistics, grey systems theory and its application in medical.

Xiaojun Guo  
School of Science, Nantong University,  
Nantong, Jiangsu, China  
Tel: +86 13862960631,  
E-mail address: guoxj159@163.com

March 18, 2021

Dear Editor-in-chief,

On behalf of my co-authors, I would like to submit the manuscript entitled “*Predicting the Trend of Infectious Diseases Using Grey Self-memory Coupling Model: a Case Study of Incidence of Tuberculosis*” as an original article to *Public Health*.

In this paper, a grey self-memory system model was constructed to predict the trend of infectious diseases with uncertain characteristics. And tuberculosis was taken as an example to predict and analyze its incidence, so as to verify the effectiveness and application of this method.

We believe that two aspects of this manuscript will make it interesting to general readers of *Public Health*. **Above all**, this work achieves organic coupling of the traditional mean GM(1,1) model and the self-memory algorithm. The limitations of the traditional grey prediction model, i.e., being sensitive to initial value, can be eliminated by using multiple time-points initial field instead of only single time-point one in the system's self-memorization equation. **Further**, the grey self-memory system model can closely capture the individual stochastic fluctuations in the whole evolution trend of the uncertain system and is much more efficient than other existing common prediction models. It is appropriate for predict the future evolution trend of the incidence of infectious diseases, and is worth popularizing and applying to other similar public health prediction problems.

This is the first submission of this manuscript and no parts of this manuscript are being considered for publication elsewhere. All authors have approved this manuscript. No author has financial or other contractual agreements that might cause conflicts of interest.

In view of the research innovation in **grey prediction models and its applications on Public Health** of this research article, we hope that the experts in the fields of **grey systems theory and its medical applications** be considered as reviewers. With thanks for your consideration.

To study the increasing complexity, uncertainty and chaos of the system's structure, the grey systems theory was put forward especially suitable for small sample time sequence. As an important theoretical component, the grey prediction models have already been effectively utilized in numerous fields, ranging from social economy, engineering science, energy environment, transportation to **medicine and health**, and have demonstrated satisfactory prediction results.

In addition, although this work was partially supported by the National Statistical Science Research Project of China (Grant No. 2020LY020), but there is no funding for this project regretfully. Therefore, the project had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Thank you in advance for consideration of this manuscript, and we look forward to receiving comments from the editors and reviewers. If you have any queries, please don't hesitate to contact me at the address below.

Yours sincerely,

Xiaojun Guo, Ph.D.

# Predicting the Trend of Infectious Diseases Using Grey Self-memory System Model: a Case Study of Incidence of Tuberculosis

Xiaojun Guo <sup>a,b,\*</sup>, Houxue Shen <sup>a</sup>, Sifeng Liu <sup>b</sup>, Naiming Xie <sup>b</sup>, Yingjie Yang <sup>c</sup>, Jingliang Jin <sup>a,\*\*</sup>

<sup>a</sup> School of Science, Nantong University, Nantong 226019, China

<sup>b</sup> College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

<sup>c</sup> Institute of Artificial Intelligence, De Montfort University, Leicester LE1 9BH, UK

## Abstract

**Background:** The prediction and early warning of infectious diseases is an important work in the field of public health. In order to predict the trend of infectious diseases affected by many uncertain factors, this study constructed the grey self-memory system prediction model. And tuberculosis was taken as an example to predict and analyze its incidence, so as to verify the effectiveness and application of this method.

**Methods:** By organically coupling the self-memory algorithm with the traditional mean GM(1,1) model, the tuberculosis incidence statistics of China from 2004 to 2018 was selected for prediction analysis. Because the self-memory equation uses the multiple time-points initial field to replace the single time-point's, it overcomes the weakness of the traditional grey prediction model which is sensitive to the initial value. Meanwhile, by comparing with the other traditional forecasting methods, three accuracy check indexes (MSE, AME, MAPE) were conducting to error analysis.

**Results:** Based on the comprehensive analysis of the three accuracy indexes, the GM self-memory model has significant forecasting advantages compared with the mean GM model and the other three statistical forecasting methods. With the help of the self-memory algorithm, the GM self-memory model makes full use of the historical data at multiple time-points, and its single-step rolling prediction accuracy is significantly better than other prediction methods. Therefore, the incidence of tuberculosis in China in the next year can be predicted as 55.30 (unit:  $1/10^5$ ).

**Conclusion:** The grey self-memory system prediction model can closely capture the individual random fluctuation phenomenon in the whole evolution trend of the uncertain system. It is appropriate for predict the future evolution trend of the incidence of infectious diseases, and is worth popularizing and applying to other similar public health prediction problems.

**Keywords:** Grey prediction theory; Grey self-memory system model; Infectious diseases; Incidence prediction

---

\* Corresponding author. guoxj159@163.com (X. Guo).

\*\* Corresponding author. chinajjl@ntu.edu.cn (J. Jin).

# Predicting the Trend of Infectious Diseases Using Grey Self-memory System Model: a Case Study of Incidence of Tuberculosis

## Abstract

**Background:** The prediction and early warning of infectious diseases is an important work in the field of public health. In order to predict the trend of infectious diseases affected by many uncertain factors, this study constructed the grey self-memory system prediction model. And tuberculosis was taken as an example to predict and analyze its incidence, so as to verify the effectiveness and application of this method.

**Methods:** By organically coupling the self-memory algorithm with the traditional mean GM(1,1) model, the tuberculosis incidence statistics of China from 2004 to 2018 was selected for prediction analysis. Because the self-memory equation uses the multiple time-points initial field to replace the single time-point's, it overcomes the weakness of the traditional grey prediction model which is sensitive to the initial value. Meanwhile, by comparing with the other traditional forecasting methods, three accuracy check indexes (MSE, AME, MAPE) were conducting to error analysis.

**Results:** Based on the comprehensive analysis of the three accuracy indexes, the GM self-memory model has significant forecasting advantages compared with the mean GM model and the other three statistical forecasting methods. With the help of the self-memory algorithm, the GM self-memory model makes full use of the historical data at multiple time-points, and its single-step rolling prediction accuracy is significantly better than other prediction methods. Therefore, the incidence of tuberculosis in China in the next year can be predicted as 55.30 (unit:  $1/10^5$ ).

**Conclusion:** The grey self-memory system prediction model can closely capture the individual random fluctuation phenomenon in the whole evolution trend of the uncertain system. It is appropriate for predict the future evolution trend of the incidence of infectious diseases, and is worth popularizing and applying to other similar public health prediction problems.

**Keywords:** Grey prediction theory; Grey self-memory system model; Infectious diseases; Incidence prediction

---

## 1. Introduction

In recent years, various new types of diseases and traditional infectious diseases still occur from time to time in the whole world, which caused significant sudden public health incidents. The current and possible social risks caused by it are serious and have a far-reaching impact, which has attracted the attention of the government and researchers. The purpose is to deploy and deal with its impact on the society as soon as possible, so as to avoid all kinds of social risks. Although the global medical and health situation has been gradually improved effectively, public health emergencies still pose a huge threat to the lives and health of the people in China and even the whole world. At present, these incidents not only endanger life, health and safety, but also easily trigger social panic and public crisis, seriously threatening the stability and healthy development of the country and society.

The prevention and control of infectious diseases has increasingly become an important global public health issue. Scientific and reasonable analysis and prediction can lay a theoretical foundation for the in-depth development of disease early warning work, which will greatly reduce the social and economic harm of infectious diseases. And it is an important prerequisite for the formulation of short-term or long-term prevention and control strategies and measures. Prediction and early warning of infectious diseases is a mathematical analysis method that combines quantitative analysis tools such as mathematics and statistics with epidemiology. According to the occurrence and development laws of infectious diseases and related factors, this method can detect abnormal

precursors or adverse trends of the development of infectious diseases in early stage, thus improving the initiative and foresight of the prevention and control of infectious diseases. Therefore, the prediction and early warning of the spread, development and epidemic trend of infectious diseases is an important work in the management of health services, and is an important and effective way to control the epidemic and spread of infectious diseases. It is very important for prevention and control and emergency decision of health department [1].

With the rapid development of prediction theory, regression analysis [2], time series analysis [3], grey system model [4,5], Markov prediction model [6], artificial neural network model [7] and other methods have been used to predict the future trend of some infectious diseases. However, the general statistical forecasting methods have different limitations, such as requiring the data with large sample size, and requiring the historical data to obey some typical statistical distribution. Because the evolution path of infectious diseases (occurrence, development, epidemic) is affected by many uncertain factors, it has obvious characteristics of gray system, so it can be regarded as a typical dynamic variable gray system. At the same time, because the grey system model is not restricted by the constraints of the general statistical model on the original data, it has no strict requirements on the sample size and the statistical distribution of data, so it has strong practicability for the prediction of infectious diseases with stable epidemic factors.

Compared with the general prediction model (such as regression analysis method), the advantage of grey system model is that only a small amount of non-negative data can be used to predict the system behavior without the need to know the statistical distribution of the system. Grey prediction models [8-10] represented by mean GM model have been widely used in many scientific fields and achieved satisfactory results. Among them, in the field of epidemiology and public health, because the evolution path of infectious diseases is usually affected by numerous internal and external unstable factors and accompanied by irregular individual fluctuations, it can be regarded as a dynamic variable gray system with random fluctuations. At present, the grey prediction models have been effectively applied to the prediction analysis of the morbidity and mortality of all kinds of infectious diseases (such as parasitic diseases and malignant tumors) [11]. Moreover, the principle of self-memory [12] is an important method to improve the accuracy of model prediction. It can overcome the limitation of "sensitive to initial value" caused by the traditional GM model which only uses the single point initial value condition to solve the differential equation for prediction.

Among all kinds of infectious diseases, the incidence of tuberculosis, hepatitis and syphilis is relatively high, which have become several major infectious diseases threatening the life and health of the Chinese people. Tuberculosis and AIDS have relatively high mortality rates, far higher than other infectious diseases in terms of the number of deaths. Therefore, from the perspective of morbidity and death, tuberculosis is still one of the infectious diseases that threaten the life and health of the Chinese people. Considering the typical uncertainty characteristics of various infectious diseases, such as various transmission routes and high degree of pathogen variation, this paper takes tuberculosis as an example to predict the incidence by using the coupling model of mean GM and self-memory algorithm. It will help effectively control and reduce the incidence of all kinds of infectious diseases in China, and provide a reference for disease management departments to grasp the spread and development laws of infectious diseases such as tuberculosis, prevent and control major public health events and formulate emergency management strategies.

## 2. Materials & Methods

### 2.1. Data and analysis

In this paper, the incidence of tuberculosis (a representative infectious disease among Class A and B statutory reported infectious diseases in China) was taken as the research object. The annual data were obtained from China Health Statistics Yearbook 2019 [13]. Since the Chinese Center for Disease Control and Prevention was established in 2002 and was tested by the sudden SARS epidemic in 2003. And the strength of the early disease control system was relatively weak. Therefore, the statistical data of tuberculosis incidence from 2004 to 2018 were selected for modeling and analysis. The relevant data are shown in [Table 1](#).

**Table 1** Incidence of tuberculosis from 2004 to 2018 (unit:  $1/10^5$ ).

Year	2004	2005	2006	2007	2008	2009	2010	2011
Incidence	74.64	96.31	86.23	88.55	88.52	81.09	74.27	71.09
Year	2012	2013	2014	2015	2016	2017	2018	
Incidence	70.62	66.80	65.63	63.42	61.00	60.53	59.27	

Obviously, the time series of tuberculosis incidence showed an obvious exponential decline in general. But at the same time, due to unstable factors brought by social and economic influences, the incidence was accompanied by irregular individual fluctuation phenomena. China is one of the 30 countries with a high tuberculosis burden in the world. Due to the country's vast territory and large population, even a low incidence means a large number of cases.

## 2.2. Method

Based on the mean GM model, the self-memory prediction technique was introduced to construct a grey self-memory system model. The modeling process can be summarized as follows:

**Step 1** Determining the self-memory dynamic equation.

The whitening differential equation  $dx^{(1)}/dt = -ax^{(1)} + b$  contained in the mean GM(1,1) model [14] is taken as the self-memory dynamic equation of the grey self-memory coupling model:

$$dx/dt = F(x, \lambda, t) \quad (1)$$

where  $x$  is the variable,  $\lambda$  is the parameter,  $t$  is the time, and the dynamic kernel is  $F(x, \lambda, t) = -ax^{(1)} + b$ .

Let the memory function  $\beta(t)$ , and  $|\beta(t)| \leq 1$ , meanwhile, the inner product operation of Hilbert space is defined as:

$$(f, g) \equiv \int_{a_0}^{b_0} f(\xi)g(\xi)d\xi \quad (f, g \in L^2) \quad (2)$$

**Step 2** Calculating the self-memory difference-integral equation.

Set the time set  $T = \{t_{-p}, t_{-p+1}, \dots, t_{-1}, t_0, t\}$ , where  $t_{-p}, t_{-p+1}, \dots, t_{-1}$  represent the historical observation time points,  $t_0$  represents the initial prediction time point,  $t$  represents the future prediction time point, and  $p$  represents the retrospective order. Meanwhile, it is assumed that the sample interval of time point is  $\Delta t$ . And the variable  $x$  and the memory function  $\beta(t)$  are assumed to meet the conditions of continuous, differentiable and integrable. Then, with the aid of inner product operation (2), the self-memory dynamic equation (1) can be transformed into as follows:

$$\int_{t_{-p}}^{t_{-p+1}} \beta(\tau) \frac{\partial x}{\partial \tau} d\tau + \int_{t_{-p+1}}^{t_{-p+2}} \beta(\tau) \frac{\partial x}{\partial \tau} d\tau + \dots + \int_{t_0}^t \beta(\tau) \frac{\partial x}{\partial \tau} d\tau = \int_{t_{-p}}^t \beta(\tau) F(x, \lambda, \tau) d\tau$$

Through integral median formula, integration by parts, and reduction of similar terms, the prediction equation of self-memory can be deduced as follows:

$$\beta_t x_t - \beta_{-p} x_{-p} - \sum_{i=-p}^0 x_i^m (\beta_{i+1} - \beta_i) - \int_{t_{-p}}^t \beta(\tau) F(x, \lambda, \tau) d\tau = 0 \quad (3)$$

where  $\beta_i \equiv \beta(t)$ ,  $x_i \equiv x(t)$ ,  $\beta_i \equiv \beta(t_i)$ ,  $x_i \equiv x(t_i)$ , median  $x_i^m \equiv x(t_m)$ ,  $t_i < t_m < t_{i+1}$ ,  $i = -p, -p+1, \dots, 0$ .

Let  $x_{-p-1}^m \equiv x_{-p}$ ,  $\beta_{-p-1} \equiv 0$ , Equation (3) can be transformed into

$$x_t = \frac{1}{\beta_t} \sum_{i=-p-1}^0 x_i^m (\beta_{i+1} - \beta_i) + \frac{1}{\beta_t} \int_{t_{-p}}^t \beta(\tau) F(x, \lambda, \tau) d\tau = S_1 + S_2 \quad (4)$$

In this equation, the retrospective order is  $p$ , the self-memory term  $S_1$  represents the influence of the historical statistical data on the predicted value  $x_t$  at  $p+1$  time points, and the other effect term  $S_2$  represents the influence of the dynamic kernel source function  $F(x, \lambda, t) = -ax^{(1)} + b$  in the retrospection period  $[t_{-p}, t_0]$ .

**Step 3** Discretizing the self-memory prediction equation.

In Equation (4), the integral is approximated by summation, the differential is approximated by difference, and the median  $x_t^m$  is approximated by two adjacent time point values, that is  $x_t^m = (x_{i+1} + x_i)/2 \equiv y_i$ . Meanwhile, equidistant time point intervals are taken at the same time, and let  $\Delta t_i = t_{i+1} - t_i = 1$ , then  $\beta_t$  and  $\beta_i$  are combined to obtain the discrete form of the self-memory prediction equation:

$$x_t = \sum_{i=-p-1}^{-1} \alpha_i y_i + \sum_{i=-p}^0 \theta_i F(x, \lambda, i) \quad (5)$$

where memory coefficients are  $\alpha_i = (\beta_{i+1} - \beta_i)/\beta_i$  and  $\theta_i = \beta_i/\beta_i$ , and the dynamic kernel source function is  $F(x, \lambda, t) = -ax^{(1)} + b$ .

**Step 4** Estimating the memory coefficients by least squares.

$F(x, \lambda, t)$  is regarded as the system input,  $x_t$  is regarded as the system output. Assuming that there is  $L(L > p)$  time points of the original data sequence, the least square method can be used to get the memory coefficients  $\alpha_i$  and  $\theta_i$ . Set

$$X_t = \begin{bmatrix} x_{t1} \\ x_{t2} \\ \vdots \\ x_{tL} \end{bmatrix}, \quad Y = \begin{bmatrix} y_{-p-1,1} & y_{-p,1} & \cdots & y_{-1,1} \\ y_{-p-1,2} & y_{-p,2} & \cdots & y_{-1,2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{-p-1,L} & y_{-p,L} & \cdots & y_{-1,L} \end{bmatrix}, \quad A = \begin{bmatrix} \alpha_{-p-1} \\ \alpha_{-p} \\ \vdots \\ \alpha_{-1} \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} F(x, \lambda, -p)_1 & F(x, \lambda, -p+1)_1 & \cdots & F(x, \lambda, 0)_1 \\ F(x, \lambda, -p)_2 & F(x, \lambda, -p+1)_2 & \cdots & F(x, \lambda, 0)_2 \\ \vdots & \vdots & \ddots & \vdots \\ F(x, \lambda, -p)_L & F(x, \lambda, -p+1)_L & \cdots & F(x, \lambda, 0)_L \end{bmatrix}, \quad \Theta = \begin{bmatrix} \theta_{-p} \\ \theta_{-p+1} \\ \vdots \\ \theta_0 \end{bmatrix}$$

Then the discrete form of the self-memory prediction equation (5) can be expressed in matrix form  $X_t = YA + \Gamma\Theta$ .

If let  $Z = [Y, \Gamma]$ ,  $W = [A, \Theta]^T$ , the equation becomes  $X_t = ZW$ , and the least square estimation of the memory coefficient matrix  $W = [A, \Theta]^T$  is obtained as:

$$W = (Z^T Z)^{-1} Z^T X_t \quad (6)$$

**Step 5** Solving the self-memory prediction model.

By substituting the memory coefficients  $\alpha_i$  and  $\beta_i$  determined by Equation (6) into the self-memory discrete prediction equation (5), the corresponding simulated and predicted value  $\hat{x}^{(1)}(t)$  can be obtained. The original data simulation and prediction sequence  $\hat{X}^{(0)}$  of the grey self-memory coupling model can be further obtained through first-order reduction

$$\hat{x}^{(0)}(t) = \hat{x}^{(1)}(t) - \hat{x}^{(1)}(t-1) \quad (7)$$

where  $t = 1, 2, \dots, n$ , and  $\hat{x}^{(1)}(0) \equiv 0$ .

**2.3. Accuracy check**

The accuracy check of simulation and prediction is an important criterion to judge the prediction model. The prediction model must pass the accuracy check before extrapolation and application, so as to judge the reliability



and robustness of the prediction model. In practice, a variety of accuracy checking methods can be used to determine whether the prediction model is reasonable or not. Considering the typical grey uncertainty characteristics of the grey self-memory system model, this paper adopts the following three general checking standards [15,16] to compare the accuracy of different prediction models: 1) Mean squared error (MSE); 2) Absolute mean error (AME); 3) Mean absolute percentage error (MAPE).

MAPE is a widely accepted accuracy check standard, which can objectively give the relative difference between the original and predicted values. MSE and AME are two kinds of check standards to measure the mean magnitude of prediction error. MSE can be used to judge the degree of dispersion between the predicted value and the original value. By calculating the absolute value, AME can eliminate the influence of extreme values in positive and negative directions and the addition and subtraction displacement of errors [16]. The three accuracy check standards are expressed as follows:

Set the original sequence as  $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ , and the predicted sequence as  $\hat{X}^{(0)} = \{\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \dots, \hat{x}^{(0)}(n)\}$ , denote the relative error of the time point  $k$  as  $APE(k)$ , i.e.

$$APE(k) = \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right| \times 100\%, \text{ then}$$

$$MSE = \frac{1}{n} \sum_{k=1}^n (x^{(0)}(k) - \hat{x}^{(0)}(k))^2$$

$$AME = \frac{1}{n} \sum_{k=1}^n |x^{(0)}(k) - \hat{x}^{(0)}(k)|$$

$$MAPE = \frac{1}{n} \sum_{k=1}^n \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right| \times 100\%$$

The prediction accuracy of the grey self-memory system model can be tested by the relative error  $APE(k)$  of the time point  $k$  and the three system mean errors (MSE, AME and MAPE), and the model error analysis can be made.

### 3. Results

The incidence of tuberculosis was predicted by constructing a grey self-memory system model, which was compared and analyzed with traditional prediction models such as mean GM model, Holt linear trend method [17], one-time exponential smoothing method [18] and simple moving average method [19]. In the modeling and analysis, the incidence data of the first fourteen time points were taken as the modeling sample, and the incidence data of the 15th time point were taken as the test sample to conduct the prediction test. The single-point prediction errors of the models were analyzed by the relative errors, and the fitting accuracies of each models were evaluated comprehensively by the three accuracy check standards including MSE, AME and MAPE. Finally, by comparing the predictive ability between the grey self-memory system model and the other four traditional models, the most suitable model was selected to infer the future incidence trend of tuberculosis.

According to the statistical data of tuberculosis incidence from 2004 to 2017, the whitening differential equation of the mean GM(1,1) model is

$$\frac{dx^{(1)}}{dt} = -0.040294x^{(1)} + 99.240605$$

If the right end of the above equation is taken as the dynamic kernel  $F(x, t)$  of the self-memory equation, then  $dx/dt = F(x, t)$ , the GM self-memory prediction model of tuberculosis incidence can be established

accordingly, and the corresponding self-memory discrete prediction equation  $x_t = \sum_{i=-2}^{-1} \alpha_i y_i + \sum_{i=-1}^0 \theta_i F(x, i)$  can be

established. The retrospective order can be determined to be the optimal value  $p = 1$  after trial calculation based

on the minimum principle of fitting root mean square error, and the memory coefficient matrix can be obtained by using the least square estimation as follows:

$$W = \begin{bmatrix} \alpha_{-2} \\ \alpha_{-1} \\ \theta_{-1} \\ \theta_0 \end{bmatrix} = \begin{bmatrix} -0.107903 \\ 1.107079 \\ 16.451061 \\ -15.738718 \end{bmatrix}$$

Respectively by using mean GM model, Holt linear trend model, exponential smoothing method, simple moving average method and the aboved GM self-memory system model, the modeling prediction and error analysis for tuberculosis incidence are carried out, the results obtained are shown in Table 2. Thereinto, because of the cause of the modeling mechanism, the simulated predicted values at part of the initial point of the model are default.

**Table 2** Comparison of predicted values and errors of tuberculosis incidence under different models (unit:  $1/10^5$ ).

Year	Actual value	GM self-memory system model		Mean GM model		Holt linear trend method		Exponential smoothing method		Simple moving average method	
		Predicted value	APE / %	Predicted value	APE / %	Predicted value	APE / %	Predicted value	APE / %	Predicted value	APE / %
2004	74.64	—	—	—	—	—	—	—	—	—	—
2005	96.31	—	—	94.320	2.066%	—	—	96.310	0.000%	—	—
2006	86.23	86.569	-0.393%	90.595	-5.062%	96.310	-11.690%	96.310	-11.690%	—	—
2007	88.55	86.690	2.101%	87.017	1.731%	87.238	1.482%	91.270	-3.072%	—	—
2008	88.52	85.896	2.964%	83.580	5.581%	88.419	0.114%	89.910	-1.570%	90.363	-2.082%
2009	81.09	83.025	-2.386%	80.280	0.999%	88.510	-9.150%	89.215	-10.020%	87.767	-8.234%
2010	74.27	77.080	-3.783%	77.109	-3.823%	81.832	-10.182%	85.153	-14.653%	86.053	-15.865%
2011	71.09	70.194	1.260%	74.064	-4.183%	75.026	-5.537%	79.711	-12.127%	81.293	-14.352%
2012	70.62	67.932	3.806%	71.139	-0.735%	71.484	-1.223%	75.401	-6.770%	75.483	-6.886%
2013	66.80	68.261	-2.187%	68.329	-2.289%	70.706	-5.847%	73.010	-9.296%	71.993	-7.774%
2014	65.63	63.972	2.526%	65.631	-0.002%	67.191	-2.378%	69.905	-6.514%	69.503	-5.901%
2015	63.42	63.232	0.296%	63.039	0.601%	65.786	-3.731%	67.768	-6.856%	67.683	-6.722%
2016	61.00	61.004	-0.007%	60.549	0.739%	63.657	-4.356%	65.594	-7.531%	65.283	-7.021%
2017	60.53	58.554	3.264%	58.158	3.919%	61.266	-1.216%	63.297	-4.571%	63.350	-4.659%
MAPE			2.081%		2.441%		4.742%		7.282%		7.950%
2018	59.27	58.169	1.857%	55.861	5.752%	60.604	-2.251%	61.913	-4.459%	61.650	-4.016%

#### 4. Discussion

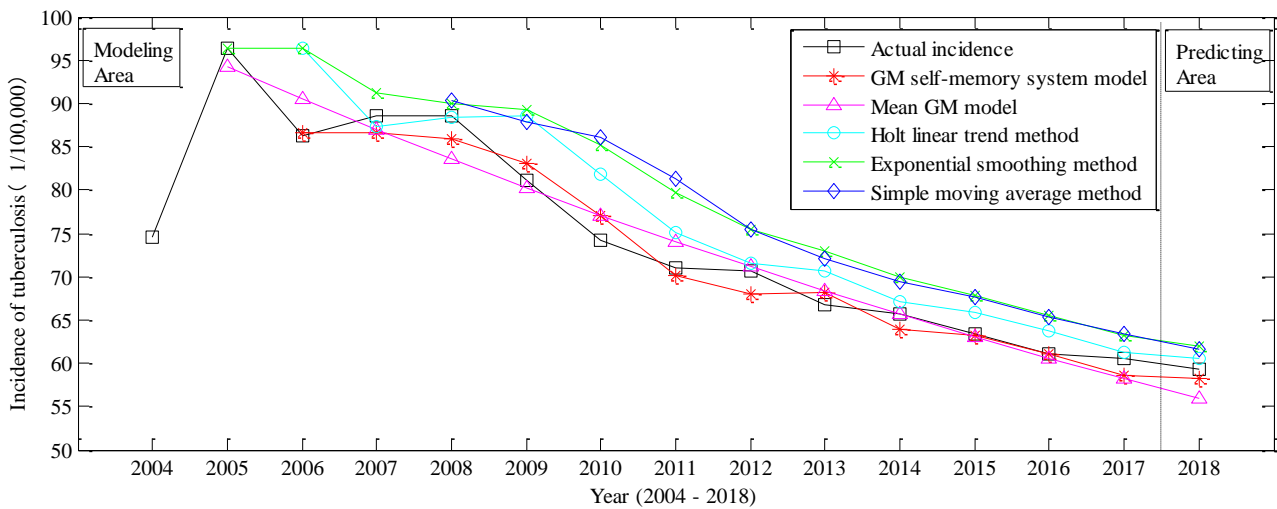
The three accuracy check standard values (MSE, AME, MAPE) of different tuberculosis incidence prediction models are listed in Table 3. From the inspection perspective of each accuracy index, the accuracy of GM self-memory system model meets the conditions of simulation prediction. Visible at the same time, the prediction error of mean GM model is lower than Holt linear trend method, exponential smoothing method and simple moving average method. And the self-memory algorithm helps the mean GM model to further reduce the prediction error and has a superior predictive ability. Therefore, the single step rolling forecast accuracy of the GM self-memory system model is significantly better than the other four traditional prediction models.

**Table 3** Accuracy check of simulation prediction results under five different models.

Prediction model	MSE	AME	MAPE
GM self-memory system model	3.245984	1.536583	2.081%
Mean GM model	5.839220	1.900308	2.441%
Holt linear trend method	21.892470	3.541750	4.742%
Exponential smoothing method	38.236752	5.291846	7.282%
Simple moving average method	40.100508	5.580100	7.950%

From the angle of the comprehensive analysis about the relative error  $APE(k)$  of single point and three kinds of accuracy check standard values (MSE, AME, MAPE), the new established GM self-memory system model was obviously better than mean GM model and the other three traditional statistical prediction models. It is appropriate for forecasting the incidence of tuberculosis of China, so as to predict the next annual incidence of tuberculosis is 55.30 (unit:  $1/10^5$ ), then it can provide effective decision-making of tuberculosis prevention and control guidance.

At the same time, Fig. 1-2 depict the fitting status with the actual curve of the simulated prediction curves of the tuberculosis incidence rate of the five comparative models, as well as the corresponding comparison of APE distribution. Comprehensive analysis shows that the new GM self-memory system model has significant predict advantage compared with the mean GM model and other statistical prediction models. While capturing the overall development trend of the original data well, the coupling model can effectively weaken the influence of the uncertain disturbance of external factors, and capture the random fluctuation trend of individuals closely with the help of the self-memory algorithm. It is a kind of stable reliable prediction model, which can be used to predict the future trend of the evolution of infectious disease incidence.



**Fig. 1.** Comparison of actual and simulated prediction values under five different models.

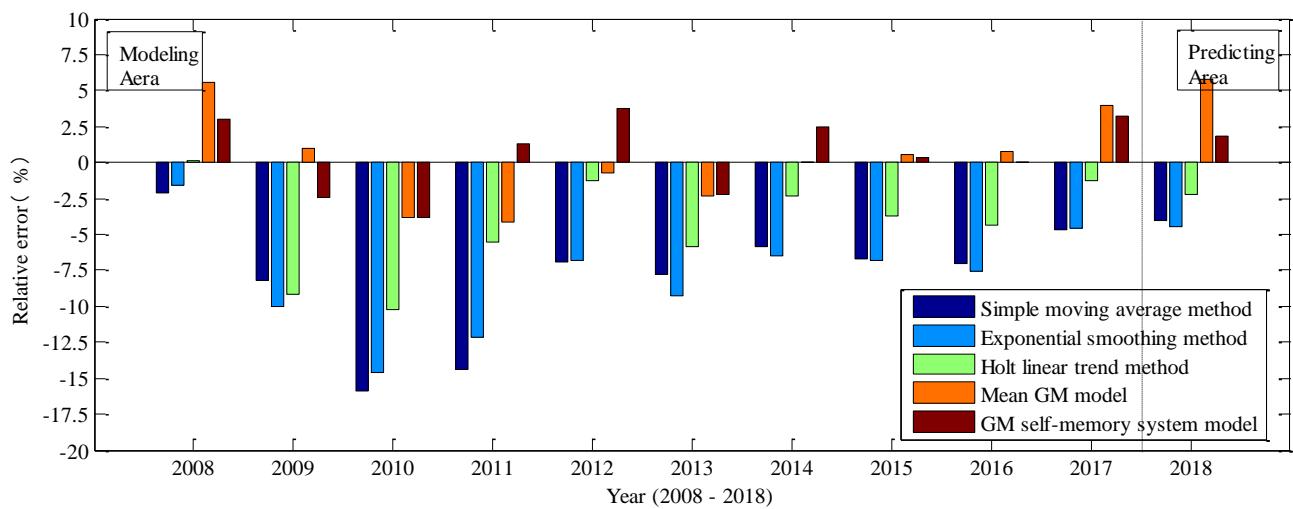


Fig. 2. Comparison of relative error distributions of five different prediction models.

## 5. Conclusions

Morbidity is the most direct and objective index to evaluate the effectiveness of prevention and control measures of an infectious disease, and whether the incidence of an infectious disease can be reduced is of great significance to the containment of the epidemic of such an infectious disease [20]. In terms of infectious disease trend prediction, the combination of Grey system model with a self-memory algorithm has better prediction effect compared with the traditional grey models and other statistical methods. At the same time, owing to the low requirements of the sample size and data distribution, the coupling model is different from the general statistical models for various constraint conditions of the original data. So it is suitable for the infectious diseases of popular factors relatively stable, and has a good application effect. The method is suitable for extrapolation prediction of infectious diseases, and it is easy to grasp the rules of occurrence, development and change. Therefore, it can better grasp the dynamic trend of infectious diseases, and provide reasonable and effective scientific basis for the formulation of emergency plan and prevention and control strategy. It is also worth popularizing and applying to the prediction problems in other similar areas of health management.

Meanwhile, Grey self-memory system model also has some limitations, which is generally suitable for short-term prediction. But with the lapse of time, some uncertain disturbance factors in the future will continue to enter the system and cause influence, then the accuracy will be significantly reduced if it is used for medium and long-term prediction. Therefore, if the incidence of infectious diseases is to be predicted in the medium and long term, the latest data can be used to modify the model and generate new information Rolling model for prediction. In addition, some old data that have little impact on the development of the system can be appropriately removed to improve the prediction accuracy.

### Declaration of competing interest

The authors have no competing interests to declare.

### Acknowledgments

This work was partially supported by the National Statistical Science Research Project of China (Grant No. 2020LY020). But the project had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### References

- [1] L.P. Zhang, Y.L. Zheng, X.L. Zhang, et al., Application of grey system theory in modern medicine, *Journal of Mathematical Medicine*. 28 (3) (2015) 430-434.

- [2] I.M. Hall, R. Gani, H.E. Hughes, et al., Real-time epidemic forecasting for pandemic influenza, *Epidemiology and Infection*. 135 (3) (2007) 372-285.
- [3] X.Y. Zhang, Y.Y. Liu, M. Yang, et al., Comparative study of four time series methods in forecasting typhoid fever incidence in China, *PLoS One*. 8 (5) (2013) e63116.
- [4] L.P. Zhang, Y.L. Zheng, K. Wang, et al., An optimized Nash nonlinear grey Bernoulli model based on particle swarm optimization and its application in prediction for the incidence of Hepatitis B in Xinjiang, China, *Computers in Biology and Medicine*. 49 (2014) 67-73.
- [5] L.P. Zhang, L. Wang, Y.L. Zheng, et al., Time prediction models for echinococcosis based on gray system theory and epidemic dynamics, *International Journal of Environmental Research and Public Health*. 14 (3) (2017) 262.
- [6] R. Yaesoubi, T. Cohen, Generalized Markov models of infectious disease spread: a novel framework for developing dynamic health policies, *European Journal of Operational Research*. 215 (3) (2011) 679-687.
- [7] V.S.H. Rao, M.N. Kumar, Estimation of the parameters of an infectious disease model using using neural networks, *Nonlinear Analysis: Real World Application*. 11 (3) (2010) 1810-1818.
- [8] S.F. Liu, Y.J. Yang, J. Forrest, *Grey Data Analysis Methods: Models and Application*, Berlin: Springer-Verlag. (2016) 142-149.
- [9] L.F. Wu, S.F. Liu, L.G. Yao, et al., Grey system model with the fractional order accumulation, *Communications in Nonlinear Science and Numerical Simulation*. 18 (7) (2013) 1775-1785.
- [10] L.F. Wu, S.F. Liu, Z.G. Fang, et al., Properties of the GM(1,1) with fractional order accumulation, *Applied Mathematics and Computation*. 252 (2015) 287-293.
- [11] Y.Q. Chen, Z.Y. Wei, Predicting the mortality rate of malignant tumors by the grey system GM(1,1) model, *Chinese Journal of Public Health*. 6 (8) (1990) 340-343.
- [12] H.X. Cao, Self-memory equation for atmosphere motion, *Science in China (Series B)*. 36 (7) (1993) 845-855.
- [13] Ministry of Health of the People's Republic of China, *China Health Statistics Yearbook 2019*, Beijing: China Union Medical College Press. (2019).
- [14] S.F. Liu, Y.J. Yang, L.F. Wu, et al., *Grey system theory and its application (The 7th edition)*, Beijing: science press. (2017) 140-146.
- [15] Z. Zhao, J.Z. Wang, J. Zhao, et al., Using a grey model optimized by differential evolution algorithm to forecast the per capita annual net income of rural households in China, *Omega*. 40 (5) (2012) 525-532.
- [16] Y.H. Lin, C.C. Chiu, P.C. Lee, et al., Applying fuzzy grey modification model on inflow forecasting, *Engineering Applications of Artificial Intelligence*. 25 (4) (2012) 734-743.
- [17] G. Yapar, S. Capar, H.T. Selamlar, et al., Modified holt's linear trend method, *Hacettepe Journal of Mathematics and Statistics*. 47 (5) (2018) 1394-1403.
- [18] B. Billah, M.L. King, R.D. Snyder, et al., Exponential smoothing model selection for forecasting, *International Journal of Forecasting*. 22 (2) (2006) 239-247.
- [19] F.R. Johnston, J.E. Boyland, M. Meadows, et al., Some properties of a simple moving average when applied to forecasting a time series, *Journal of the Operational Research Society*. 50 (1999) 1267-1271.
- [20] X.J. Shen, L.M. Ou, X.J. Chen, et al., The application of the grey disaster model to forecast epidemic peaks of typhoid and paratyphoid fever in China, *PLoS One*. 8 (4) (2013) e60601.